25

AMINO ACIDS, PEPTIDES, PROTEINS, ENZYMES, AND NUCLEIC ACIDS

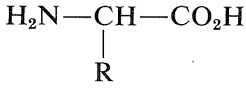
The chemistry of life is largely the chemistry of polyfunctional organic compounds. The functional groups usually are of types that interact rather strongly as, for example, the hydroxyl and carbonyl functions of carbohydrates (Chapter 20). The interaction between amino and carboxyl functions of amino acids figures greatly in the present chapter. We will approach the very important chemistry of amino acids and their derivatives in three stages. First, simple α-amino acids will be considered with emphasis on how the properties of amine functions and of acid functions are modified in molecules that possess both groups. Then we shall discuss some important properties of peptides and proteins, which are substances made up of amino acids linked together by amide bonds. Attention also will be given to the chemical problems presented by enzymes, which are protein molecules able to act as efficient catalysts for specific chemical reactions, and to the role of nucleic acids in protein synthesis.

25-1 TYPES OF BIOLOGICALLY IMPORTANT AMINO ACIDS

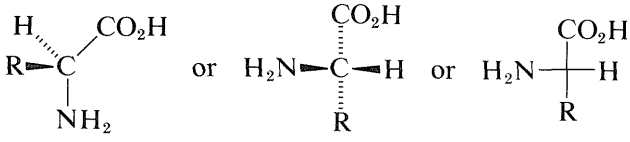
25-1A Protein Amino Acids

The amino acids that occur naturally as constituents of proteins have an amino group (NH_2) and a carboxylic acid group (CO_2H) attached to the *same*

carbon. They are called **α-amino acids** and have the general formula



They differ only in the nature of the R group on the α carbon and, with few exceptions, they are chiral molecules with the L configuration at the chiral α carbon:[1]



L-amino acid
(S-amino acid)

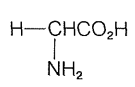
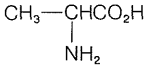
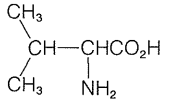
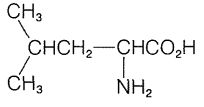
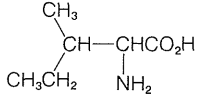
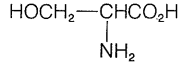
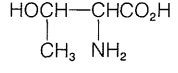
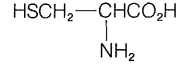
The structures and names of some particularly important α-amino acids are shown in Table 25-1. You will notice that the names in common use for amino acids are not descriptive of their structural formulas; but at least they have the advantage of being shorter than the systematic names. The abbreviations Gly, Glu, and so on, that are listed in Table 25-1 are particularly useful in designating the sequences of amino acids in proteins and peptides, as will become evident later in the chapter.

The nature of the substituent R varies considerably. In some amino acids, R is a hydrocarbon group, whereas in others it possesses functional groups such as OH, SH, SCH_3, CO_2H, or NH_2. Amino acids that have amine or other basic functions in the R group are called **basic amino acids** (lysine and arginine), whereas those with acidic groups are called **acidic amino acids** (aspartic and glutamic acids). Three of the amino acids listed in Table 25-1 (cysteine, cystine, and methionine) contain sulfur in —SH, —S—S—, and —SCH_3 groups. Cysteine and cystine can be interconverted readily with a wide variety of oxidizing and reducing agents according to the general reaction $2RSH \underset{[H]}{\overset{[O]}{\rightleftharpoons}} RSSR$. This is an important process in the biochemistry of sulfur-containing peptides and proteins (Section 25-8A).

The α-amino function of the common amino acids is primary —NH_2 in all except proline and hydroxyproline. Several of the amino acids have aromatic R groups (phenylalanine, tyrosine, tryptophan), while histidine and tryptophan have azarene R groups.

[1]A number of D-amino acids have been found to be constituents of peptides in the cell walls of bacteria.

Table 25-1
Amino Acids Important as Constituents of Proteins

Name	Abbreviations 3-letter	1-letter	Structure[a]	$pK_{a_{(CO_2H)}}$	$pK_{a_{(\alpha\text{-}NH_3^{\oplus})}}$	$pK_{a_{(R)}}$	pI	Solubility[b]	Mp, °C
glycine	Gly	G		2.34	9.60		5.97	22.5	292 dec
alanine	Ala	A		2.35	9.69		6.02	15.8	297 dec
valine[c]	Val	V		2.32	9.62		5.97	6.8	315 dec
leucine[c]	Leu	L		2.36	9.60		5.98	2.4	337 dec
isoleucine[c]	Ile	I		2.36	9.68		6.02	2.1	285 dec
serine	Ser	S		2.21	9.15		5.68	4.3	228 dec
threonine[c]	Thr	T		2.09	9.10		5.60	1.6	253 dec
cysteine	Cys or CySH	C		1.71	10.8	8.3	5.02	very sol.	–

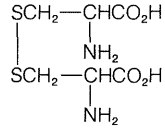
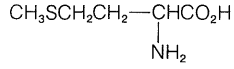
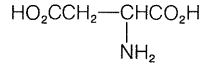
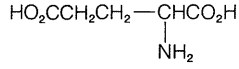
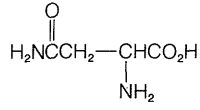
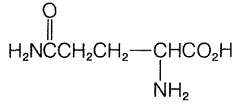
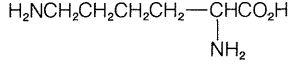
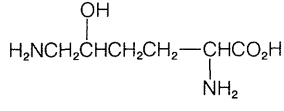
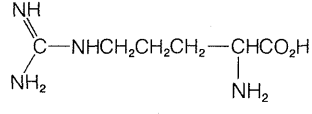
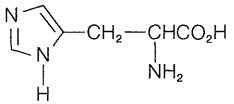
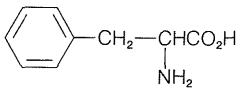
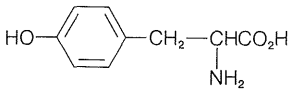
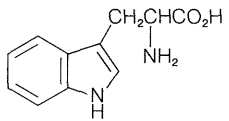
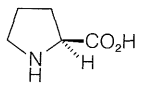
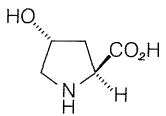
cystine	Cys or Cys	Cy S S Cy		1.65 2.26	7.86 9.85		5.06	0.009	258
methionine	Met	M		2.28	9.21		5.06	3.0	283
aspartic acid	Asp	D		2.09	9.82	3.86	2.98	0.4	269
glutamic acid	Glu	E		2.19	9.67	4.25	3.22	0.7	247
asparagine	Asn	N		2.02	8.8		5.41	2.4	236
glutamine	Gln	Q		2.17	9.13		5.70	3.6[18]	184
lysine[c]	Lys	K		2.18	8.95	10.53	9.74	very sol.	224
hydroxylysine	Hyl			2.13	8.62	9.67	9.15	very sol.	—
arginine	Arg	R		2.17	9.04	12.48	10.76	very sol.	230–244 dec

Table 25-1 (continued)
Amino Acids Important as Constituents of Proteins

Name	Abbreviations 3-letter	1-letter	Structure[a]	$pK_{a_{(CO_2H)}}$	$pK_{a_{(\alpha\text{-}NH_3^{\oplus})}}$	$pK_{a_{(R)}}$	p*I*	Solubility[b]	Mp, °C
histidine	His	H		1.82	9.17	6.0	7.59	4.0	287
phenylalanine[c]	Phe	F		1.83	9.13		5.48	2.7	283
tyrosine	Tyr	Y		2.20	9.11	10.07	5.67	0.04	342
tryptophan	Trp	W		2.38	9.39		5.88	1.1	283
proline	Pro	P		1.99	10.60		6.30	154.5	220
hydroxyproline[d]	Hyp			1.92	9.73		6.33	34.5	270

[a] For convenience only, the structures are represented as neutral nonpolar molecules. In reality, ionic and dipolar forms are present in aqueous solution in amounts dependent on the pH (Section 25-2A).

[b] Water solubility at isoelectric point of the L isomer in g/100 g at 20°C. The D,L mixtures are usually less soluble.

[c] Must be included in diet for maintenance of proper nitrogen equilibrium in normal adult humans.

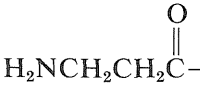
[d] Found only in collagen.

Exercise 25-1 Select the amino acids in Table 25-1 that have more than one chiral center and draw projection formulas for all the possible stereoisomers of each which possess the L configuration at the α carbon.

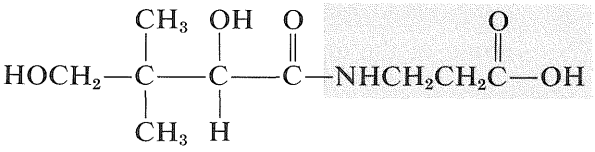
Exercise 25-2 Which of the amino acids in Table 25-1 are *acidic* amino acids and which *basic* amino acids? Which of the structures shown would have the most basic nitrogen? The least basic amino nitrogen? Give the reasons for your choices. (Review Section 23-7.)

25-1B Nonprotein Amino Acids

The most abundant amino acids are those that are protein constituents and these are always α-amino acids. However, there are many other amino acids that occur naturally in living systems that are not constituents of proteins, and are not α-amino acids. Many of these are rare, but others are common and play important roles in cellular metabolism. For example, 3-aminopropanoic acid is a precursor in the biosynthesis of the vitamin, pantothenic acid,[2]

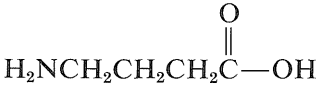


3-aminopropanoic acid (β-alanine)



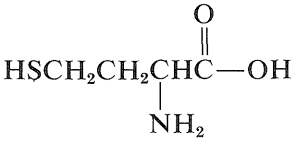
pantothenic acid

and 4-aminobutanoic acid is involved in the transmission of nerve impulses.

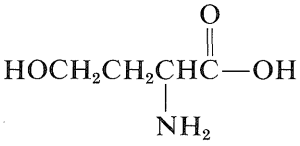


4-aminobutanoic acid (γ-aminobutyric)

Homocysteine[3] and homoserine are among the important α-amino acids that are not constituents of proteins. These substances are precursors in the biosynthesis of methionine.



homocysteine



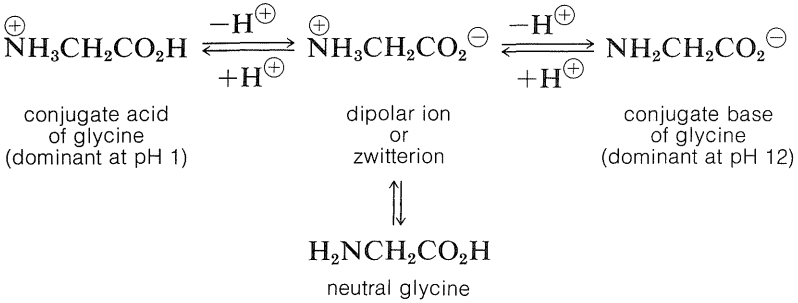
homoserine

[2]Pantothenic acid is in turn a precursor for the synthesis of coenzyme A, which is essential for the biosynthesis of fats and lipids (Sections 18-8F and 30-5A).

[3]The prefix *homo* implies an additional carbon in the longest chain.

25-2 THE ACID-BASE PROPERTIES OF α-AMINO ACIDS

The behavior of glycine is reasonably typical of that of the simple amino acids. Because glycine is neither a strong acid nor a strong base, we shall expect a solution of glycine in water to contain four species in rapid equilibrium. The proportions of these species are expected to change with pH, the cationic conjugate acid being the predominant form at low pH and the anionic conjugate base being favored at high pH:



Spectroscopic measurements show that the equilibrium between neutral glycine and the dipolar ion favors the dipolar ion by at least 100 to 1. This is to be expected because the $\overset{\oplus}{H_3}N-$ group of the dipolar ion will stabilize the $-CO_2^{\ominus}$ end while the $-CO_2^{\ominus}$ group will stabilize the $\overset{\oplus}{H_3}N-$ end.

The acid-ionization constant of $\overset{\oplus}{H_3}NCH_2CO_2H$ is 4.5×10^{-3} ($pK_a = 2.34$, Equation 25-1), which is about 25 times greater than K_a for ethanoic acid. (Section 18-2). This is expected because of the electron-attracting electrostatic effect of the $\overset{\oplus}{H_3}N-$ group. Ionization of the $\overset{\oplus}{H_3}N-$ group of the dipolar ion ($K_a = 2.0 \times 10^{-10}$; $pK = 9.60$; Equation 25-2) is oppositely affected by the electrostatic effect of the $-CO_2^{\ominus}$ group and is 10 times less than of ethanammonium ion (Section 23-7B). The manner in which the concentrations of the charged glycine species change with pH is shown in Figure 25-1. Notice that, between pH 3 and pH 8, almost all of the glycine is in the form of the dipolar ion. The pH at the center of this range, where the concentration of $\overset{\oplus}{H_3}NCH_2CO_2H$ is equal to the concentration of $H_2NCH_2\overset{\ominus}{C}O_2$, is called the **isoelectric point**, pI, and usually corresponds to the pH at which the amino acid has minimum water solubility. Isoelectric points for the amino acids are shown in Table 25-1. The isoelectric points are the average of the pK_a values for dissociation of the monocation and the dipolar ion forms of the amino acid. For glycine, $pI = (2.34 + 9.60)/2$.

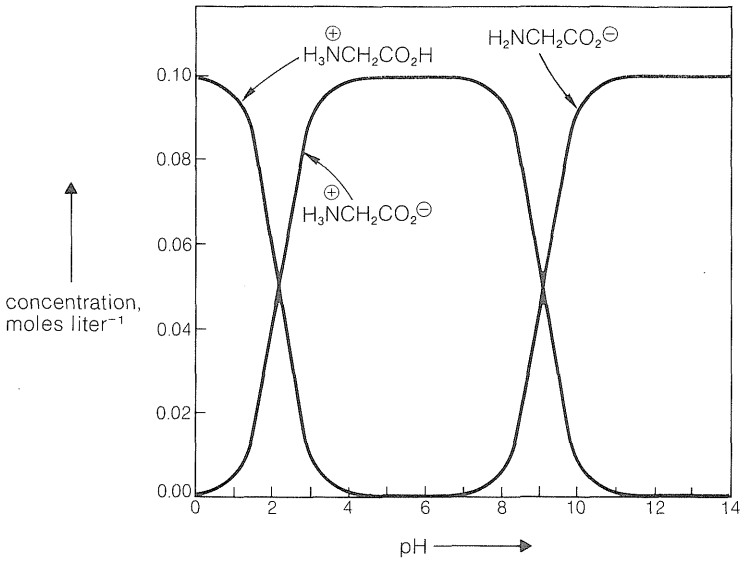


Figure 25-1 Concentrations of $H_3\overset{\oplus}{N}CH_2CO_2H$, $H_3\overset{\oplus}{N}CH_2CO_2^{\ominus}$, and $H_2NCH_2CO_2^{\ominus}$ as a function of pH for a 0.1*M* solution of glycine in water

$$pK_a = pH + \log_{10} \frac{[H_3\overset{\oplus}{N}CH_2CO_2H]}{[H_3\overset{\oplus}{N}CH_2CO_2^{\ominus}]} = 2.34 \qquad (25\text{-}1)$$

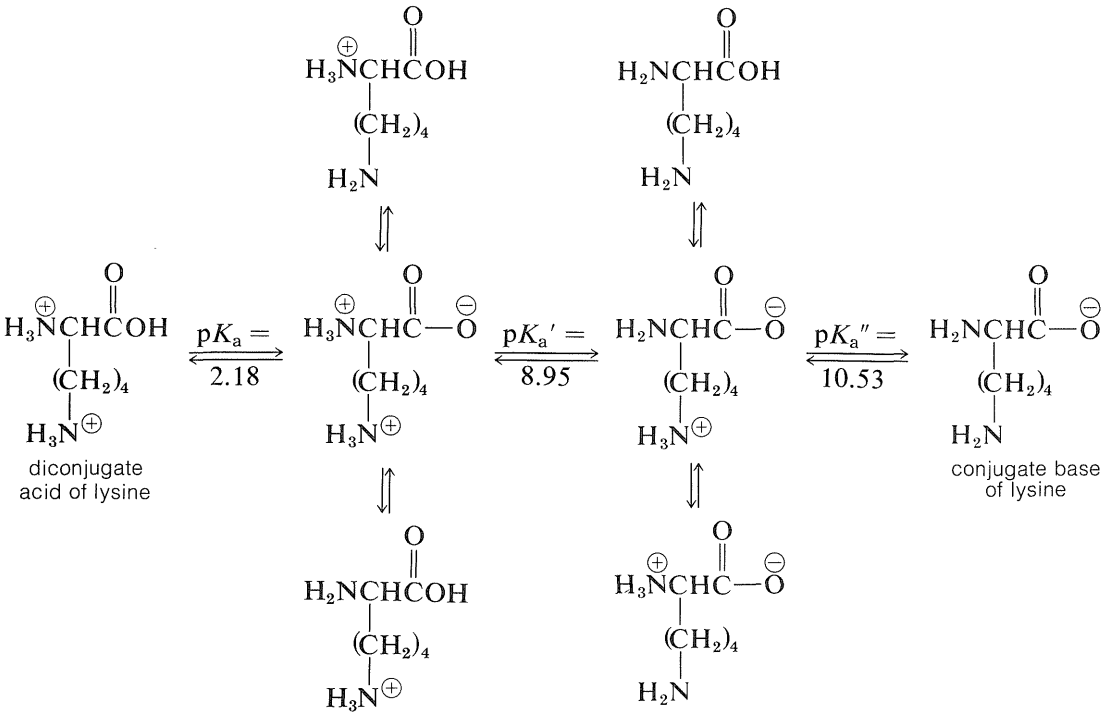
$$pK_a' = pH + \log_{10} \frac{[H_3\overset{\oplus}{N}CH_2CO_2^{\ominus}]}{[H_2NCH_2CO_2^{\ominus}]} = 9.60 \qquad (25\text{-}2)$$

Exercise 25-3 How would the general features of the plot of concentration of dipolar ion and charged species versus pH for glycine (Figure 25-1) change for 6-aminohexanoic acid, which has pK_a values of 4.43 and 10.75? Give special attention to the position of the isoelectric point and the width of the pH range over which the dipolar ion is expected to be the most stable species present.

Exercise 25-4 Use Equations 25-1 and 25-2 to show that the isoelectric point of glycine is the average of the two pK_a values for the acid dissociation of glycine.

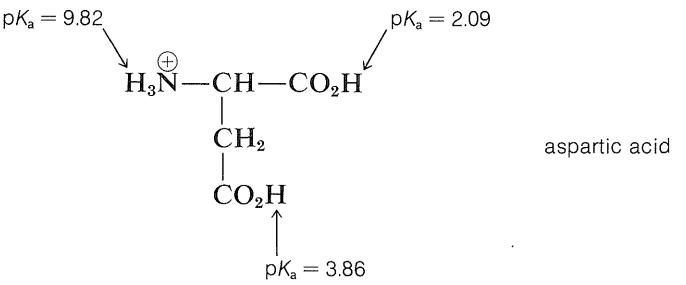
The pH behavior of amino acids with either acidic or basic functional groups attached to the side chains is more complicated than of simple amino

acids. For example, there are three acid dissociations starting with the diconjugate acid of lysine:

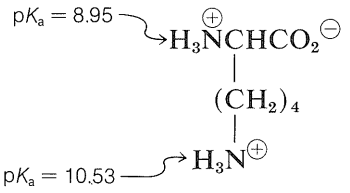


The pK_a values for the side-chain functions of acidic and basic amino acids are given in Table 25-1.

We already have mentioned how the $H_3\overset{\oplus}{N}$— group of the conjugate acid of glycine enhances the acid strength of the carboxyl group compared to ethanoic acid and how the —$\overset{\ominus}{C}O_2$ group reduces the acidity of the $H_3\overset{\oplus}{N}$— group of the dipolar ion relative to ethanammonium ion. These effects will be smaller the farther away the charged group is from the ionizable group. As a result, one would predict that the carboxyl groups of aspartic acid would have different pK_a values, and indeed this is so:



Similarly, the side-chain ammonium group of lysine is less acidic than that of the ammonium group close to the carboxyl group:



conjugate acid of lysine

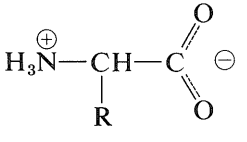
Exercise 25-5 **a.** The equations for the acid-base equilibria of lysine on p. 1214 show possible involvement of three forms of the monocation and three forms of the neutral acid. Arrange the three forms of each set in expected order of stability. Give your reasoning.

b. The conjugate acid of glutamic acid (Table 25-1) has three acid dissociation steps with pK_a values of 2.19, 4.25 and 9.67. Write equations for the equilibria involved and assign pK_a values to each. Do the same for arginine (Table 25-1) with pK_a values of 2.17, 9.04 and 12.48. Calculate the isoelectric point for glutamic acid and for arginine.

25-3 PHYSICAL AND SPECTROSCOPIC PROPERTIES

The α-amino acids crystallize as the dipolar forms, $H_3\overset{\oplus}{N}—CHR—\overset{\ominus}{CO_2}$, and the strong intermolecular electrical forces in the crystals lead to higher melting points than those of simple amines or monocarboxylic acids (see Table 25-1). The melting points are so high that decomposition often occurs on melting. The solubility characteristics of amino acids in water are complex because of the acid-dissociation equilibria involved, but they are least soluble at their isoelectric points. The dipolar structures of amino acids greatly reduce their solubility in nonpolar organic solvents compared to simple amines and carboxylic acids.

The infrared spectra of α-amino acids in the solid state or in solution do not show a carbonyl absorption band at 1720 cm^{-1} characteristic of a carboxyl group. Rather, they show a strong absorption near 1600 cm^{-1} typical of the carboxylate anion. The N—H stretch appears as a strong, broad band between 3100–2600 cm^{-1}:



$\nu_{CO_2^\ominus}$, 1600 cm^{-1}
$\nu_{\oplus N—H}$, 3100–2600 cm^{-1}

Exercise 25-6 Indicate the approximate positions of C=O and N—H absorptions you would expect in the infrared spectra of (a) $ClH_3\overset{\ominus}{N}\overset{\oplus}{C}H_2CO_2H$ (b) $H_2\overset{\ominus}{N}\overset{\oplus}{C}H_2CO_2Na$.

Exercise 25-7 Sketch the nmr spectrum showing the splitting pattern and chemical shifts you would anticipate for alanine dissolved in an excess of D_2O. Do not neglect H–D exchange (Section 9-10E and 9-10I).

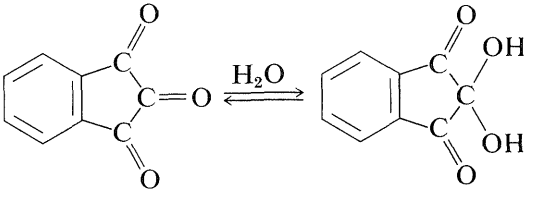
Amino acids do not give any very useful ultraviolet absorption spectra unless they possess aromatic groups as in phenylalanine, tryptophan, and tyrosine. The absorption characteristics of these groups are more useful in monitoring chemical and conformational changes in proteins than they are in the simple amino acids.

It is not easy to obtain the mass spectra of amino acids because of their low volatility. However, a number of special techniques now make possible determination of the mass spectra of amino acids and also of peptides. Because very small amounts of sample are required, this is becoming a particularly useful method of amino acid and peptide analysis.

25-4 ANALYSIS OF AMINO ACIDS

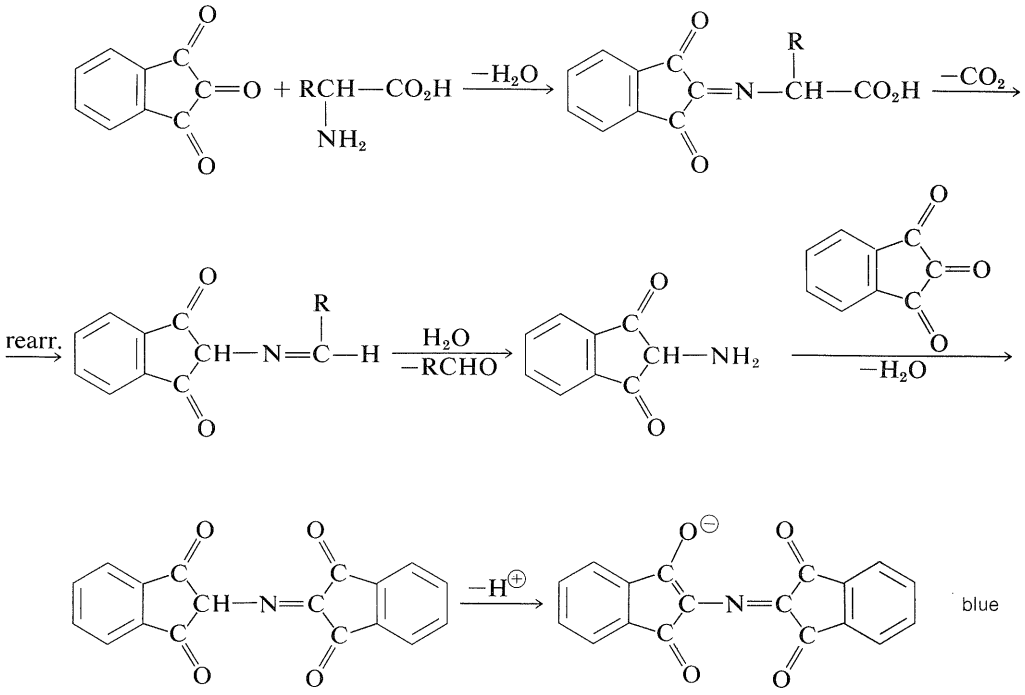
25-4A The Ninhydrin and Related Tests

In many kinds of research it is important to have simple and sensitive means for analysis of amino acids, particularly in small quantities. Detection of amino acids can be achieved readily by the "ninhydrin color test," whereby an alcoholic solution of the triketone, "ninhydrin," is heated with an amino acid and produces an intense blue-violet color. The sensitivity and reliability of this test is such that 0.1 micromole of amino acid gives a color intensity reproducible to a few per cent, provided that a reducing agent such as stannous chloride is present to prevent oxidation of the colored salt by dissolved oxygen.

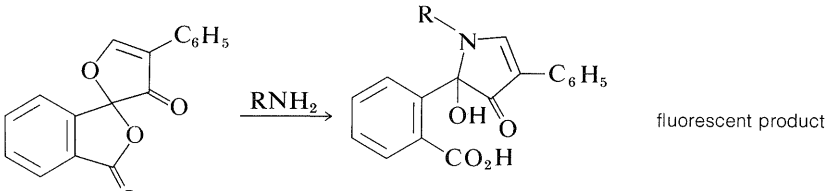


indane-1,2,3-trione "ninhydrin"

The color-forming reaction is interesting because most α-amino acids give the same color irrespective of their structure.[4] The sequence of steps that leads to the color is as follows:



A new, very sensitive method of detection and analysis of amino acids, which is useful down to the 10^{-12} mole (picomole) level, depends on the formation from RNH_2 and "fluorescamine," **1**, of substances that are intensely fluorescent in ultraviolet light:



1

[4]Proline and hydroxyproline are exceptions because neither has the necessary primary NH_2 group needed for the reaction. However, these compounds do react with ninhydrin to give yellow compounds, and these colors can be used to identify them satisfactorily.

Exercise 25-8 The reactions that lead to the blue color produced between ninhydrin and α-amino acids are examples of reactions discussed previously in the context of carbonyl chemistry (see, for instance, Section 16-4C). Write mechanisms, based insofar as possible on analogy, for each of the steps involved in the ninhydrin test, using glycine as an example. Would you expect ammonia or methanamine to give the blue color? Explain.

25-4B Paper Chromatography

Ninhydrin (or fluorescamine) is very useful in chromatographic methods for the analysis of amino acids. One of these is paper chromatography, wherein amino acids are separated as the consequence of differences in their partition coefficients between water and an organic solvent. The aqueous phase is held stationary in the pores of the paper because of strong interaction of the water with the hydroxyl functions of the cellulose. The differences in partition coefficients show up as differences in rates of migration on the surface of moist (but not wet) paper over which there is a slow flow of a water-saturated organic solvent. We shall discuss one of several useful modes of operation. In this example, a drop of the solution to be analyzed is placed on the corner of a sheet of moist paper (often filter paper), which is then placed in an apparatus like that of Figure 25-2, arranged so that the organic solvent can migrate upward by capillarity across the paper, carrying the amino acids with it along one edge. The acids that have the greatest solubility in the organic solvent move most rapidly and when the solvent reaches the top of the paper, the paper is removed, dried, then turned sidewise, and a different solvent allowed to migrate upward. This double migration process gives a better separation of the amino acids than a single migration and results in concentration of the different amino acids in

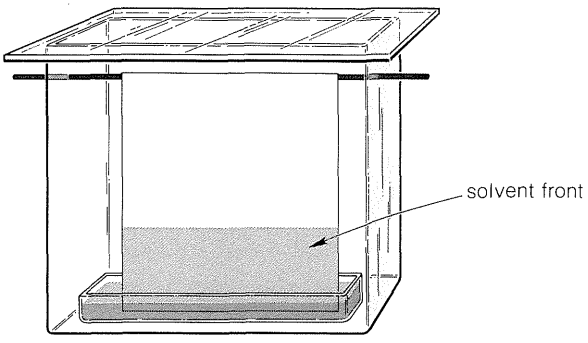


Figure 25-2 Diagram of apparatus used to develop a paper chromatogram, Paper is suspended from its top edge within an airtight container, here a glass box closed with a glass plate, having an atmosphere saturated with solvent vapor; the lower edge of paper dips into a trough containing the liquid solvent.

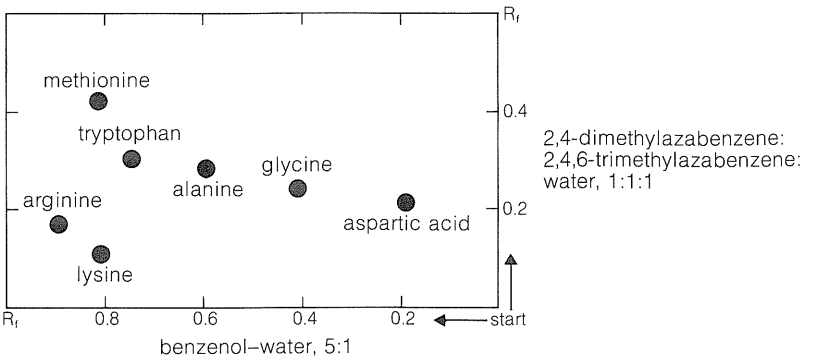
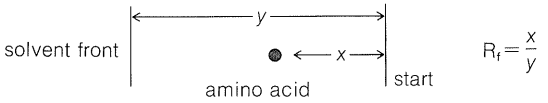


Figure 25-3 Idealized two-dimensional paper chromatogram of a mixture of amino acids. The horizontal and vertical scales represent the distance of travel of a component of the mixture in a given solvent relative to that of the solvent itself. This is known as the R_f value and is fairly constant for a particular compound in a given solvent. A rough identification of the amino acids present in the mixture may therefore be made on the basis of their R_f values.



rather well-defined spots. These spots can be made visible by first drying and then spraying the paper with ninhydrin solution. The final result is as shown in Figure 25-3 and usually is quite reproducible under a given set of conditions. The identities of the amino acids that produce the various spots are established by comparison with the behavior of known mixtures.

Analysis by thin-layer chromatography (see Section 9-2B) can be carried out in the same way as paper chromatography. The partitioning is now between a solid stationary phase (the coating on the plate) and the moving solvent front.

25-4C Ion-Exchange Chromatography

The advent of ion-exchange chromatography has revolutionized the separation and analysis of amino acids as well as that of many inorganic substances. As the name implies, it involves the exchange of ions between a stationary and a moving phase. The stationary phase is an insoluble polymer (or resin) having chains on which are located ionic functions such as sulfonate groups $-SO_3^{\ominus}$ or quaternary ammonium groups, $-\overset{\oplus}{N}R_3$. The counterions to these groups, such as $Na^{\oplus}$ or $Cl^{\ominus}$, are not bound to the resin and can be exchanged for other ions in the mobile phase as the mobile phase travels through the resin. A common application of this principle is in household water softeners, in which the calcium and magnesium ions in ordinary "hard" water are replaced by sodium

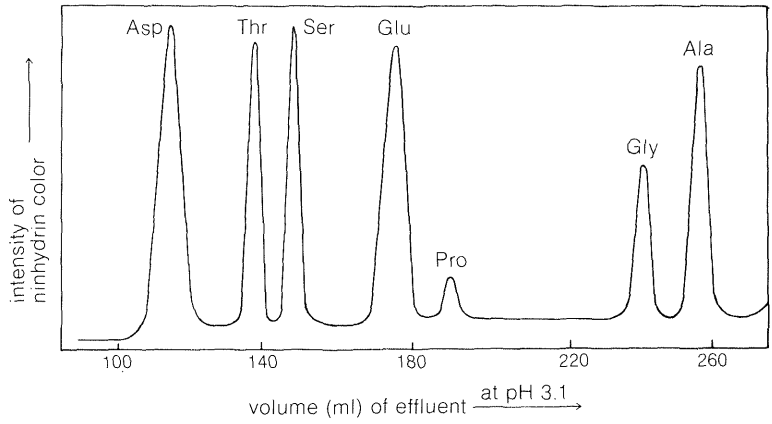


Figure 25-4 Part of amino-acid chromatogram obtained by the method of automatic amino-acid analysis from a hydrolyzed sample of the enzyme ribonuclease. The component amino acids listed are present in the ratio Asp:Thr:Ser:Glu:Pro:Gly:Ala = 15:10:15:12:4:3:12, as determined by peak intensity. The volume of effluent is a measure of the retention time of the amino acids on the column.

ions from the resin (Equation 25-3). The resulting "soft" water can be freed of metal ions, if desired, by exchanging the $Na^{\oplus}$ ions for protons (Equation 25-4):

$$2(\text{resin}{-}SO_3^{\ominus}Na^{\oplus}) + Ca^{2\oplus} \rightleftharpoons 2(\text{resin}{-}SO_3^{\ominus})Ca^{2\oplus} + 2Na^{\oplus} \qquad (25\text{-}3)$$

$$\text{resin}{-}SO_3^{\ominus}H^{\oplus} + Na^{\oplus} \rightleftharpoons \text{resin}{-}SO_3^{\ominus}Na^{\oplus} + H^{\oplus} \qquad (25\text{-}4)$$

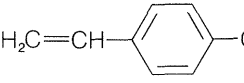
In strongly acidic solutions (pH ~0), the amine and carboxyl groups of an amino acid are completely protonated. This cationic form of the amino acid can be exchanged with the cations associated with the sulfonate groups of the resin:

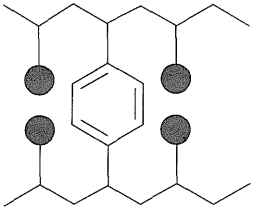
$$\text{resin}{-}SO_3^{\ominus}Na^{\oplus} + H_3\overset{\oplus}{N}CRHCO_2H \rightleftharpoons \text{resin}{-}SO_3^{\ominus}H_3\overset{\oplus}{N}CRHCO_2H + Na^{\oplus} \qquad (25\text{-}5)$$

The process is reversible, and the amino acid cations can in turn be exchanged off the columns. However, different amino acids have different affinities for the resin, and these are considerably influenced by the pH of the moving phase (eluent). The basic amino acids (arginine, lysine), which form cations most readily, are more strongly held by cation-exchange resins than are acidic amino acids (asparatic and glutamic acids). There is a spectrum of affinities of the other amino acid cations for the resin between these extremes. Thus a mixture of amino acids can be separated by ion-exchange chromatography by elution with

buffered aqueous solutions. The effluent from the column is mixed with ninhydrin solution and the intensity of the blue color is measured and plotted as a function of time at constant flow rates (Figure 25-4). The identity of an amino acid is determined by the volume of solvent required to elute the amino acid from the column, and the concentration is determined from the intensity of the color developed.

Exercise 25-9 Explain why arginine elutes from an ion-exchange column using a buffer at pH 5–6, whereas glutamic acid elutes at pH 3.

Exercise 25-10 A cation-exchange resin can be prepared by radical-addition polymerization of phenylethene (styrene, Section 10-8) in the presence of about 2–10% 1,4-diethenylbenzene (1,4-divinylbenzene),  , followed by electrophilic sulfonation of the resulting polymer with H_2SO_4–SO_3 (see Section 22-4G). Explain how these reactions lead to a *three-dimensional* insoluble polymer with linkages as shown below. Indicate the reaction mechanisms involved.



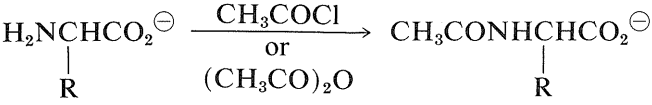
Exercise 25-11 Consider a "hard" water comprised of dilute $MgCl_2$. Ion exchange with resin$-SO_3^{\ominus}Na^{\oplus}$ replaces $Mg^{2\oplus}$ with $Na^{\oplus}$, and with resin$-SO_3H$, $Na^{\oplus}$ is replaced by $H^{\oplus}$, thereby producing a dilute HCl solution. What kind of an ion-exchange resin would you need to remove the $Cl^{\ominus}$ from the HCl solution and produce "deionized" water? (Consider exchanging $Cl^{\ominus}$ for $^{\ominus}OH$.)

25-5 REACTIONS OF AMINO ACIDS

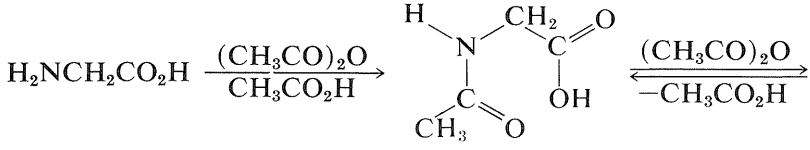
25-5A Ester and Amide Formation

To some degree the reactions of amino acids are typical of isolated carboxylic acid and amine functions. Thus the carboxyl function can be esterified with an excess of an alcohol under acidic conditions, and the amine function can be

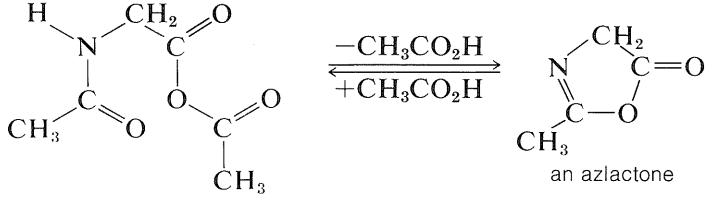
acylated with acid chlorides or anhydrides under basic conditions:



The products, however, are not indefinitely stable because the functional groups can, and eventually will, react with each other. For example, in the acylation of glycine with ethanoic anhydride, the first-formed product may cyclize to the "azlactone" if the reaction is prolonged or excess anhydride is used:

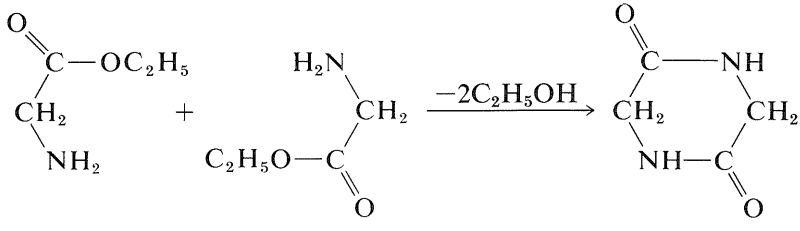


N-ethanoylamino-ethanoic acid (*N*-acetylglycine)



an azlactone

Esters of amino acids also cyclize, but they do so intermolecularly to give "diketopiperazines." These compounds are cyclic amides:

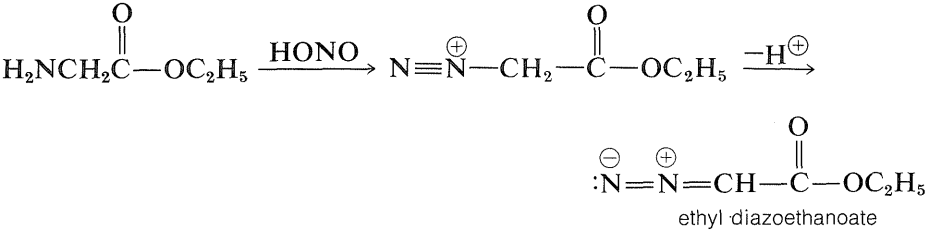


2,5-diazacyclohexane-1,4-dione (diketopiperazine)

Exercise 25-12 a. Draw the structure of the azlactone derived from L-phenylalanine and ethanoic anhydride.
b. Which of the hydrogens in this azlactone would you expect to be the most acidic? Explain.
c. Why do chiral azlactones derived from amino acids such as L-phenylalanine racemize easily on heating in ethanoic acid in the presence of ethanoate ion?

25-5B Nitrous Acid Reaction

The amine function of α-amino acids and esters reacts with nitrous acid in a manner similar to that described for primary amines (Section 23-10A). The diazonium ion intermediate loses molecular nitrogen in the case of the acid, but the diazonium ester loses a proton and forms a relatively stable diazo compound known as ethyl diazoethanoate:



ethyl diazoethanoate

This diazo ester is formed because loss of N_2 from the diazonium ion results in formation of a quite unfavorable carbocation.

Exercise 25-13 Explain why glycine itself, as the dipolar ion, reacts with nitrous acid to eliminate nitrogen, whereas the ethyl ester of glycine forms ethyl diazoethanoate.

25-5C Amino Acids with Aldehydes

α-Amino acids react with aldehydes to form decarboxylation and/or deamination products. The reaction sequence is shown in Figure 25-5 and closely resembles the ninhydrin reaction (Section 25-4A). In the first step the amine condenses with the aldehyde to give an imine or Schiff base, **2**. What happens next depends on the relative rates of proton shift and decarboxylation of **2**.

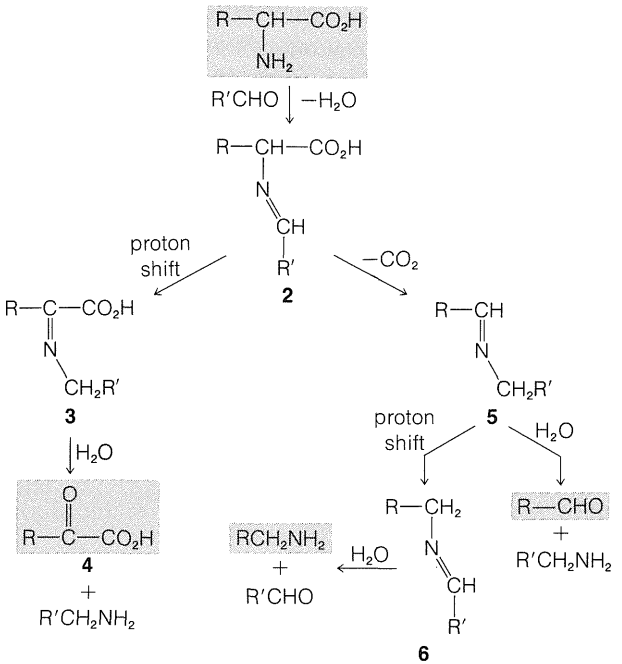


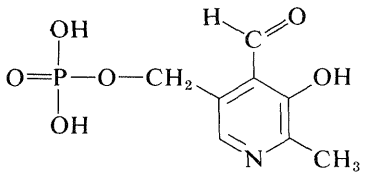
Figure 25-5 Reactions of α-amino acids with an aldehyde, R'CHO. The products are the result of decarboxylation and/or deamination; the fraction of the products formed by each route is determined by the ratio of the rate of proton shift to the rate of decarboxylation of **2**.

Proton shift produces a rearranged imine, **3**, which can hydrolyze to the keto acid **4**. The keto acid is a deamination product. Alternatively, decarboxylation can occur (see Section 18-4) and the resulting imine, **5**, can either hydrolyze or rearrange by a proton shift to a new imine, **6**. Hydrolysis of **5** or **6** gives an aldehyde and an amine.

There is an important biochemical counterpart of the deamination reaction that utilizes pyridoxal phosphate, **7**, as the aldehyde. Each step in the sequence is catalyzed by a specific enzyme. The α-amino group of the amino acid combines with **7** and is converted to a keto acid. The resulting pyridoxamine then reacts to form an imine with a different α-keto acid, resulting in formation of a new α-amino acid and regenerating **7**. The overall process is shown in Equation 25-6 and is called **transamination.** It is a key part of the process whereby amino acids are metabolized.

$$\underset{\underset{NH_2}{|}}{RCHCO_2H} + \underset{\underset{O}{\|}}{R'CCO_2H} \underset{\text{enzymes}}{\overset{\text{pyridoxal phosphate}}{\rightleftharpoons}} \underset{\underset{O}{\|}}{RCCO_2H} + \underset{\underset{NH_2}{|}}{R'CHCO_2H} \qquad (25\text{-}6)$$

The biochemical process occurs with complete preservation of the L configuration at the α carbon. The same reactions can be carried out nonenzymatically using pyridoxal phosphate, but they are nonstereospecific, require metal ions as a catalyst, and give mixtures of products.



pyridoxal phosphate, **7**

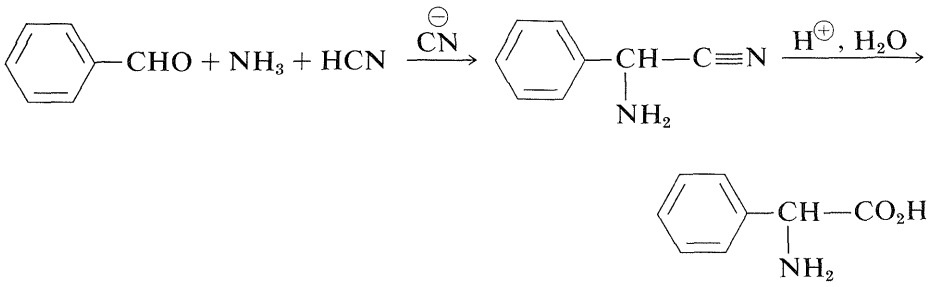
25-6 SYNTHESIS OF α-AMINO ACIDS

Many of the types of reactions that are useful for the preparation of amino acids have been discussed previously in connection with separate syntheses of carboxylic acids (Chapter 18) and amino compounds (Chapter 23). Examples include the S_N2 displacement of halogen from α-halo acids by ammonia,

$$ClCH_2CO_2H \xrightarrow[50°]{NH_3\ (excess)} H_2NCH_2C\overset{\ominus}{O_2}\overset{\oplus}{N}H_4 \xrightarrow{H^{\oplus}} H_2NCH_2CO_2H$$

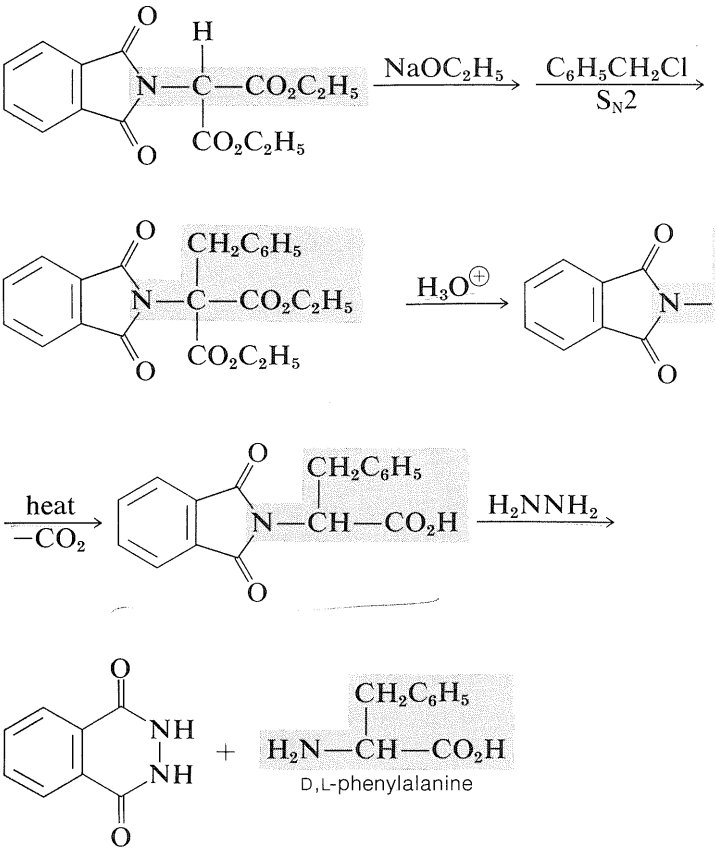
glycine

and the **Strecker synthesis,** which, in its first step, bears a close relationship to cyanohydrin formation (Section 16-4A):

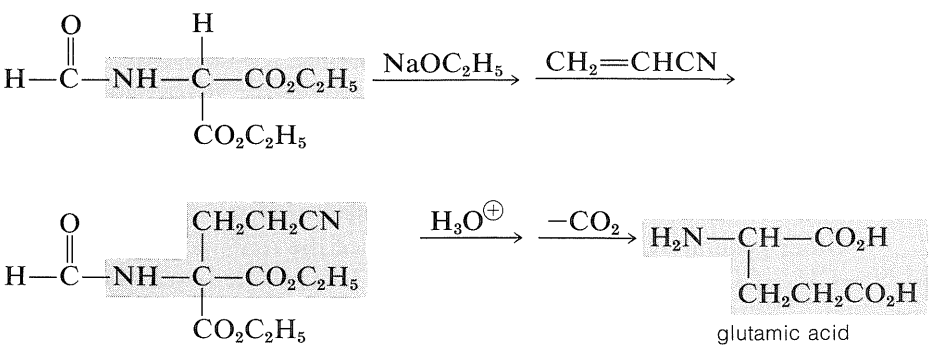


Other general synthetic methods introduce the α-amino acid grouping, $H_2N-\underset{|}{CH}-CO_2H$, by way of enolate anions. Two selected examples follow. Notice that in each a carbanion is generated and alkylated. Also the H_2N- group is introduced as a protected amide or imide group.

1. **phthalimidomalonic ester synthesis**



2. ***N*-formylaminomalonic ester synthesis**



The key step is the base-catalyzed addition of $CH_2{=}CHCN$, which is a Michael addition (Section 18-9D).

With those amino acids that are very soluble in water, it usually is necessary to isolate the product either by evaporation of an aqueous solution or by precipitation induced by addition of an organic solvent like alcohol. Difficulty may be encountered in obtaining a pure product when inorganic salts are coproducts of the synthesis. The best general method for removal of inorganic salts involves passage of the solutions through columns of suitable ion-exchange resins (Section 25-4C).

The products of laboratory syntheses, starting with achiral reagents, are of course racemic α-amino acids. To obtain the natural amino acids, the D,L mixtures must be resolved (Section 19-3).

Exercise 25-14 Show how the following amino acids may be prepared from the indicated method and starting materials:

a. glutamic acid from 2-oxopentanedioic acid (α-ketoglutaric acid) by the Strecker method

b. leucine from 2-methyl-1-propanol by the phthalimidomalonic ester synthesis

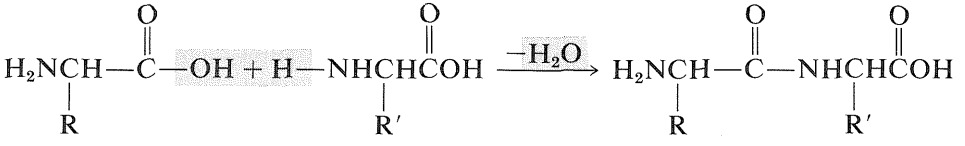
c. aspartic acid from ethyl chloroethanoate by the *N*-formylaminomalonic ester synthesis

Exercise 25-15 Suggest a synthetic route to proline from hexanedioic acid (adipic acid) that involves the transformations $-CO_2H \longrightarrow -NH_2$, and $-CH_2CO_2H$ to $-CHBrCO_2H$. Specify the reagents required to accomplish each step.

25-7 PEPTIDES AND PROTEINS

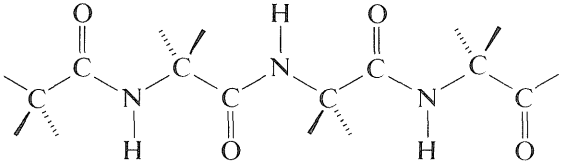
25-7A Classification

Amino acids are the building blocks of the polyamide structures of peptides and proteins. Each amino acid is linked to another by an amide (or peptide) bond formed between the NH_2 group of one and the CO_2H group of the other:



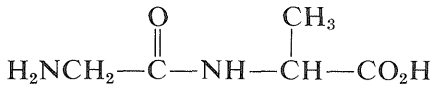
In this manner a polymeric structure of repeating amide links is built into a chain or ring. The amide groups are planar and configuration about the C–N bond is usually, but not always, trans (Section 24-1). The pattern of *covalent*

bonds in a peptide or protein is called its **primary structure:**

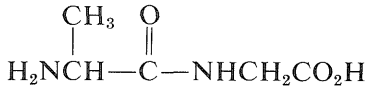


The distinction between a protein and a peptide is not completely clear. One arbitrary choice is to call proteins only those substances with molecular weights greater than 5000. The distinction might also be made in terms of differences in physical properties, particularly hydration and conformation. Thus proteins, in contrast to peptides, have very long chains that are coiled and folded in particular ways, with water molecules filling the voids in the coils and folds. Hydrogen bonding between the amide groups plays a decisive role in holding the chains in juxtaposition to one another, in what is sometimes called the **secondary** and **tertiary structure.**[5] Under the influence of heat, organic solvents, salts, and so on, protein molecules undergo changes, often irreversibly, called **denaturation.** The conformations of the chains and the degree of hydration are thereby altered, with the result that solubility and ability to crystallize decreases. Most importantly, the physiological properties of the protein usually are destroyed permanently on denaturation. Therefore, if a synthesis of a protein is planned, it would be necessary to duplicate not only the amino-acid sequences but also the exact conformations of the chains and the manner of hydration characteristic of the native protein. With peptides, the chemical and physiological properties of natural and synthetic materials usually are identical, provided the synthesis duplicates all of the structural and configurational elements. What this means is that a peptide automatically assumes the secondary and tertiary structure characteristic of the native peptide on crystallization or dissolution in solvents.

Representation of peptide structures of any length with conventional structural formulas is cumbersome. As a result, abbreviations are universally used that employ three-letter symbols for the component amino acids. It is important that you know the conventions for these abbreviations. The two possible dipeptides made up of one glycine and one alanine are



glycylalanine (Gly–Ala)



alanylglycine (Ala–Gly)

Notice that in the conventions used for names and abbreviated formulas the amino acid with the free amino group (*the* N-*terminal amino acid*) always is written on the *left*. The amino acid with the free carboxyl group (*the* C-*termi-*

[5]The distinction between secondary and tertiary structure is not sharp. Secondary structure involves consideration of the interactions and spatial relationships of the amino acids in the peptide chains that are *close* together in the primary structure, whereas tertiary structure is concerned with those that are *far apart* in the primary structure.

nal amino acid) always is written on the *right*. The dash between the three-letter abbreviations for the acids designates that they are linked together by an *amide* bond.

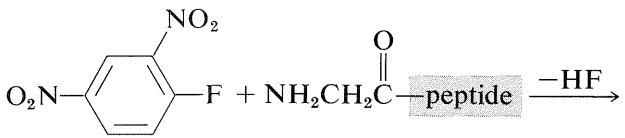
Exercise 25-16 The structure of the hormonal peptide oxytocin is abbreviated to Cys–Tyr–Ile–Gln–Asn–Cys–Pro–Leu–GlyNH$_2$. Draw its full covalent structure.

25-7B Determination of Amino-Acid Sequences

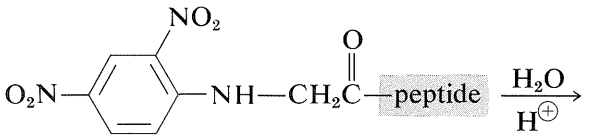
The general procedure for determining the primary structure of a peptide or protein consists of three main steps. First, the number and kind of amino-acid units in the primary structure must be determined. Second, the amino acids at the ends of the chains are identified, and third, the sequence of the component amino acids in the chains is determined.

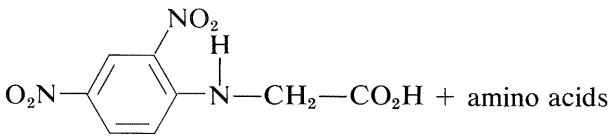
The amino-acid composition usually is obtained by complete acid hydrolysis of the peptide into its component amino acids and analysis of the mixture by ion-exchange chromatography (Section 25-4C). This procedure is complicated by the fact that tryptophan is destroyed under acidic conditions. Also, asparagine and glutamine are converted to aspartic and glutamic acids, respectively.

Determination of the *N*-terminal acid in the peptide can be made by treatment of the peptide with 2,4-dinitrofluorobenzene, a substance very reactive in nucleophilic displacements with amines but not amides (see Section 14-6B). The product is an *N*-2,4-dinitrophenyl derivative of the peptide which, after hydrolysis of the amide linkages, produces an *N*-2,4-dinitrophenyl-amino acid:



2,4-dinitrofluorobenzene

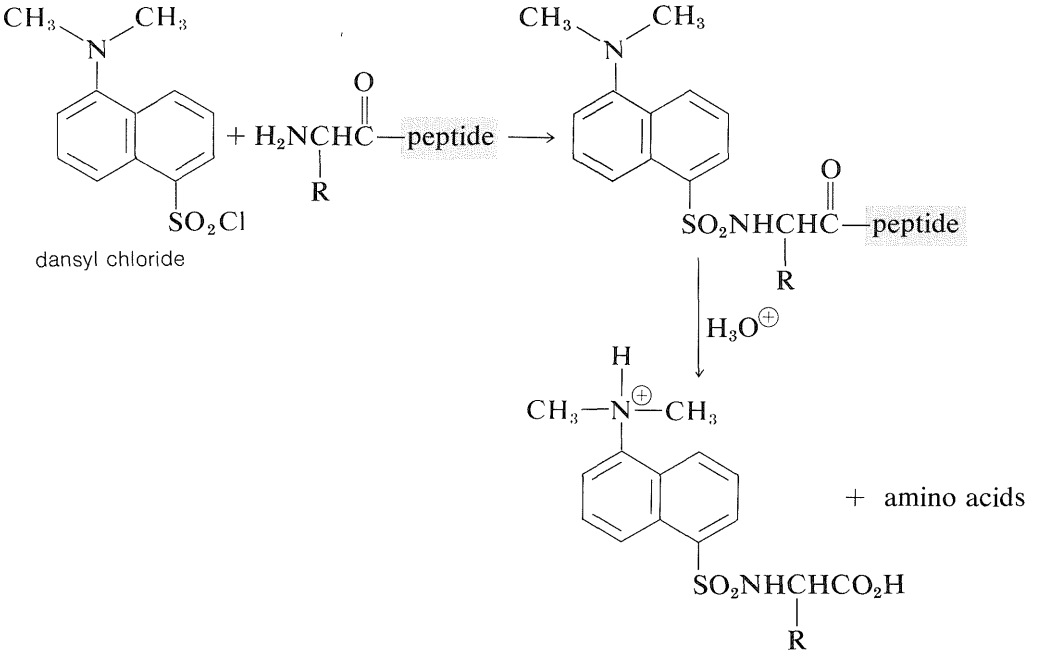




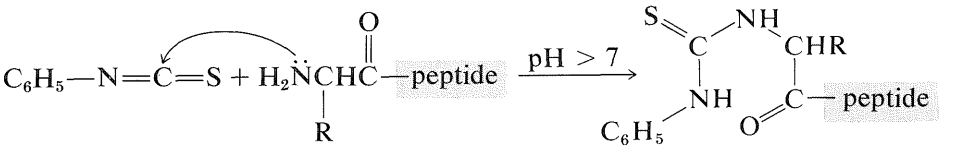
N-2,4-dinitrophenylglycine
(low solubility in acid solutions)

These amino-acid derivatives can be separated from the ordinary amino acids resulting from hydrolysis of the peptide because the low basicity of the 2,4-dinitrophenyl-substituted nitrogen (Section 23-7C) greatly reduces the solubility of the compound in acid solution and alters its chromatographic behavior. The main disadvantage to the method is that the entire peptide must be destroyed in order to identify the one *N*-terminal acid.

A related and more sensitive method makes a sulfonamide of the terminal NH_2 group with a reagent called "dansyl chloride." As with 2,4-dinitrofluorobenzene, the peptide must be destroyed by hydrolysis to release the *N*-sulfonated amino acid, which can be identified spectroscopically in microgram amounts:



A powerful method of sequencing a peptide from the *N*-terminal end is the **Edman degradation** in which phenyl isothiocyanate, $C_6H_5N{=}C{=}S$, reacts selectively with the terminal amino acid under mildly basic conditions. If the reaction mixture is then acidified, the terminal amino acid is cleaved from the peptide as a cyclic **thiohydantoin, 8**:



$$\xrightarrow{pH < 7} \text{peptide}' + \underset{\text{a thiohydantoin } \mathbf{8}}{\text{S=C(NH)(CHR)N(C}_6\text{H}_5\text{)C=O}} \longleftarrow \text{elements of } N\text{-terminal amino acid}$$

The advantage of the Edman procedure is that the residual peptide after *one* degradation now has a new *N*-terminal amino acid that can react further with phenyl isothiocyanate. In practice, it is possible to carry out sequential Edman degradations by an automated procedure that identifies each of the amino acids in sequence from the *N*-terminus as a thiohydantoin. Figure 25-6 illustrates how the procedure works. If the *N*-terminal nitrogen is not a free amino group, but for example is an ethanoylamide, $CH_3CONH—$, the Edman degradation does not proceed.

There are simple reagents that react selectively with the carboxyl terminus of a peptide, but they have not proved as generally useful for analysis of the *C*-terminal amino acids as has the enzyme *carboxypeptidase A*. This enzyme catalyzes the hydrolysis of the peptide bond connecting the amino acid with the terminal carboxyl groups to the rest of the peptide. Thus the amino acids at the carboxyl end will be removed one by one through the action of the enzyme. Provided that appropriate corrections are made for different rates of hydrolysis of peptide bonds for different amino acids at the carboxyl end of the peptide, the sequence of up to five or six amino acids in the peptide can be deduced from the order of their release by carboxypeptidase. Thus a sequence such as peptide–Ser–Leu–Tyr could be established by observing that carboxypeptidase releases amino acids from the peptide in the order Tyr, Leu, Ser:

$$\text{peptide–Ser–Leu–Tyr} \xrightarrow[H_2O]{\text{carboxypeptidase}} \text{peptide–Ser–Leu} + \text{Tyr}$$

$$\longrightarrow \text{peptide–Ser} + \text{Leu} + \text{Tyr} \longrightarrow \text{peptide} + \text{Ser} + \text{Leu} + \text{Tyr}$$

Determining the amino-acid sequences of large peptides and proteins is very difficult. Although the Edman degradation and even carboxypeptidase can be used to completely sequence small peptides, they cannot be applied successfully to peptide chains with several hundred amino acid units. Success has been obtained with long peptide chains by employing reagents, often enzymes, to selectively cleave certain peptide bonds. In this way the chain can be broken down into several smaller peptides that can be separated and sequenced. The problem then is to determine the sequence of these small peptides in the original structure. To do this, alternative procedures for selective cleavages are carried out that produce different sets of smaller peptides. It is not usually necessary to sequence completely all of the peptide sets. The overall amino-acid composition and the respective end groups of each peptide may suffice to

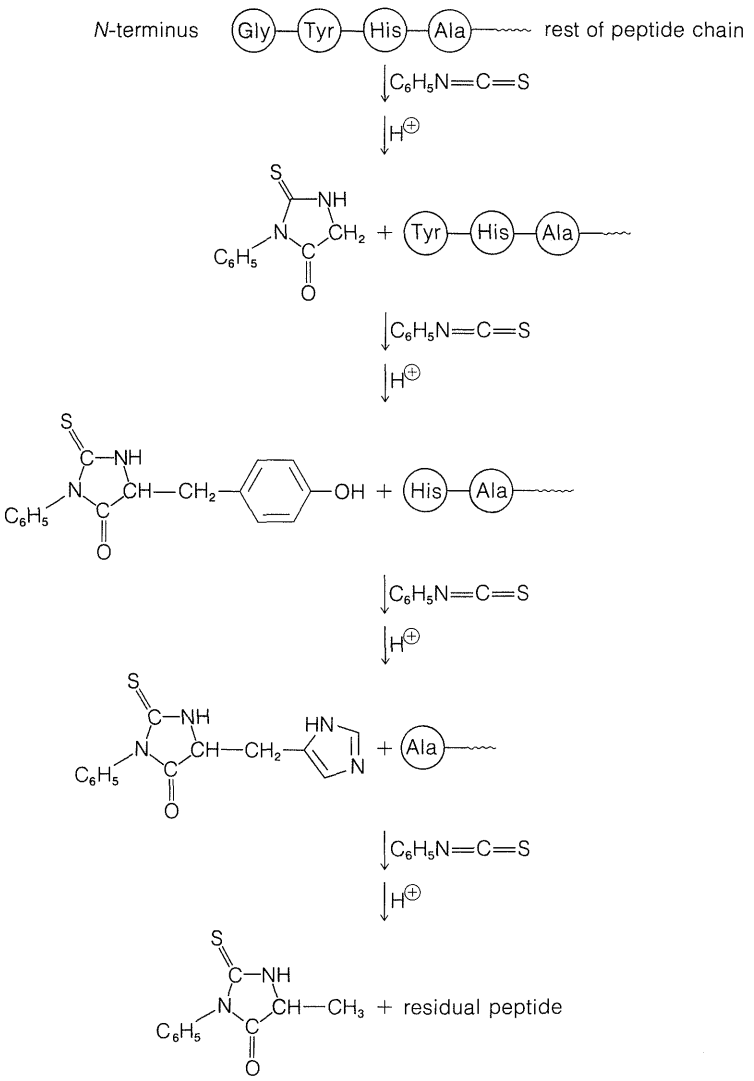
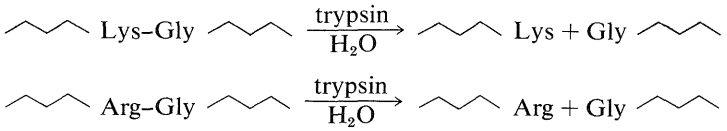


Figure 25-6 Result of a series of Edman degradations on an *N*-terminal Gly–Tyr–His–Ala–peptide

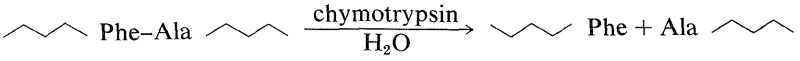
show overlapping sequences from which the complete amino-acid sequence logically can be deduced.

The best way to show you how the overlap method of peptide sequencing works is by a specific example. In this example, we will illustrate the use of the two most commonly used enzymes for selective peptide cleavage. One is *trypsin,* a proteolytic enzyme of the pancreas (MW 24,000) that selectively catalyzes the hydrolysis of the peptide bonds of *basic* amino acids, lysine and

arginine. Cleavage occurs on the *carboxyl side* of lysine or arginine:



Chymotrypsin is a proteolytic enzyme of the pancreas (MW 24,500) that catalyzes the hydrolysis of peptide bonds to the aromatic amino acids, tyrosine, tryptophan, and phenylalanine, more rapidly than to other amino acids. Cleavage occurs on the *carboxyl side* of the aromatic amino acid:



Our example is the sequencing of a peptide (P) derived from partial hydrolysis of a protein which, on complete acid hydrolysis, gave Ala, 3 Gly, Glu, His, 3 Lys, Phe, Tyr, 2 Val, and one molar equivalent of ammonia.

1. Treatment of the peptide (P) with carboxypeptidase released alanine, and with 2,4-dinitrofluorobenzene followed by hydrolysis gave the 2,4-dinitrophenyl derivative of valine. These results establish the *N*-terminus as valine and the *C*-terminus as alanine. The known structural elements now are

P = Val ———————— Ala

2. Partial hydrolysis of the peptide (P) with trypsin gave a hexapeptide, a tetrapeptide, a dipeptide, and one molar equivalent of lysine. The peptides, which we will designate respectively as M, N, and O, were sequenced by Edman degradation and found to have structures:

Gly–Ala	O
Val–Tyr–Glu–Lys	N
Val–Gly–Phe–Gly–His–Lys	M

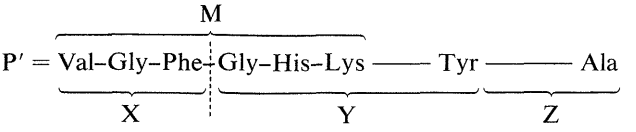
With this information, four possible structures can be written for the original peptide P that are consistent with the known end groups and the fact that trypsin cleaves the peptide P on the *carboxyl* side of the lysine unit. Thus

N–M–Lys–O	M–N–Lys–O
N–Lys–M–O	M–Lys–N–O

3. Partial hydrolysis of the peptide P using chymotrypsin as catalyst gave three peptides, X, Y, and Z. These were not sequenced, but their amino-acid composition was determined:

Gly, Phe, Val	X
Gly, His, Lys, Tyr, Val	Y
Ala, Glu, Gly, 2 Lys	Z

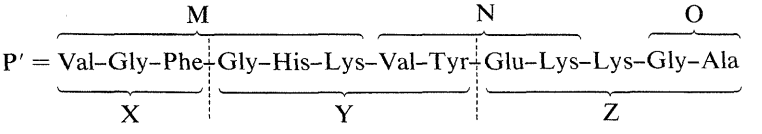
This information can be used to decide which of the alternative structures deduced above is correct. Chymotrypsin cleaves the peptide on the carboxyl side of the phenylalanine and tyrosine units. Only peptide M contains Phe, and if we compare M with the compositions of X, Y, and Z, we see that only X and Y overlap with M. Peptide Z contains the only Ala unit and must be the *C*-terminus. If we put together these pieces to get a peptide, P′ (which differs from P by not having the nitrogen corresponding to the ammonia formed on complete hydrolysis) then P′ must have the structure X–Y–Z:



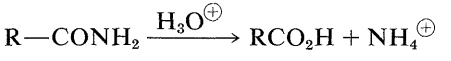
This may not be completely clear, and it will be well to consider the logic in some detail. Peptides M and N both have *N*-terminal valines, and one of them must be the *N*-terminal unit. Peptide M overlaps with X and Y, and because X and Y are produced by a cleavage on the *carboxyl* side of Phe, the X and Y units have to be connected in the order X–Y. Because the other Val is in Y, the *N*-terminus must be M. This narrows the possibilities to

M–N–Lys–O
M–Lys–N–O

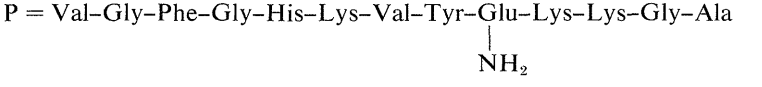
There are two Lys units in Z, and this means that only the sequence M–N–Lys–O is consistent with the sequence X–Y–Z, as shown:



The final piece of the puzzle is the placement of the mole of ammonia released from the original peptide on acid hydrolysis. The ammonia comes from a primary amide function:



The amide group cannot be at the *C*-terminus because the peptide would then be inert to carboxypeptidase. The only other possible place is on the side-chain carboxyl of glutamic acid. The complete structure may be written as



Exercise 25-17 What problems might be encountered in using the 2,4-dinitrofluorobenzene method for determination of end groups on Gly–Lys–Ala? Explain.

Exercise 25-18 The tripeptide, eisenine, has only one free carboxyl group, does not react with 2,4-dinitrofluorobenzene, and on complete hydrolysis yields 2 moles of L-glutamic acid, 1 mole of L-alanine, and 1 mole of ammonia. Alanine is indicated to be the *C*-terminal amino acid. Write a structure for eisenine that is in accord with the above facts.

Exercise 25-19* Eledoisin is a peptide isolated from the salivary glands of *eledone,* a Mediterranean eight-armed cephalopod. The peptide is a powerful hypotensive agent. Deduce a possible structure from the following information: (1) Complete hydrolysis gives equal amounts of ammonia, Ala, Asp, Glu, Gly, Ile, Leu, Lys, Met, Phe, Pro, and Ser. (2) No free amino *N*-terminal group or free carboxyl *C*-terminal group can be detected. (3) Chymotrypsin hydrolysis forms two peptides, L and M. Their compositions are

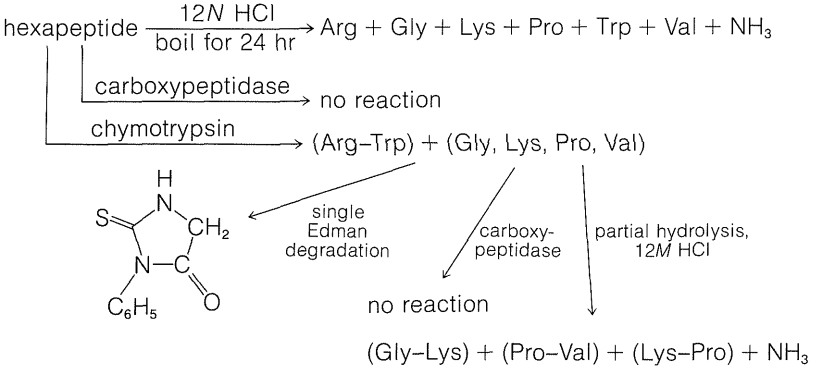
L = Ala, Asp, Glu, Lys, Phe, Pro, Ser (unsequenced)
M = Ile–Gly–Leu–MetNH_2 (sequenced)

(At this point you should be able to deduce the sequence of five amino acids at the *C*-terminus of eledoisin.) (4) Trypsin hydrolysis gives two peptides, P and Q, with the indicated compositions:

P = Glu, Lys, Pro, Ser
Q = Ala, Asp, Gly, Ile, Leu, Met, Phe

(At this point, you can deduce two possible sequences for Q.) (5) Trypsin hydrolysis of L gives a peptide of composition Ala, Asp, Phe which, with 2,4-dinitrofluorobenzene, gives the 2,4-dinitrophenyl derivative of aspartic acid. (6) Partial acid hydrolysis of eledoisin gives several dipeptides, among them Ser–Lys and Pro–Ser.

Exercise 25-20 A hexapeptide was subjected to the transformations diagrammed below. (The commas between the amino acids indicate the sequence is unknown or unspecified.) Deduce the structure of the hexapeptide.

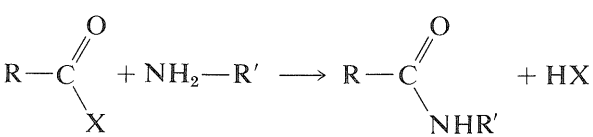


Using procedures such as those outlined in this section more than 100 proteins have been sequenced. This is an impressive accomplishment considering the complexity and size of many of these molecules (see, for example, Table 25-3). It has been little more than two decades since the first amino acid sequence of a protein was reported by F. Sanger, who determined the primary structure of insulin (1953). This work remains a landmark in the history of chemistry because it established for the first time that proteins have definite primary structures in the same way that other organic molecules do. Up until that time, the concept of definite primary structures for proteins was openly questioned. Sanger developed the method of analysis for *N*-terminal amino acids using 2,4-dinitrofluorobenzene and received a Nobel Prize in 1958 for his success in determining the amino-acid sequence of insulin.

25-7C Methods for Forming Peptide Bonds

The problems involved in peptide syntheses are of much practical importance and have received considerable attention. The major difficulty in putting together a chain of say 100 amino acids in a particular order is one of overall yield. At least 100 separate synthetic steps would be required and, if the yield in each step were equal to $n \times 100\%$, the overall yield would be $(n^{100} \times 100\%)$. If the yield in each step were 90%, the overall yield would be only 0.003%. Obviously, a practical laboratory synthesis of a peptide chain must be a highly efficient process. The extraordinary ability of living cells to achieve syntheses of this nature, not of just one but of a wide variety of such substances, is truly impressive.

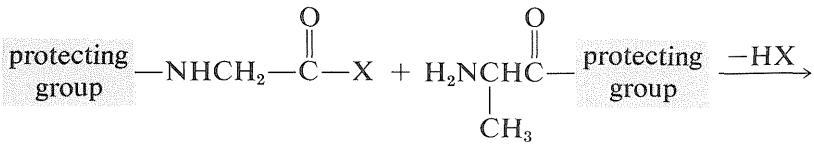
Several methods for the formation of amide bonds have been discussed in Sections 18-7A and 24-3A. The most general reaction is shown below, in which X is some reactive leaving group (see Table 24-1):

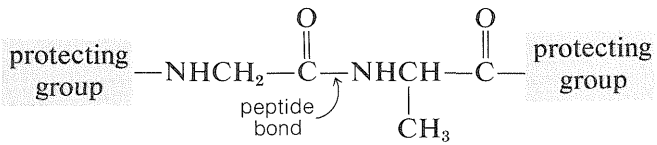


When applied to coupling two different amino acids, difficulty is to be expected because these same reactions can link two amino acids in a total of four different ways. Thus if we started with a mixture of glycine and alanine, we could generate four dipeptides, Gly–Ala, Ala–Gly, Gly–Gly, and Ala–Ala.

To avoid unwanted coupling reactions a protecting group is substituted on the amino function of the acid that is to act as the acylating agent. Furthermore, all of the amino, hydroxyl, and thiol functions that may be acylated to give undesired products usually must be protected. For instance, to synthesize

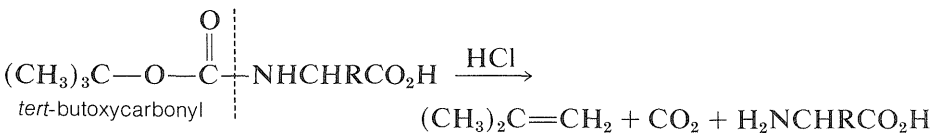
Gly–Ala free of other possible dipeptides, we would have to protect the amino group of glycine and the carboxyl group of alanine:

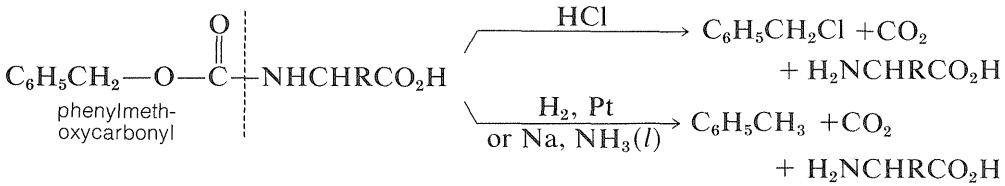




In general, peptide synthesis is a sequence of steps involving (a) protection of functional groups, (b) conversion of the carboxyl group to a more reactive group, (c) coupling, and (d) removal of the protecting group, as shown in Figure 25-7.

Some methods of protecting amine and hydroxyl functions were discussed previously in Sections 23-13 and 15-9, respectively. A summary of some commonly used protecting groups for NH_2, OH, SH, and CO_2H functions is in Table 25-2, together with the conditions by which the protecting groups may be removed. The best protecting groups for NH_2 functions are phenylmethoxycarbonyl (benzyloxycarbonyl) and *tert*-butoxycarbonyl. Both groups can be removed by treatment with acid, although the *tert*-butoxycarbonyl group is more reactive. The phenylmethoxycarbonyl group can be removed by reduction with either hydrogen over a metal catalyst or with sodium in liquid ammonia. This method is most useful when, in the removal step, it is necessary to avoid treatment with acid:





In most cases, formation of the ethyl ester provides a satisfactory protecting group for the carboxyl function.

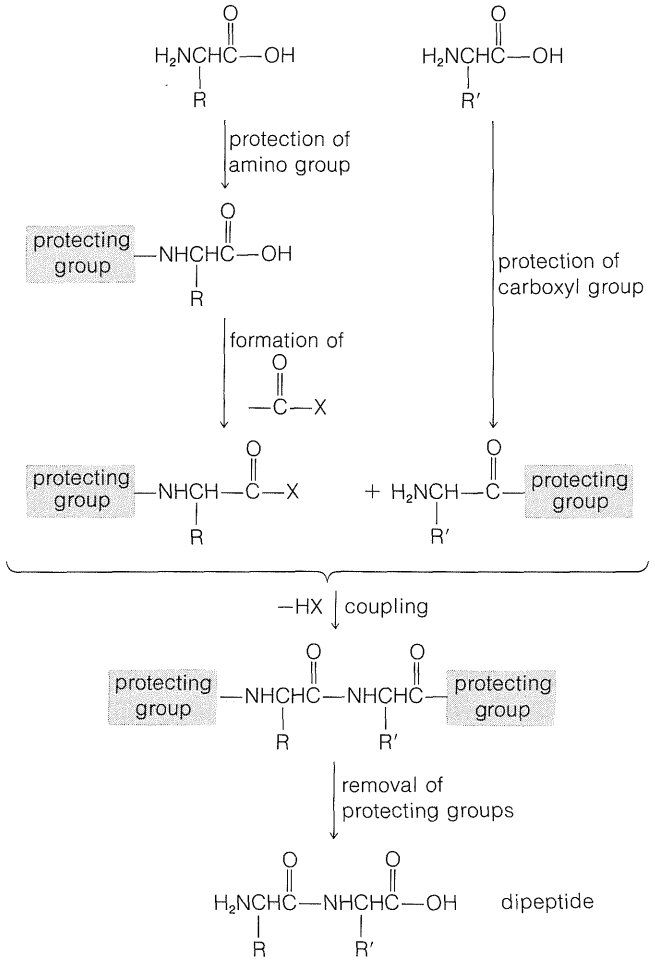


Figure 25-7 Sequence of reactions for forming a peptide bond from two different amino acids. The same type of procedure can be used to make peptide bonds between two peptides or between an amino acid and a peptide.

Conversion of the carboxyl group to a more reactive group and coupling are key steps in peptide synthesis. The coupling reaction must occur readily and quantitatively, and with a minimum of racemization of the chiral centers in the molecule. This last criterion is the Achilles' heel of many possible coupling sequences. The importance of nonracemization can best be appreciated by an example. Consider synthesis of a tripeptide from three protected L-amino acids, A, B, and C, in two sequential coupling steps, $C \xrightarrow{B} B\text{–}C \xrightarrow{A} A\text{–}B\text{–}C$. Suppose that the coupling yield is quantitative, but there is 20% formation of the D isomer in the *acylating* component in each coupling step.

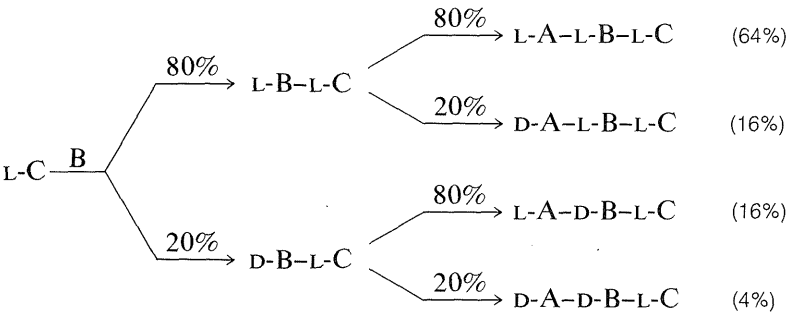
Table 25-2
Some Amine and Carboxyl Protecting Groups Used in Peptide Syntheses

Common name	Structure	Synthesis	Removed by[a]
Amino Protecting Groups			
tert-butoxycarbonyl	$(CH_3)_3COC(=O)-$	$(CH_3)_3COC(=O)Cl + H_2NR$	1, 2, 3
benzyloxycarbonyl	$C_6H_5CH_2OC(=O)-$	$C_6H_5CH_2OC(=O)Cl + H_2NR$	1, 3, 4, 5
para-toluenesulfonyl	$4\text{-}CH_3C_6H_4SO_2-$	$4\text{-}CH_3C_6H_4SO_2Cl + H_2NR$	5
triphenylmethyl (trityl)	$(C_6H_5)_3C-$	$(C_6H_5)_3CCl + H_2NR$	1, 2, 3, 4, 5
Carboxyl Protecting Groups			
ethyl	C_2H_5-	$C_2H_5OH + RCO_2H$ ($H^{\oplus}$)	6
tert-butyl	$(CH_3)_3C-$	$(CH_3)_2C{=}CH_2 + RCO_2H$ (H_2SO_4)	1, 2, 3
benzyl	$C_6H_5CH_2-$	$C_6H_5CH_2OH + RCO_2H$ ($H^{\oplus}$)	1, 3, 4, 5, 6
Hydroxyl and Thiol			
tert-butyl	$(CH_3)_3C-$	$(CH_3)_2C{=}CH_2 + ROH$ (H_2SO_4)	1, 2, 3
benzyl	$C_6H_5CH_2-$	$C_6H_5CH_2Cl + RSNa$	1, 3, 4, 5

1 = HBr, CH_3CO_2H
2 = CF_3CO_2H
3 = HF
4 = H_2, Pd
5 = Na, NH_3
6 = NaOH

[a]Reagents at room temperature except for Na,NH_3, which is carried out at the bp of $NH_3(l)$, −33°.

Then the tripeptide will consist of a mixture of four diastereomers, only 64% of which will be the desired L,L,L diastereomer (Equation 25-7):



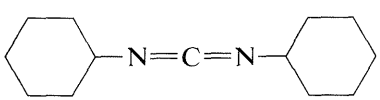
(25-7)

This is clearly unacceptable, especially for longer-chain peptides. Nine coupling steps with 20% of the wrong isomer formed in each would give only 13% of the decapeptide with the correct stereochemistry.

Exercise 25-21 How could an optically pure *N*-acylamino acid racemize and lead to racemic *N*-acylpeptides as the result of a peptide coupling reaction wherein the carboxyl group of the amino acid was converted to an anhydride group? (Review Section 25-5A.)

Exercise 25-22 Suppose there is 1% formation in each step of the wrong isomer of the acylating component in an otherwise quantitative 100-step peptide synthesis. What is the yield of the desired polypeptide isomer?

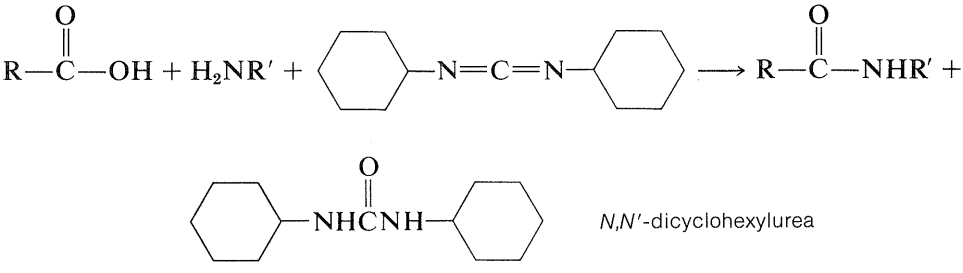
The most frequently used carboxyl derivatives in amide coupling are azides, RCO—N_3, mixed anhydrides, RCO—O—COR′, and esters of moderately acidic phenols, RCO—OAr (see Table 24-1). It also is possible to couple free acid with an amine group using a diimide, R—N=C=N—R, most frequently *N*,*N*′-dicyclohexylcarbodiimide.



N,*N*′-dicyclohexylcarbodiimide

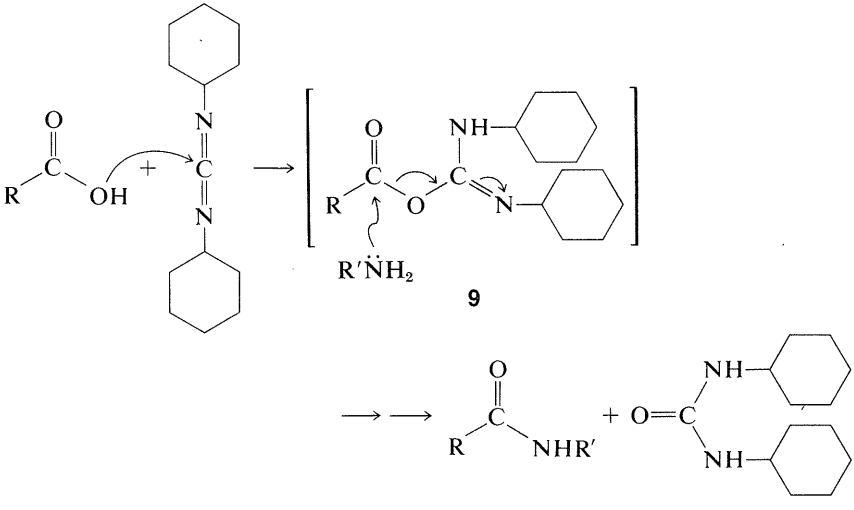
The diimide reagent may be thought of as a dehydrating agent. The "elements of water" eliminated in the coupling are consumed by the diimide to form a

substituted urea. The overall reaction is



N,N′-dicyclohexylurea

This reaction takes place because diimides, $-N=C=N-$, have reactive cumulated double-bond systems like those of ketenes, $\diagdown C=C=O$; isocyanates, $-N=C=O$; and isothiocyanates, $-N=C=S$; and are susceptible to nucleophilic attack at the central carbon. In the first step of the diimide-coupling reaction, the carboxyl function adds to the imide to give an acyl intermediate, 9. This intermediate is an activated carboxyl derivative RCO—X and is much more reactive toward an amino function than is the parent acid. The second step therefore is the aminolysis of **9** to give the coupled product and *N,N′*-dicyclohexylurea:



After completion of a coupling reaction, and before another amino acid can be added to the *N*-terminus, it is necessary to remove the protecting group. This must be done by selective reactions that do not destroy the peptide bonds or side-chain protecting groups. This part of peptide synthesis is discussed in Section 23-13, and some reactions useful for removal of the *N*-terminal protecting groups are summarized in Table 25-2.

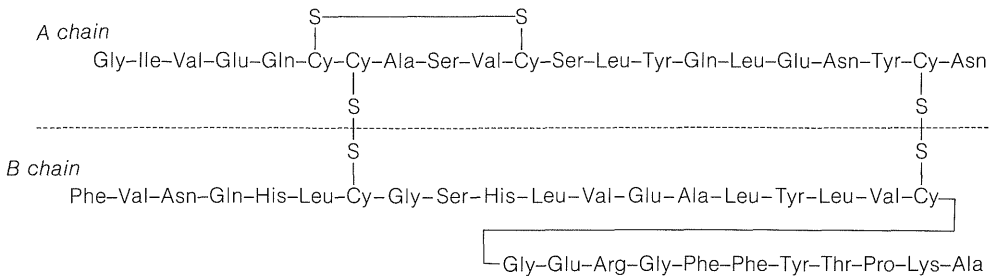
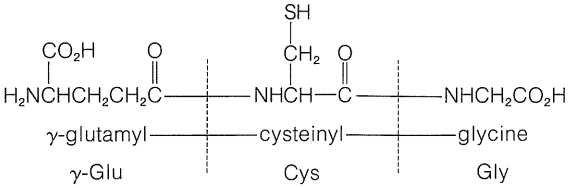


Figure 25-8 Amino-acid sequence in beef insulin. The *A* chain of 21 amino-acid residues is linked to the *B* chain of 30 residues by way of two disulfide bridges.

In spite of the large number of independent steps involved in the synthesis of even small peptides, each with its attendant problems of yield, racemization, and selectivity, remarkable success has been achieved in the synthesis of large peptides and certain of the smaller proteins. The synthesis of insulin (Figure 25-8) with its 51 amino acid units and 3-disulfide bridges has been achieved by several investigators. Several important hormonal peptides, namely glutathione, oxytocin, vasopressin, and thyrotropic hormone (see Figure 25-9) have been synthesized. A major accomplishment has been the synthesis of an enzyme with ribonuclease activity reported independently by two groups of investigators, led by R. Hirschman (Merck) and R. B. Merrifield (Rockefeller University). This enzyme is one of the simpler proteins, having a linear structure of 124 amino-acid residues. It is like a peptide, not a protein, in that it assumes the appropriate secondary and tertiary structure without biochemical intervention (Section 25-7A). As a specific example of the strategy involved in peptide synthesis, the stepwise synthesis of oxytocin is outlined in Figure 25-10, using the abbreviated notation in common usage.

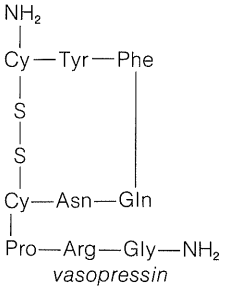
25-7D Solid-Phase Peptide Synthesis

The overall yield in a multistep synthesis of a peptide of even modest size is very poor unless each step can be carried out very efficiently. An elegant modification of classical peptide synthesis has been developed by R. B. Merrifield, which offers improved yields by minimizing manipulative losses that normally attend each step of a multistage synthesis. The key innovation is to anchor the *C*-terminal amino acid to an insoluble support, and then add amino-acid units by the methods used for solution syntheses. After the desired sequence of amino acids has been achieved, the peptide can be cleaved from the support and recovered from solution. All the reactions involved in the synthesis must, of course, be brought to essentially 100% completion so that a homogeneous product can be obtained. The advantage of having the peptide anchored to a solid support is that laborious purification steps are virtually eliminated; solid material is purified simply by washing and filtering without transferring the material from one container to another. The method has become known as **solid-phase peptide synthesis.** More of the details of the solid-phase synthesis follow.

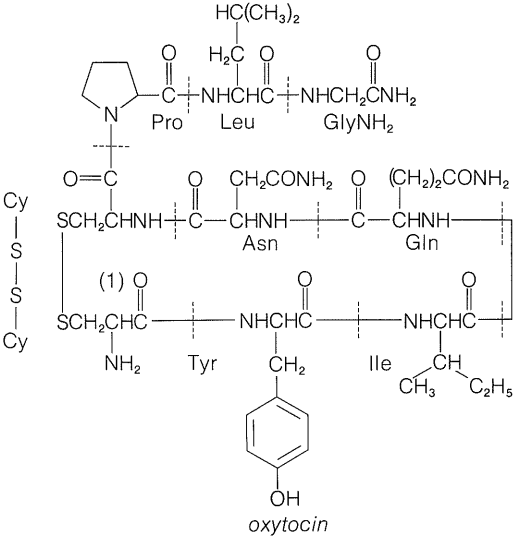


glutathione

Glutathione (GSH) is widely distributed in cell tissue. Its biological function is not completely understood but it is thought to be a coenzyme for a Cannizzaro-type reaction interconverting glyoxal and lactic acid, $CH_3COCHO \xrightleftharpoons{GSH,\ water} CH_3CH(OH)CO_2H$. GSH is very easily oxidized by air to the disulfide, $2GSH \xrightleftharpoons{-2H} GSSG$. Notice also that the peptide bond to the glutamyl residue is to C5, *not* to C1.

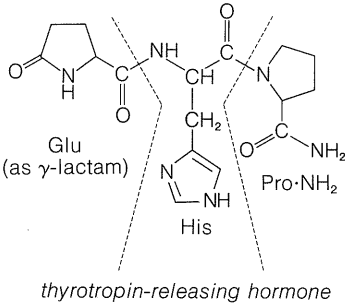


vasopressin



oxytocin

Vasopressin and oxytocin are peptide hormones secreted from the posterior lobe of the pituitary gland. They function primarily to raise blood pressure (vasopressin), as antidiuretic (vasopressin), and to promote contraction of uterus and lactation muscles (oxytocin). The isolation, identification, and synthesis of these hormones was accomplished by Vincent du Vigneaud, for which he was awarded the Nobel Prize in chemistry in 1955.



thyrotropin-releasing hormone

Thyrotropin-releasing hormone is one of several small peptide hormones secreted by the anterior lobe of the pituitary gland. These are the "master" hormones that function to stimulate hormone secretion from other endocrine glands. Thyrotropin stimulates the functioning of the thyroid gland.

Figure 25-9 Some important physiologically active peptides

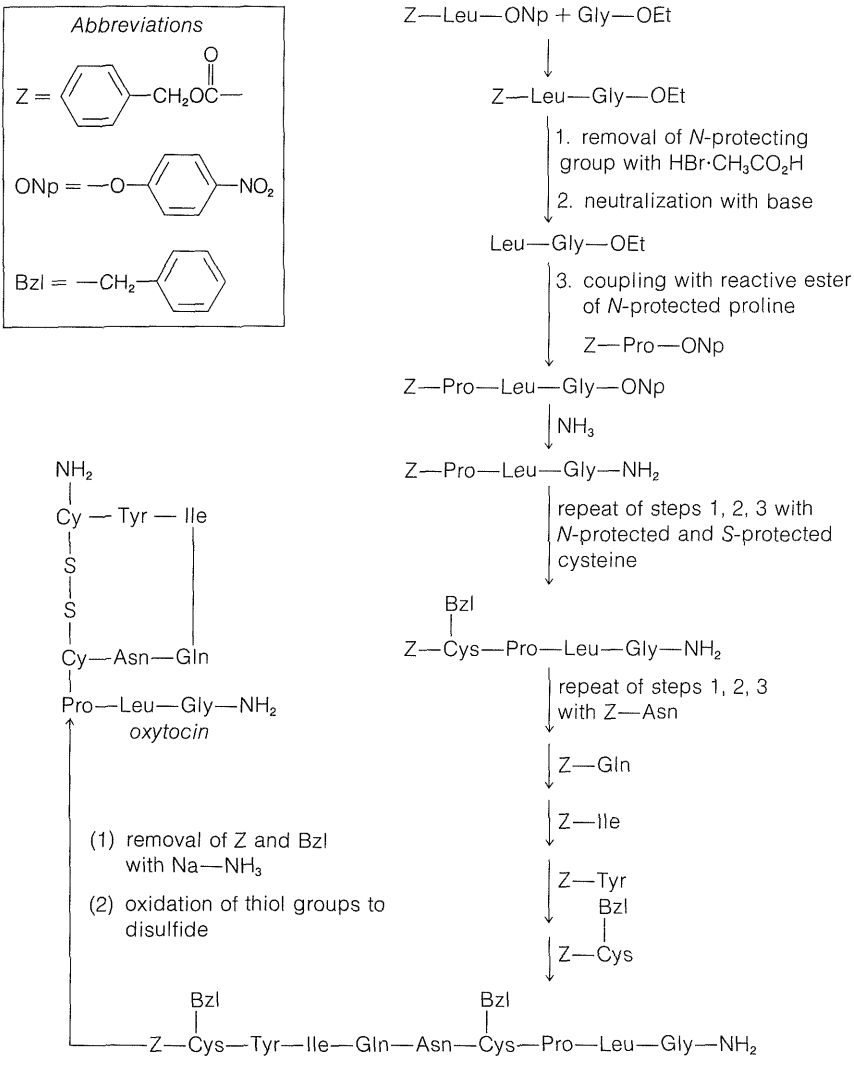
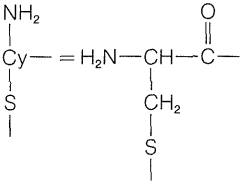
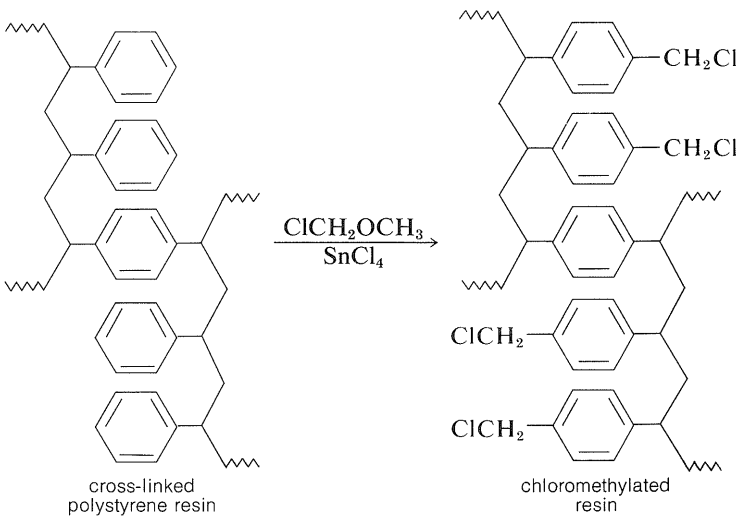


Figure 25-10 Stepwise synthesis of oxytocin by the reactive ester method. In the abbreviations used here —Gly—NH_2 = —$NHCH_2CONH_2$ and

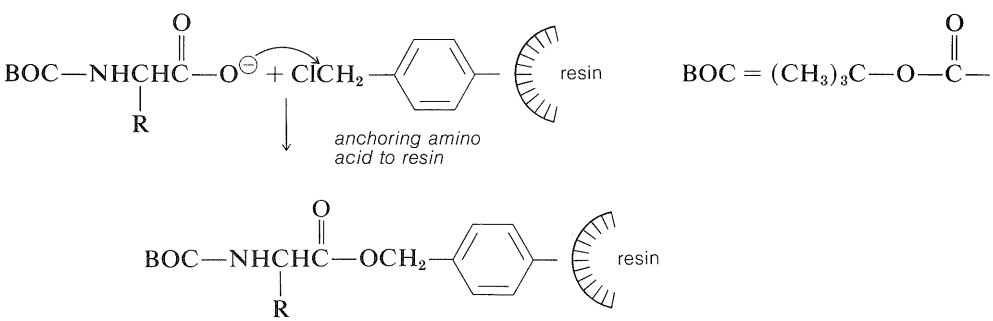


The nature of the polymer support is of great importance for a successful peptide synthesis. One that is widely used is a cross-linked polystyrene resin

of the type employed in ion-exchange chromatography (Section 25-4C). It is necessary that the resin be insoluble but have a loose enough structure to absorb organic solvents. Otherwise, the reagents will not be able to penetrate into the spaces between the chains. This is undesirable because the reactions occur on the surface of the resin particles and poor penetration greatly reduces the number of equivalents of reactive sites that can be obtained per gram of resin. Finally, to anchor a peptide chain to the resin, a reactive functional group (usually a chloromethyl group) must be introduced into the resin. This can be done by a Friedel–Crafts chloromethylation reaction (Exercise 22-21), which substitutes the $ClCH_2$— group in the 4-position of the phenyl groups in the resin:

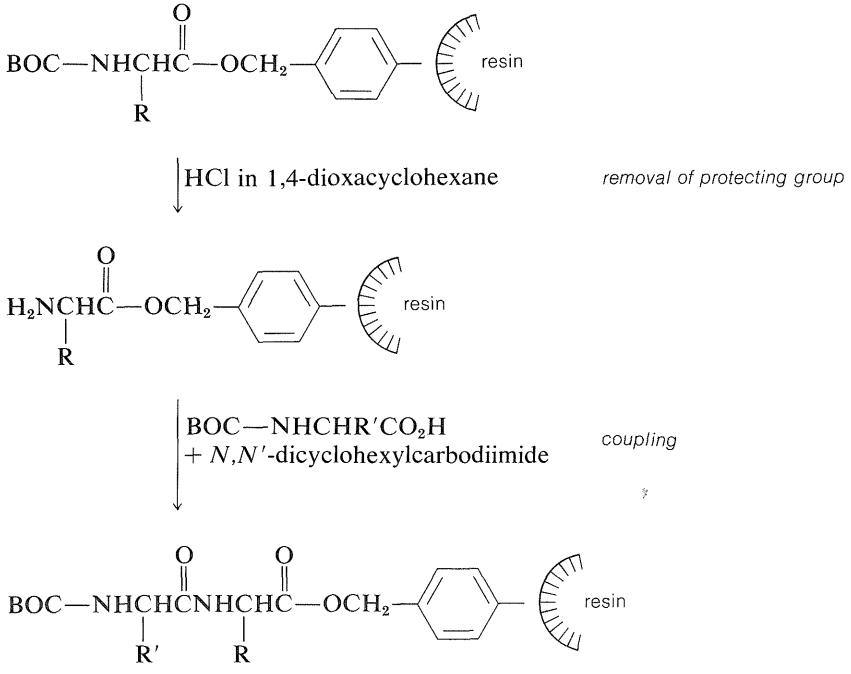


At the start of the peptide synthesis, the *C*-terminal amino acid is bonded through its carboxyl group to the resin by a nucleophilic attack of the carboxylate ion on the chloromethyl groups. The α-amino group must be suitably protected, as with *tert*-butoxycarbonyl, before carrying out this step:

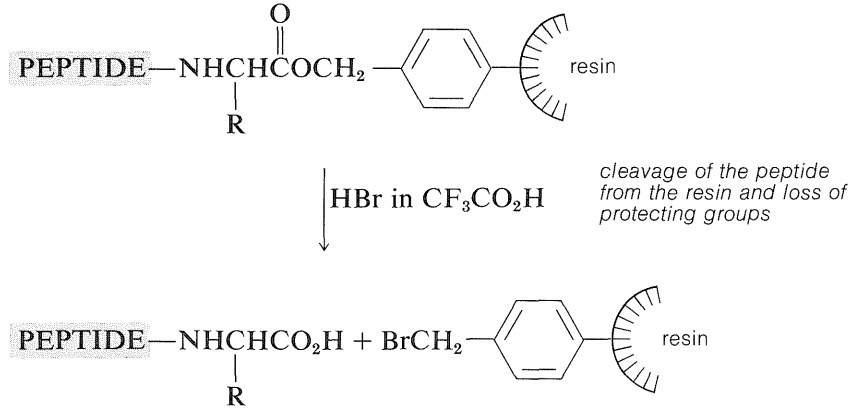


Next, the amine protecting group must be removed without cleaving the ester bond to the resin. The coupling step to a second *N*-protected amino acid

follows, with *N,N'*-dicyclohexylcarbodiimide as the coupling reagent of choice:



The peptide-bond-forming steps are repeated as many times as needed to build up the desired sequence. Ultimately, the peptide chain is removed from the resin, usually with HBr in anhydrous trifluoroethanoic acid, CF_3CO_2H, or with anhydrous HF. This treatment also removes the other acid-sensitive protecting groups.

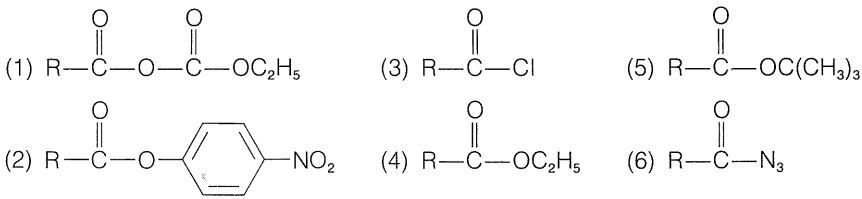


The method lends itself beautifully to automatic control, and machines suitably programmed to add reagents and wash the product at appropriate times have been developed. At present, the chain can be extended by six or so amino-

acid units a day. It is necessary to check the homogeneity of the growing peptide chain at intervals because if any step does not proceed properly, the final product can be seriously contaminated with peptides with the wrong sequence.

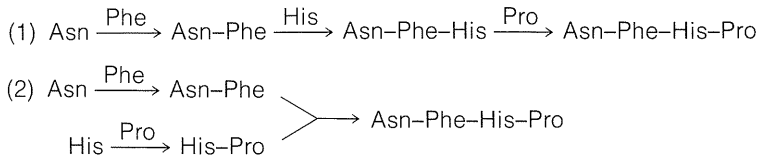
In the synthesis of the enzyme ribonuclease by the Merrifield method, the 124 amino acids were arranged in the ribonuclease sequence through 369 reactions and some 12,000 individual operations of the automated peptide-synthesis machine without isolation of any intermediates.

Exercise 25-23* The following are acyl derivatives known to react with amines to give *N*-acylated amines:



Arrange these reagents in expected order of reactivity with the free amino group of a carboxyl-protected peptide, H_2N–peptide–X, where X is the carboxyl-protecting group. Give your reasoning and indicate what disadvantages each reagent may have as to side reactions, undesirable by-products, and so on. It may be useful to review Sections 18-7A and 24-3.

Exercise 25-24 Consider two routes to the synthesis of a tetrapeptide:



If each coupling step proceeds in 80% yield, which of the two routes would give the highest overall yield?

Exercise 25-25 Indicate the steps that would be necessary to attach each of the amino acids listed to the *N*-terminus of a peptide chain. Assume that any side-chain functions in the peptide are suitably protected, but do not assume that the amino acids will couple with the peptide without suitable protection of their functional groups.
a. lysine **b.** aspartic acid **c.** cystine **d.** serine

Exercise 25-26 Show how each of the following substances may be synthesized starting with the individual amino acids. Indicate the reagents needed in each step.
a. glutamylglycine (Glu–Gly) **b.** Tyr–Ala–Val

25-7E Separation of Peptides and Proteins

In many problems of peptide sequencing and peptide synthesis it is necessary to be able to separate mixtures of peptides and proteins. The principal methods used for this purpose depend on acid-base properties or on molecular sizes and shapes.

Ultracentrifugation is widely used for the purification, separation, and molecular-weight determination of proteins. A centrifugal field, up to 500,000 times that of gravity, is applied to the solution, and molecules move downward in the field according to their mass and size.

Large molecules also can be separated by **gel filtration** (or gel chromatography), wherein small molecules are separated from large ones by passing a solution over a gel that has pores of a size that the small molecules can penetrate into and be trapped. Molecules larger than the pore size are carried on with the solvent. This form of chromatographic separation is based on "sieving" rather than on chemical affinity. A wide range of gels with different pore sizes is available, and it is possible to fractionate molecules with molecular weights ranging from 700 to 200,000. The molecular weight of a protein can be estimated by the sizes of the pores that it will, or will not, penetrate. The chemical structure of a gel of this type is described in Exercise 25-27.

The acid-base properties, and hence ionic character, of peptides and proteins also can be used to achieve separations. **Ion-exchange chromatography,** similar to that described for amino acids (Section 25-4C), is an important separation method. Another method based on acid-base character and molecular size depends on differential rates of migration of the ionized forms of a protein in an electric field (**electrophoresis**). Proteins, like amino acids, have isoelectric points, which are the pH values at which the molecules have no net charge. At all other pH values there will be some degree of net ionic charge. Because different proteins have different ionic properties, they frequently can be separated by electrophoresis in buffered solutions. Another method, which is used for the separation and purification of enzymes, is **affinity chromatography,** which was described briefly in Section 9-2B.

Exercise 25-27* A resin known as Sephadex that is useful in gel filtration is prepared from a polysaccharide that is cross-linked into a three-dimensional matrix with "epichlorohydrin,"  . The degree of cross-linking determines the pore size of the gel. Write equations, specifying the conditions as closely as possible, for reactions whereby a glucose unit of one polysaccharide chain could be linked to the glucose of another chain through an epichlorohydrin molecule.

Exercise 25-28 Hemoglobin, the protein responsible for carrying oxygen from the lungs to the body tissues, contains 0.355% iron. Hydrolysis of 100 g of hemoglobin gives 1.48 g of tryptophan. Calculate the minimum molecular weight of hemoglobin that is consistent with these results.

25-8 STRUCTURE AND FUNCTION OF PROTEINS

The biological functions of proteins are extremely diverse. Some act as hormones that regulate various metabolic processes. An example is insulin, which regulates blood-sugar levels. Enzymes act as catalysts for biological reactions, and other proteins serve as biological structural materials—for example, collagen and elastin in connective tissue and keratin in hair. Iron-containing proteins (hemoglobin and myoglobin in mammals) and copper-containing proteins (hemocyanins in shellfish) transport molecular oxygen. Some blood proteins form antibodies, which provide resistance to disease, while the so-called nucleoproteins are important constituents of the genes that supply and transmit genetic information in cell division. Motion by means of muscle contraction and the generation and transmission of nerve impulses also involve proteins.

How can a group of compounds, made from a common basis set of amino acids, be so remarkably heterogeneous and exhibit such varied yet specific functions? Clearly, the primary structure and the presence or absence of special functional groups, metals, and so on, are of paramount importance. Of complementary importance are the *three-dimensional structures* of proteins, which are dictated not just by the primary structure but by the way the primary structure is put together biochemically. The polypeptide chains are seldom, if ever, fully extended, but are coiled and folded into more or less stable conformations. As a result, amino-acid side chains in distant positions in the linear sequence are brought into close proximity, and this juxtaposition often is crucial for the protein to fulfill its specific biological function.

25-8A Three-Dimensional Structure of Proteins

The elucidation of the detailed *shape* of protein molecules—in fact, the spatial locations of the individual atoms in a protein—is accomplished primarily by x-ray crystallography. The three-dimensional structures of more than twenty proteins have now been established by this technique. The importance of x-ray crystallography to structural and biological chemistry has been recognized in the award of six Nobel Prizes in this area.[6] A number of important proteins and their properties are listed in Table 25-3.

[6]The following Nobel laureates received their awards for contributions to the use of x-ray crystallography for structure determination: 1914, Max von Laue (physics), diffraction of x rays in crystals; 1915, William Bragg and Lawrence Bragg (physics), study of crystal structure by means of x rays; 1954, Linus Pauling (chemistry), study of structure of proteins; 1962, Max Perutz and John Kendrew (chemistry), structures of myoglobin and hemoglobin; 1962, Francis Crick, James Watson, and Maurice Wilkins (physiology and medicine), double helix of DNA; 1964, Dorothy Hodgkin (chemistry), determination of structure of vitamin B_{12} and penicillin by x-ray methods. She later determined the three-dimensional structure of insulin.

Table 25-3
A Few Important Proteins of Known Structure

Name	Approx. MW	Amino acids	Disulfide bonds	Prosthetic group	Isoelectric point	Occurrence	Function
insulin	5,800	51	3	—	5.4	pancreas	regulation of blood sugar levels
ribonuclease	13,700	124	4	—	7.8	pancreas	enzyme that hydrolyzes RNA
myoglobin	17,800	153	—	heme		muscle	respiratory protein; stores O_2 in muscle tissue
hemoglobin	64,500	α-141[a] β-146	—	heme	6.7	red blood corpuscles	respiratory protein; transports O_2 from lungs; transports CO_2 to lungs
cytochrome *c*	12,800	104	—	heme[b]		all cells	respiratory protein; electron carrier for *oxidative phosphorylation*[c]
lysozyme	14,600	129	4	—	10.7	egg white	enzyme that breaks down the cell walls of bacteria by hydrolysis of $\beta(1 \rightarrow 4)$ glycoside linkages
α-chymotrypsin	24,500	241	5	—	8.4	pancreas	digestive enzyme; hydrolyzes ester and peptide bonds
carboxypeptidase A	34,600	307	—	Zn		pancreas	digestive enzyme; hydrolyzes carboxyterminal peptide bond in proteins

[a]Hemoglobin has four subunits, two α chains and two β chains, each with a heme group.

[b]Cytochrome *c* has two cysteine units covalently bonded to the two ethenyl side chains of the heme group:
protein—SH + CH_2=CH—heme ⟶ protein—S—CH_2CH_2—heme

[c]Oxidative phosphorylation is the process in which ATP is formed as electrons are transferred (by way of the cytochromes) from NADH or $FADH_2$ to O_2. For example, $\frac{1}{2}O_2 + NADH + H^{\oplus} \longrightarrow H_2O + NAD^{\oplus}$, and the energy from this process is used to synthesize ATP (see Section 20-10).

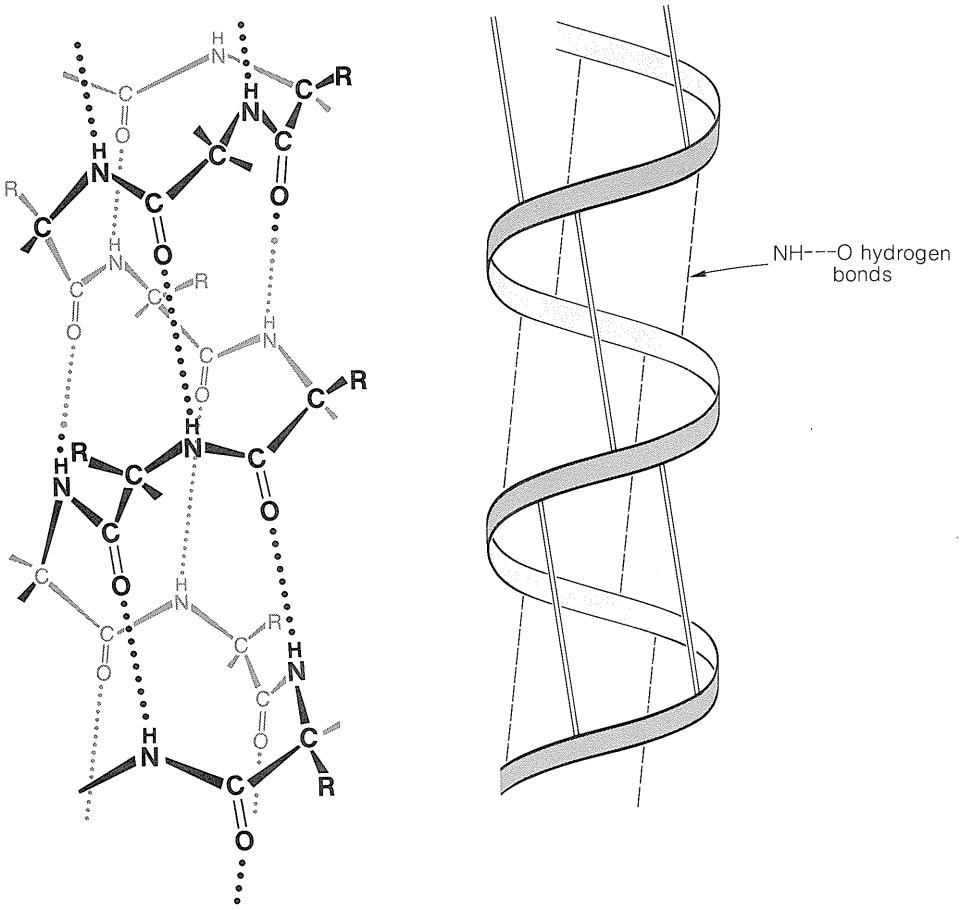


Figure 25-11 Peptide chain of a protein coiled to form a right-handed alpha helix. Configuration of the helix is maintained by hydrogen bonds, shown as vertical dotted (or solid) lines. The helix on the left shows the detailed atom structure of the peptide chain. The helix on the right is a schematic representation without structural detail.

An especially favorable conformation of a polypeptide chain that was originally deduced by Pauling and Corey is the **alpha helix** (Figure 25-11). The principal feature of the α helix is the coiling of the polypeptide chain in such a way as to form hydrogen bonds of the type $\backslash N—H\cdots O=C/$ between amide N—H and amide carbonyl groups that are *four* amino-acid units apart. The coiling is possible because the chain can twist about the C_α—C and C_α—N single bonds of most amino acid units, as shown in Figure 25-12.

There are several other points to notice about the α helix shown in Figure 25-11. The amide groups are planar and normally retain the stable *trans* configuration in the helical structure; bond lengths and bond angles are

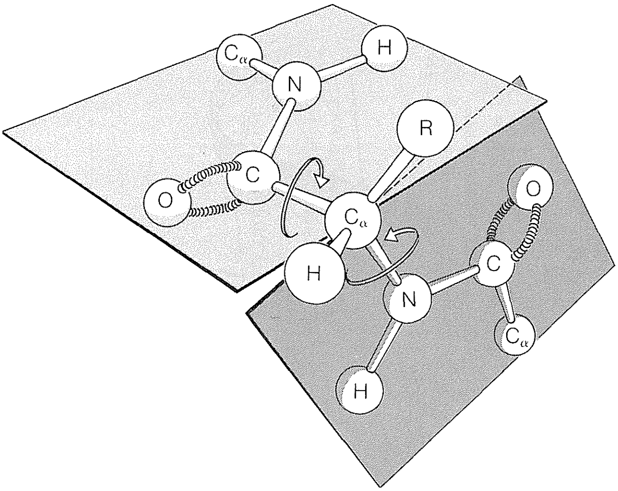


Figure 25-12 Ball-and-stick model of a peptide unit showing the coplanarity of the CNCC atoms of the amide linkage, here in the trans configuration, and the possibility of rotation about the C–C_α and N–C_α bonds.

normal, and the $>NH\cdots O=C<$ hydrogen bonds are nearly linear. However, the hydrogen bonds are not quite parallel to the long axis of the coil, so there are 3.6 rather than 4 amino-acid units per helical turn, and the spacing between turns is about 5.4 A. The α helix in proteins has *right-handed* turns like a right-hand screw thread.

The amino acids of the side chains lie outside the coil of the α helix and are in close proximity to the side chains three and four amino-acid units apart. Because of this proximity, steric hindrance between larger side chains can be sufficient to reduce the stability of the normal α helix. When such hindrance occurs, there is a discontinuity in the helical structure, and the peptide chain may assume more random arrangements about the $C—C_\alpha$ and $N—C_\alpha$ bonds (see Figure 25-12), thereby allowing the molecule to fold back on itself and form new hydrogen bonds. The helical structure apparently is always interrupted at proline or hydroxyproline residues because the $C_\alpha—N$ bonds of these amino acids are not free to rotate (they are incorporated in five-membered rings) and also because the proline and hydroxyproline amide nitrogens have no hydrogens to participate in hydrogen bonding to carbonyl groups.

Pauling and Corey recognized a second stable conformation of polypeptide chains—the extended chain or β-pleated sheet (Figure 25-13). In this conformation the chains are fully extended with trans amide configurations. In this arrangement the distance is maximized between adjacent amino-acid side chains. Hydrogen bonding of the type $>N—H\cdots O=C<$ is now *between*

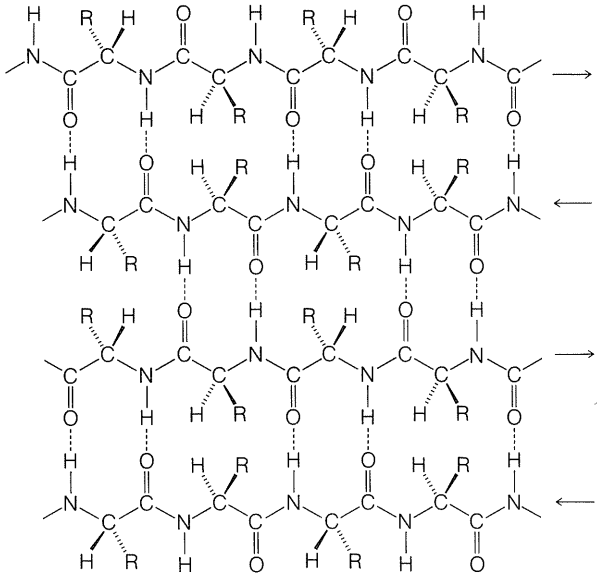
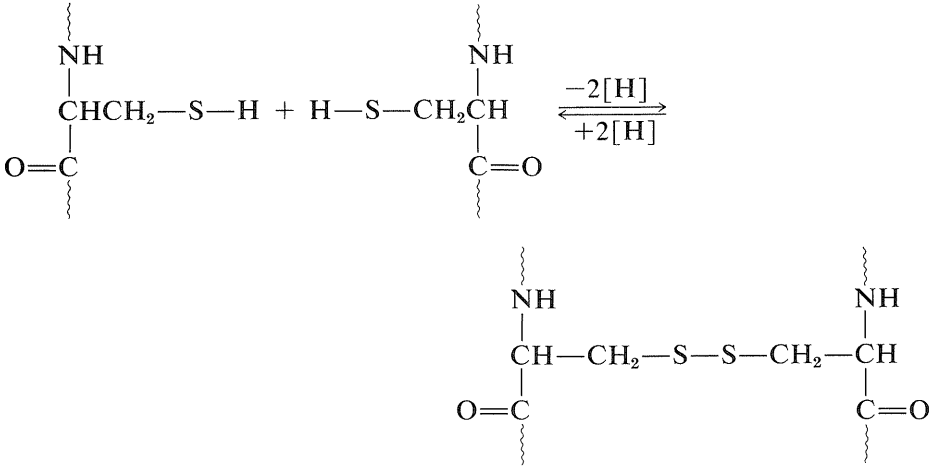


Figure 25-13 Hydrogen-bonded structure of silk fibroin. Notice that the peptides run in different directions in alternate chains. This structure is called an antiparallel β-pleated sheet.

chains rather than between amino acids in a single chain (as in the α helix). This type of structure is not as common as the α helix and is found extensively only in silk fibroin. However, a number of proteins with a single polypeptide chain can form short sections of "antiparallel" β-pleated sheets by folding back on themselves, as illustrated in Figure 25-14.

Another very important factor in protein architecture is the disulfide —S—S— link. Remote parts of the polypeptide chain can be held close together through the oxidative coupling of two cysteine thiol groups to form a disulfide bridge:



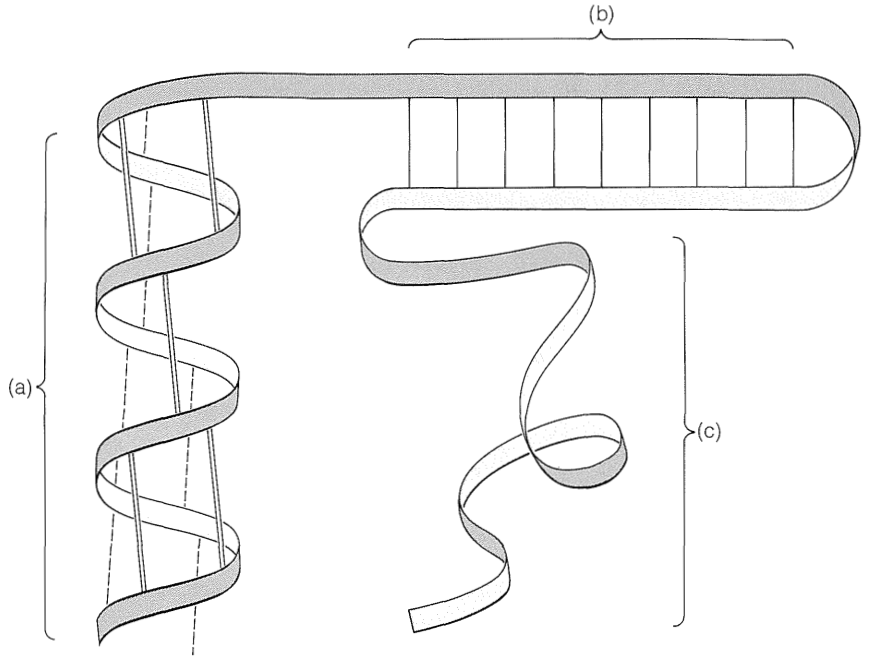


Figure 25-14 Diagrammatic representation of the coiling of a protein chain showing areas of (a) α helix, (b) β-pleated sheet, and (c) random coiling

Such —S—S— bridges greatly restrict the number of conformations available to a protein and are of fundamental importance in determining the shape of a protein, and hence, its biological activity. Lysozyme, which can be isolated from hen egg-white, provides an excellent example. This substance is an enzyme that catalyses hydrolysis of the glycoside links in polysaccharide components of bacterial cell walls. It is a relatively small protein of 129 amino acid units in a single polypeptide chain that is cross-linked by four disulfide bridges (Figure 25-15). It becomes inactive if the disulfide bridges are cleaved or combined in other combinations than the ones shown.

The disulfide bridges in some proteins are between different peptide chains. Insulin, for instance, has two interchain as well as one intrachain S—S bridges (Figure 25-8).

25-8B Myoglobin and Hemoglobin

Some idea of the complexity of protein conformations can be gained from the structure of myoglobin. This protein is responsible for the storage and

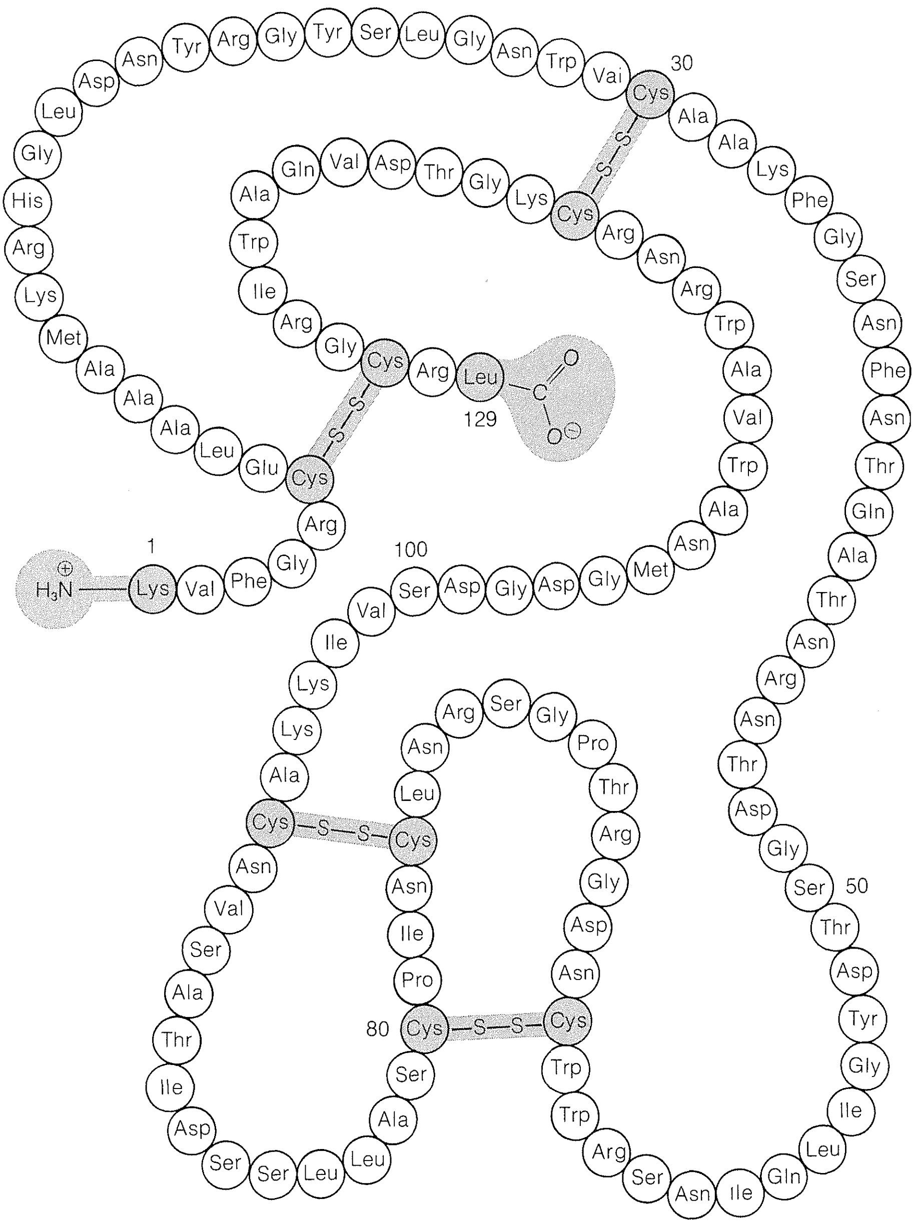


Figure 25-15 Lysozyme from hen egg-white showing the amino-acid sequence (primary structure) and the four intrachain disulfide bridges. [Adapted from D. C. Phillips, *Sci. Amer.* **5,** 215 (1966).]

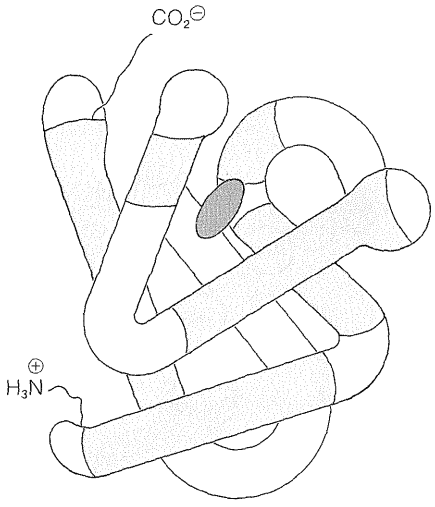


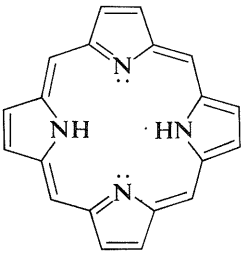
Figure 25-16 A model of myoglobin to show the way in which the polypeptide chain is coiled and folded. The shaded sections correspond to regions in which the chain is coiled into an α helix. Each fold, and the regions near the C-terminus and the N-terminus, represent discontinuities in the helical structure. The position of the heme group is represented by the disclike shape.

transport of molecular oxygen in the muscle tissue of mammals. It is a compact molecule of 153 amino-acid units in a chain that is extensively coiled as an α helix. There are eight regions of discontinuity in the helical structure, and in these regions the chain folds on itself as shown in Figure 25-16. Four of the eight nonhelical regions occur at proline residues; the reason for the discontinuity at the other regions is not entirely clear. With the exception of two histidine units, the interior regions of myoglobin accommodate only the nonpolar side chains; the interior, therefore, is mostly hydrocarbonlike and repellent to water and other polar molecules. In contrast, the polar side chains are on the exterior of the protein.

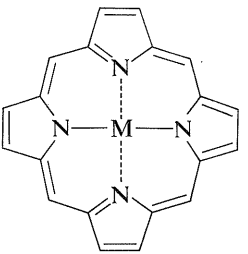
A number of proteins, including myoglobin, possess one or more nonpeptide components associated with specific sites on the polypeptide chain. These components are called **prosthetic groups** and are essential to the biological activity. When the prosthetic group is removed, the residual protein is referred to as an **apoprotein.**

In myoglobin the prosthetic group is a molecule of **heme.** The heme group belongs to a class of interesting compounds called **metalloporphyrins,** which are metal complexes of a highly conjugated ring system composed of four azacyclopentadiene (pyrrole) rings linked by —CH= bridges between the 2 and 5 positions. The parent compound is known as **porphin.** Porphyrins have

highly stabilized electronic excited states and absorb visible light. As a result they usually are brightly colored compounds (e.g., chlorophyll, Figure 20-6).

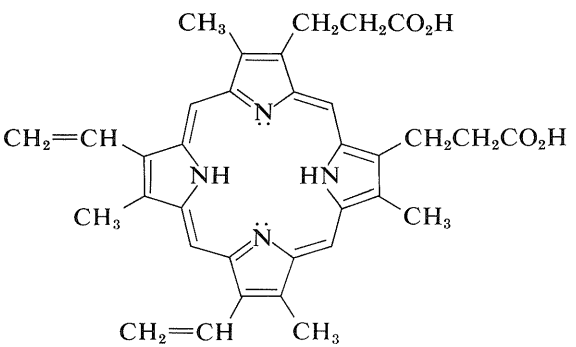


porphin

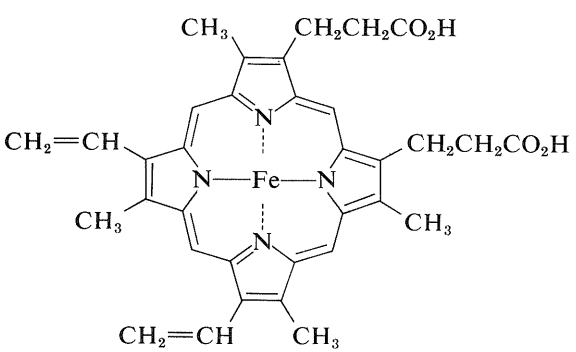


metalloporphyrin
M = Fe, Cu, Mg, Zn, Cr, and other metals

The porphyrin of heme is known as protoporphyrin IX, and the associated metal is iron [as Fe(II) or Fe(III)]. You will notice that the porphyrin ring carries methyl, ethenyl, and propanoic acid side chains:



protoporphyrin IX

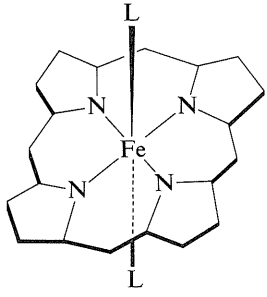


heme

A major effort on the part of several eminent chemists in the early part of the century led to the elucidation of the structure of heme. The German chemist Hans Fischer successfully synthesized heme in 1929, a feat for which, in 1930,

he received the Nobel Prize in chemistry. [Some years earlier (1915), Richard Willstätter received a Nobel Prize for structural studies of chlorophyll and plant pigments.]

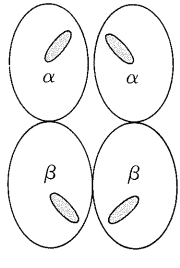
A very important question is, how does the particular combination of protein and iron-porphyrin allow myoglobin to reversibly bind molecular oxygen? The answer to this question is not known in all its details, but it is well established that Fe(II)-porphyrins will complex readily and reversibly with oxygen. There are two additional coordination sites around the iron in heme besides the four ring nitrogens. These are indicated below as the general ligands L:



The disclike heme molecule fits into a cleft in the protein structure and is bound to it through one of the L coordination sites to a histidine nitrogen. The remaining coordination site on the other side of the ring is occupied by molecular oxygen. In the absence of the coordination by histidine, the porphyrin iron would be oxidized rapidly to the ferric state, which does not bind oxygen.

A number of model compounds have been synthesized which have Fe(II)-porphyrin rings carrying a side chain with histidine arranged to be able to coordinate with the metal on one side. Several of these substances show promise as oxygen carriers with properties similar to myoglobin.

Hemoglobin is related to myoglobin in both its structure and function. It reversibly binds molecular oxygen which it transports in the red corpuscles of blood rather than in muscle tissue. However, hemoglobin is made up of *four* polypeptide chains, in contrast to myoglobin which has only *one* chain. Two of the hemoglobin chains are of one kind with 141 amino acid residues, called the α chains, and two are of another kind with 146 amino acids, called the β chains. Each chain, or *subunit,* contains one heme group identical with the heme in myoglobin. The subunits are held in the hemoglobin by noncovalent interactions and provide four hemes and hence four binding sites for molecular oxygen. The α and β hemes have different affinities for oxygen but function in a cooperative way to increase oxygen availability to the cells.



arrangement of the four chains and four heme groups of hemoglobin

In spite of the fact that the α and β chains of hemoglobin are nonidentical with the myoglobin chain, the three-dimensional structures of all three chains are strikingly similar; myoglobins and hemoglobins differ slightly in amino acid composition, depending on the species, but the protein shape remains essentially the same.

25-8C Quaternary Structures of Proteins

Many factors contribute to the three-dimensional structures of proteins. We already have mentioned hydrogen bonding between amide groups, location and character of prosthetic groups, and disulfide bonds. Other important influences include electrostatic interactions between ionic groups ($—NH_3^{\oplus}$, $—CO_2^{\ominus}$), hydrogen-bonding involving side-chain substituents ($—CH_2OH$), and nonbonded interactions. Except for the disulfide linkages, most of these interactions are weak compared to covalent bond strengths, and the conformations of many proteins can be altered rather easily. In fact, several have conformations that clearly are in dynamic equilibrium under physiological conditions. Such structural flexibility may be necessary for the protein to be functional, but if the conformation is altered *irreversibly*—that is, if it is denatured—its biological activity usually is destroyed.

In many cases there are important interactions between protein molecules that may lead to highly organized structures such as the pleated sheet of silk fibroin (Figure 25-13) or the coiling of α helices, as found in α-keratins, the fibrous proteins of hair, horn, and muscles (Figure 25-17). This sort of organization of protein molecules is called **quaternary** structure and is an important feature of many proteins that associate into dimers, tetramers, and so on. The tetrameric structure of hemoglobin is an important example.

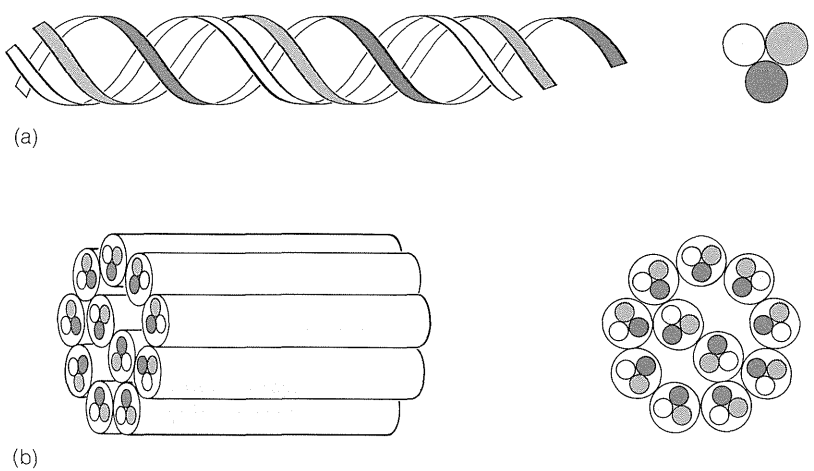


Figure 25-17 Representation of the quaternary structure of α-keratin showing (a) three α-helical polypeptide strands coiled into a rope and (b) eleven units of the three-stranded rope arranged to form one microfibril

25-9 ENZYMES

Virtually all biochemical reactions are catalyzed by proteins called enzymes. The catalytic power and specificity of enzymes is extraordinarily high. The reactions that they catalyze are generally enhanced in rate many orders of magnitude, often as much as 10^7, over the nonenzymatic process. Consequently enzymatic reactions may occur under much milder conditions than comparable laboratory reactions. For example, the simple hydrolysis of an amide proceeds at a practical rate only on heating the amide in either strongly acidic or strongly basic aqueous solution, and even then reaction may not be complete for several hours. In contrast, hydrolysis of amide or peptide bonds catalyzed by typical proteolytic enzymes, such as trypsin, chymotrypsin, or carboxypeptidase A, occurs rapidly at physiological temperatures and physiological pH.[7] It is one of the remarkable attributes of many enzymes that they catalyze reactions that otherwise would require strongly acidic or basic conditions. Enzymes are strictly catalysts, however, and affect only the *rate* of reaction, not the position of equilibrium; they lower the energy of the transition state, not the energies of the reactants or products (see Figure 4-4).

Many enzymes appear to be tailor-made for one specific reaction involving only one reactant, which is called the *substrate*. Others can function more generally with different reactants (substrates). But there is no such thing as a universal enzyme that does all things for all substrates. However, nothing seems to be left to chance; even the equilibration of carbon dioxide with water is achieved with the aid of an enzyme known as carbonic anhydrase.[8] Clearly, the scope of enzyme chemistry is enormous, yet the structure and function of relatively few enzymes are understood in any detail. We can give here only a brief discussion of the mechanisms of enzyme action—first some general principles then some specific examples.

25-9A Aspects of the Mechanisms of Enzyme Reactions

An enzyme usually catalyzes a single chemical operation at a very specific position, which means that only a small part of the enzyme is intimately involved. The region of the enzyme structure where key reactions occur as the

[7]The slowness with which amide bonds are hydrolyzed in the presence of either strong acids or strong bases, and their susceptibility to hydrolysis under the influence of enzymes, clearly is a key advantage in the biological functioning of peptides. Amide hydrolysis in neutral solution has a favorable, but not large, equilibrium constant. Therefore it does not take a great deal of biochemical energy either to form or to hydrolyze peptide bonds. The resistance to ordinary hydrolysis provides needed stability for proteins, and yet when it is necessary to break down the peptide bonds of proteins, as in digestion, this can be done smoothly and efficiently with the aid of the proteolytic enzymes.

[8]Many enzymes are named by adding the suffix *-ase* to a word, or words, descriptive of the type of enzymatic activity. Thus, *esterases* hydrolyze esters, *proteinases* hydrolyze proteins, *reductases* achieve reductions, and *synthetases* achieve syntheses of polypeptide chains, nucleic acid chains, and other molecules.

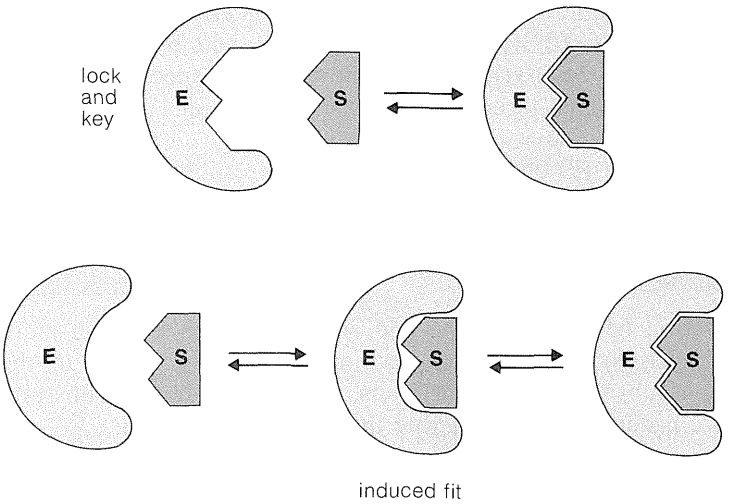


Figure 25-18 Illustration of the lock-and-key concept of enzyme-substrate interaction (top) and of the induced-fit theory, whereby the enzyme molds to the substrate through conformational changes (bottom)

result of association of the substrate with the enzyme is called the **active site.** The initial association of the enzyme (E) and the substrate (S) is formation of an enzyme-substrate complex (ES):

$$E + S \rightleftharpoons ES$$

Complexation could occur in many different ways, but for the intimate complexation required for catalysis, the enzyme must have, or must be able to assume, a shape complementary to that of the substrate. Originally, it was believed that the substrate fitted the enzyme somewhat like a key in a lock; this concept has been modified in recent years to the *induced-fit* theory, whereby the enzyme can adapt to fit the substrate by undergoing conformational changes (Figure 25-18). Alternatively, the substrate may be similarly induced to fit the enzyme. The complementarity is three-dimensional, an important factor in determining the specificity of enzymes to the structure and stereochemical configuration of the substrates.

Detailed structures for the active sites of enzymes are difficult to obtain and have been worked out only for a few enzymes that have been studied extensively by both chemical and x-ray methods. Very revealing information has been obtained by x-ray diffraction studies of complexes between the enzyme and *nonsubstrates,* which are molecules similar to actual substrates and complex with the enzyme at the active site, but do not react further. These substances often *inhibit* reaction of the normal substrate by associating strongly with the enzyme at the active site and not moving onward to products. The x-ray studies of enzymes complexed with nonsubstrates show that the active site generally is a cleft or cavity in the folded structure of the enzyme that is largely hydrophobic in character. The enzyme-substrate complex can

be inferred to be held together largely by van der Waals attractive forces between like groups (Section 12-3C), hydrogen-bonding, and by electrostatic attraction between ionic or polar groups. To achieve a stereospecific catalyzed reaction, there must be at least three points of such interactions to align properly the substrate within the cavity of the enzyme.

The reaction of the ES complex may convert the substrate to product (P) directly, and simultaneously free the enzyme (E) to react with more of the substrate:

$$ES \longrightarrow P + E$$

However, the reaction between enzyme and substrate often is much more complex. In many cases, the substrate becomes covalently bound to the enzyme. Then, in a subsequent step, or steps, the enzyme-bound substrate (ES′) reacts to give products and regenerate the active enzyme (E):

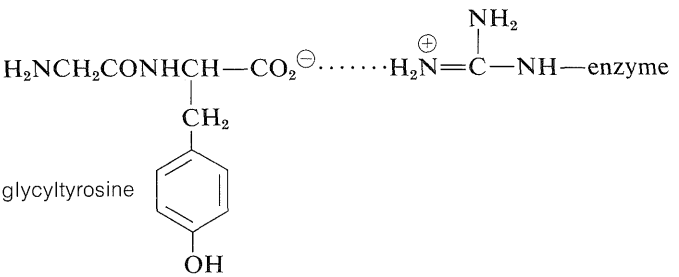
$$ES \longrightarrow ES' \longrightarrow P + E$$

25-9B Carboxypeptidase A

The considerable detail to which we now can understand enzyme catalysis is well illustrated by what is known about the action of carboxypeptidase A. This enzyme (Section 25-7B and Table 25-3) is one of the digestive enzymes of the pancreas that specifically hydrolyze peptide bonds at the *C*-terminal end. Both the amino-acid sequence and the three-dimensional structure of carboxypeptidase A are known. The enzyme is a single chain of 307 amino-acid residues. The chain has regions where it is associated as an α helix and others where it is associated as a β-pleated sheet. The prosthetic group is a zinc ion bound to three specific amino acids and one water molecule near the surface of the molecule. The amino acids bound to zinc are His 69, His 196, and Glu 72; the numbering refers to the position of the amino acid along the chain, with the amino acid at the *N*-terminus being number 1. The zinc ion is essential for the activity of the enzyme and is implicated, therefore, as part of the active site.

X-ray studies[9] of carboxypeptidase complexed with glycyltyrosine (with which it reacts only slowly) provide a detailed description of the active site, which is shown schematically in Figure 25-19a and is explained below.

1. The tyrosine carboxylate group of the substrate is associated by electrostatic attraction with the positively charged side chain of arginine 145 (W):



[9]W. N. Lipscomb, *Accounts of Chemical Research* **3**, 81 (1970); E. T. Kaiser and B. L. Kaiser, *ibid.* **5**, 219 (1972). Lipscomb received the 1976 Nobel Prize in chemistry for structural work on boranes.

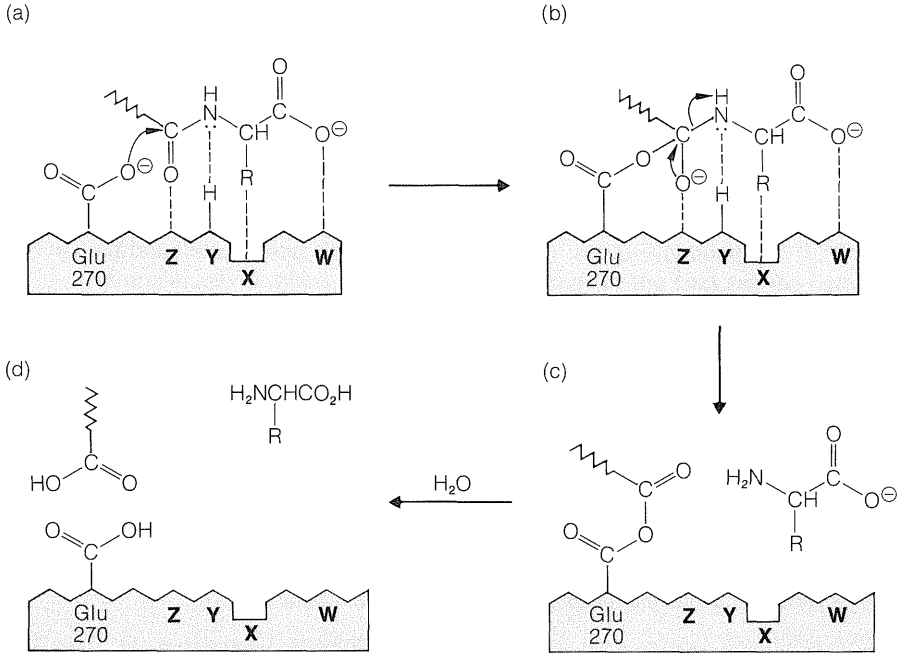


Figure 25-19 Steps in a possible mechanism of carboxypeptidase action. (a) The substrate is shown complexed to the enzyme surface through X, Y, Z, and W; X is a nonpolar pocket; Y is a hydrogen bond, possibly from OH of Tyr 248; Z is the prosthetic group, Zn; and W is an ionic interaction with $=\overset{\oplus}{N}H_2$ of Arg 145. The C-terminal amide bond of the substrate is held close to the catalytic site, which is the carboxyl of Glu 270. (b) A tetrahedral intermediate could be formed by attack of Glu 270 carboxylate anion at the amide carbonyl of the substrate. (c) Cleavage of the tetrahedral intermediate of (b) releases the C-terminal amino acid and forms an acyl-enzyme intermediate. (d) The residue of the substrate chain is released from the enzyme by hydrolysis of the acyl-enzyme intermediate. These drawings are deficient in that they try to reproduce a three-dimensional situation in two dimensions. The third dimension is especially important in understanding the stereospecificity of the enzyme.

2. The tyrosine side chain of the substrate associates with a nonpolar pocket in the enzyme (**X**).

3. Hydrogen bonding possibly occurs between the substrate tyrosine amide unshared pair and the side-chain HO groups of the enzyme tyrosine 248 (**Y**).

4. The glycyl carbonyl oxygen in the substrate probably is coordinated with the zinc ion (**Z**), displacing the water molecule coordinated to the zinc in the uncomplexed enzyme.

5. A side-chain carboxylate anion of glutamic acid 270 is so situated with respect to the reaction center that it could well function as a nucleophile by attacking the glycine carbonyl carbon.

The arrangement of the enzyme-substrate complex suggests a plausible reaction mechanism analogous to nonenzymatic mechanisms of amide hydrolysis (Section 24-4). The carboxyl group of Glu 270 can add to the amide carbonyl

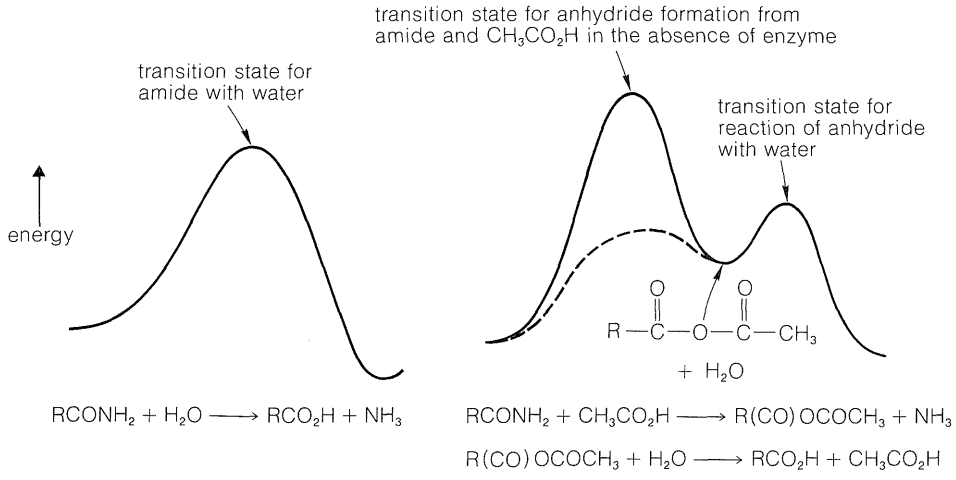


Figure 25-20 Possible energy profiles for different pathways for hydrolysis of a simple amide. The dashed line represents an enzyme-catalyzed formation of $R(CO)OCOCH_3$, which is here hypothesized to have a lower-energy transition state than the reaction of the anhydride with water (see text).

to form a tetrahedral intermediate that then rapidly dissociates to release the terminal amino acid, leaving the rest of the substrate bound to the enzyme as a mixed anhydride which can be symbolized as $E—C(=O)—O—C(=O)—R_1$. Reaction of the acyl-enzyme intermediate with water will release the peptide, minus the terminal amino acid, and regenerate the enzyme.

This postulated sequence of events may leave you wondering why the enzyme speeds up the hydrolysis, especially because the sequence proceeds through an energetically unfavorable reaction, the formation of a *carboxylic anhydride* from an *amide* and a *carboxylic acid:*

$$\underset{\text{peptide}}{R'C(=O)NHR} + \underset{\text{Glu 270}}{HO_2CR''} \longrightarrow \underset{\text{acyl-enzyme intermediate}}{R'C(=O)OC(=O)R''} + \underset{\substack{\text{peptide–}\\ \text{C-terminal}\\ \text{amino acid}}}{H_2NR}$$

The key point is that there is nothing necessarily wrong with formation of an energetically unfavorable intermediate. For effective catalysis, the energy of the least-favorable transition state between starting materials and products must be *lower* than the least-favorable transition state for the uncatalyzed reaction. The only way that formation of an unfavorable intermediate can slow the rate of the reaction is when its energy, or that of a transition state leading to it, is the highest energy point along the reaction path. Figure 25-20 illustrates this for hydrolysis of a simple amide by direct attack of water on the carbonyl carbon, or through forming an anhydride with a carboxylic acid. In either case, the overall energy change is the same if both reactions are carried out at the same pH. The reaction with CH_3CO_2H in Figure 25-20 is shown with a higher-

energy transition state than with water, as would be expected, because CH_3CO_2H is a poorer nucleophile than water.

How, then, can we possibly expect that the enzyme could make the anhydride route be faster than the simple water route? The answer is the way in which interactions in the enzyme-substrate complex can stabilize the transition state for the anhydride-forming reaction. Thus for carboxypeptidase the zinc can act as a strong electrophile to facilitate attack on the amide carbonyl. Hydrogen bonding of the amide nitrogen to Tyr 248 both will facilitate attack on the carbonyl group and assist in the breaking of the C—N bond. Furthermore, the nonpolar environment of the alkyl side-chains of the enzyme will increase the nucleophilicity of the $-CO_2^{\ominus}$ group that forms the anhydride (see Section 8-7F). Inspection of Figure 25-20 shows qualitatively that if the energy of the transition state for formation of the anhydride is lowered greatly, the overall rate will be determined by the rate of hydrolysis of the anhydride! In this circumstance (assuming the anhydride hydrolysis is uncatalyzed), the efficiency of the enzyme in breaking the peptide bond is as great as it can be, at least by this particular pathway of peptide hydrolysis. Such efficiencies have been established for other enzymes.

25-9C Participation of Serine in Enzyme Action

Mechanisms similar to the one described for carboxypeptidase appear to operate in the hydrolysis of amide and ester bonds catalyzed by a number of proteinases and esterases. The substrate, here generalized as $R-\overset{\overset{O}{\|}}{C}-X$, acts to acylate the enzyme, which subsequently reacts with water to give the observed products:

$$\text{enzyme(H)} + R\overset{\overset{O}{\|}}{C}-X \longrightarrow \text{enzyme}-\overset{\overset{O}{\|}}{C}-R + HX$$

$$\downarrow H_2O$$

$$\text{enzyme(H)} + HO\overset{\overset{O}{\|}}{C}R$$

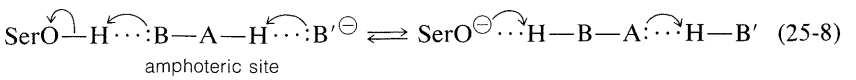
In the acylation step a nucleophilic group on one of the amino-acid side chains at the active site behaves as the nucleophile. As we have seen in Section 25-9B, the nucleophile of carboxypeptidase is the free carboxyl group of glutamic acid 270. In several other enzymes (chymotrypsin, subtilisin, trypsin, elastase, thrombin, acetylcholinesterase), it is the hydroxyl group of a *serine* residue:

$$\text{(E)}-\text{SerOH} + X-\overset{\overset{O}{\|}}{C}-R \rightleftharpoons \text{(E)}-\text{Ser}-O-\overset{\overset{O}{\|}}{C}-R + HX$$

$$\Updownarrow H_2O$$

$$\text{(E)}-\text{SerOH} + HO-\overset{\overset{O}{\|}}{C}-R$$

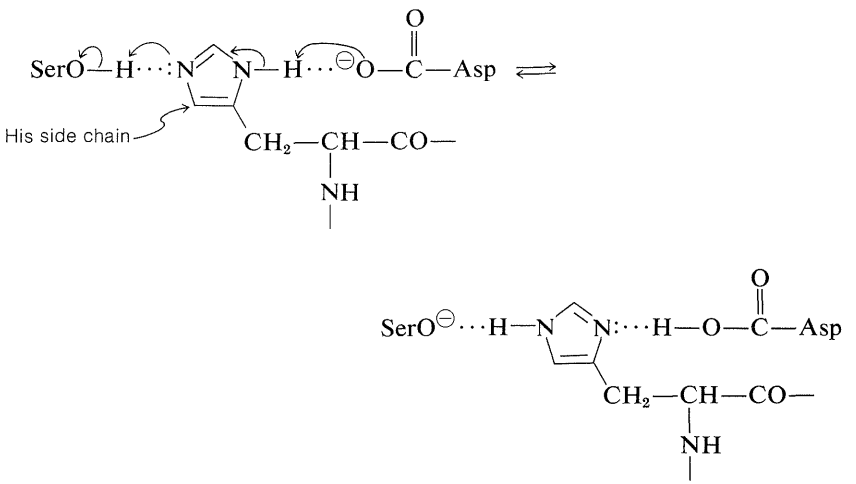
This raises another question. Why is the serine hydroxyl an effective nucleophile when water and other hydroxylic compounds clearly are not similarly effective? Apparently, the nucleophilicity of the serine $-CH_2OH$ is enhanced by acid-base catalysis involving proton transfers between acidic and basic side-chain functions in the vicinity of the active site. The serine is believed to transfer its OH proton to an amphoteric[10] site :B—A—H on the enzyme at the same instant that the proton of :B—A—H is transferred to another base $:B'^{\ominus}$ (Equation 25-8). These proton transfers are, of course, reversible:



(25-8)

Loss of its proton makes the serine hydroxyl oxygen a much more powerful nucleophile, and even though the equilibrium of Equation 25-8 must lie far to the left at physiological pH, it can increase greatly the reactivity of the serine hydroxyl.

In chymotrypsin and subtilisin, this **charge-relay** network system, as it is called, is made up of a specific aspartic acid residue, acting as $B'^{\ominus}$, and a specific histidine residue (acting as the amphoteric :B—A—H):



Exercise 25-29 Devise a way to use a stereospecific hydrolytic enzyme for resolution of D,L-alanine.

Exercise 25-30 The proteolytic enzyme, papain, differs from chymotrypsin in having cysteine, or a labile derivative thereof, as part of its active site. The enzyme is deactivated by substances that form complexes with, or react with, —SH groups and the activity is restored by reactions expected to regenerate an —SH group. Work out a schematic mechanism for cleavage of a peptide chain with papain that involves acylation of the critical —SH group of papain.

[10]Amphoteric means that a substance can act either as an acid or as a base.

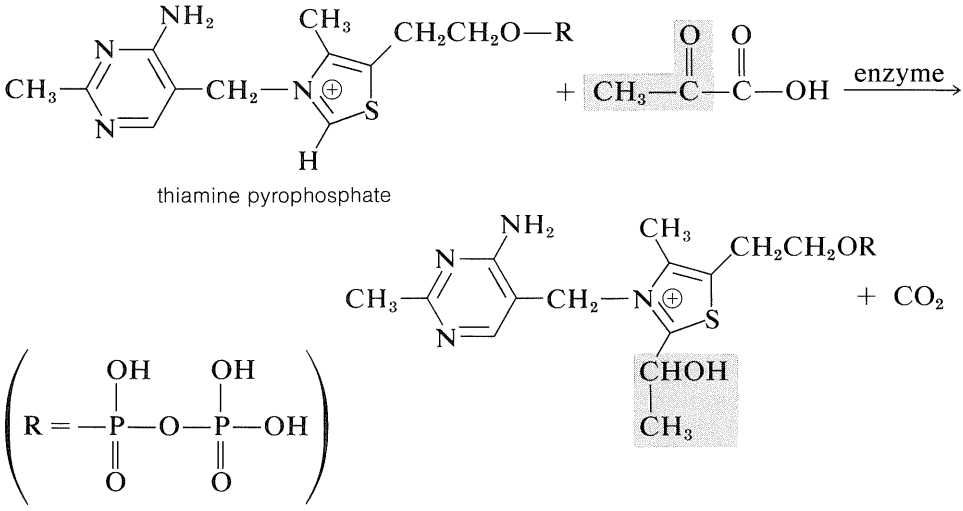
(One of the most interesting features of papain is that more than 100 of its total of 185 amino-acid residues may be removed with the aid of an aminopeptidase to give a fragment with considerable enzymatic activity.)

25-10 COENZYMES

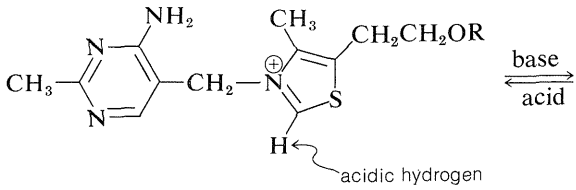
Many enzymes only operate in combination with organic molecules that are actually reagents for the reaction. These substances are called **coenzymes** or **cofactors.** Some coenzymes function with more than one enzyme and are involved in reactions with a number of different substrates.

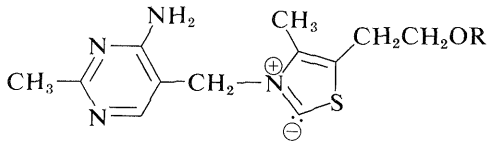
Several of the B vitamins function as coenzymes or as precursors of coenzymes; some of these have been mentioned previously. Nicotinamide adenine dinucleotide $(NAD)^{\oplus}$ which, in conjunction with the enzyme *alcohol dehydrogenase,* oxidizes ethanol to ethanal (Section 15-6C), also is the oxidant in the citric acid cycle (Section 20-10B). The precursor to $NAD^{\oplus}$ is the B vitamin, niacin or nicotinic acid (Section 23-2). Riboflavin (vitamin B_2) is a precursor of flavin adenine nucleotide FAD, a coenzyme in redox processes rather like $NAD^{\oplus}$ (Section 15-6C). Another example of a coenzyme is pyridoxal (vitamin B_6), mentioned in connection with the deamination and decarboxylation of amino acids (Section 25-5C). Yet another is coenzyme A (**CoASH**), which is essential for metabolism and biosynthesis (Sections 18-8F, 20-10B, and 30-5A).

An especially interesting coenzyme is thiamine pyrophosphate (vitamin B_1) which, in conjunction with the appropriate enzyme, decarboxylates 2-oxopropanoic acid (pyruvic acid; Section 20-10B). We can write the overall reaction as follows:



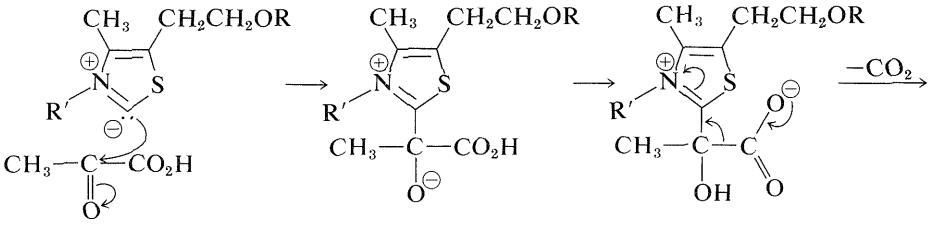
Although we do not know just how thiamine binds to the enzyme, the essential features of the reaction are quite well understood. Thiamine has an acidic hydrogen at the 2-position of the azathiacyclopentadiene ring, and you should recognize that the conjugate base, **10**, is both a nitrogen ylid and a sulfur ylid (Section 16-4A):

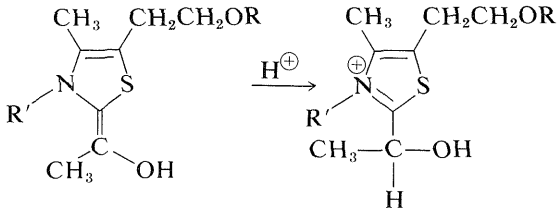




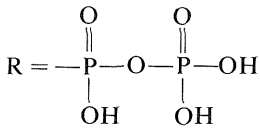
conjugate base, **10**

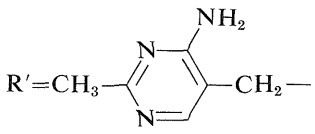
The acidity of the ring proton of the thiamine ring is a consequence of the adjacent positive nitrogen and the known ability of sulfur to stabilize an adjacent carbanion. Nucleophilic attack of the anionic carbon of **10** on C2 of 2-oxopropanoic acid is followed by decarboxylation:





11





The overall reaction introduces a two-carbon chain at the C2 position of the thiamine ring and the resulting modified coenzyme, **11**, functions in subsequent biological reactions as a carrier of a $CH_3-\underset{\displaystyle OH}{\underset{|}{CH}}-$ group and a potential source of a $CH_3-\overset{\displaystyle O}{\overset{\parallel}{C}}\diagdown$ group. The metabolism of glucose (Section 20-10) requires the conversion of pyruvate to ethanoyl **CoA** by way of **11**; and, in fermentation, the hydroxyethyl group of **11** is released as ethanal, which is reduced to ethanol by NADH (see Section 15-6C for discussion of the reverse reaction).

Thiamine pyrophosphate also plays a key role in the biosynthetic reactions that build (or degrade) pentoses from hexoses. We have mentioned these reactions previously in connection with the Calvin cycle (Section 20-9) and the pentose-phosphate pathway (Section 20-10C).

Exercise 25-31 Why would the intermediate addition product of thiamine to pyruvic acid be expected to decarboxylate readily? Support your answer by analogy; see Section 18-4.

Exercise 25-32 Write equations for a base-induced decomposition of the modified thiamine coenzyme, **11**, to ethanal and thiamine pyrophosphate.

25-11 ENZYME REGULATION

You may have wondered how the proteolytic enzymes such as trypsin, pepsin, chymotrypsin, carboxypeptidase, and others keep from self-destructing by catalyzing their own hydrolysis or by hydrolyzing each other. An interesting feature of the digestive enzymes is that they are produced in an inactive form in the stomach or the pancreas—presumably to protect the different kinds of proteolytic enzymes from attacking each other or other proteins.

The inactive precursors are called trypsinogen, pepsinogen, chymotrypsinogen, and procarboxypeptidase. These precursors are converted to the active enzymes by hydrolytic cleavage of a few specific peptide bonds under the influence of other enzymes (trypsin, for example, converts chymotrypsinogen to chymotrypsin). The digestive enzymes do not appear to self-destruct, probably because they are so constructed that it is sterically impossible to fit a part of one enzyme molecule into the active site of another. In this connection, it is significant that chymotrypsin attacks denatured proteins more rapidly than natural proteins with their compact structures of precisely folded chains.

Presumably all enzymes must have some regulatory mechanism that turns them on and off as needed. Less is known about regulation mechanisms than

about the enzymatic reactions themselves, but one type of control has been recognized. This occurs when a reaction product inhibits one of the reaction steps producing it by tying up the enzyme as a nonreactive complex (*feedback inhibition*). As the simplest example, suppose that the product (P) as well as the substrate (S) complexes with the enzyme (E); then we can write the following set of equilibria for the net reaction:

$$E + S \rightleftharpoons ES \longrightarrow E + P$$

$$E + P \rightleftharpoons EP$$

Clearly, a reaction of this type will decrease in rate as the product accumulates. It may stop altogether if the active sites are saturated with the product, and it will start again only on removal of the product.

25-12 ENZYME TECHNOLOGY

Because enzymes function nearly to perfection in living systems, there is great interest in how they might be harnessed to carry on desired reactions of practical value outside of living systems. The potential value in the use of enzymes (separate from the organisms that synthesize them) is undeniable, but how to realize this potential is another matter.

Practical use of separated enzymes is not new. Hydrolytic enzymes isolated from bacteria were widely used for a brief period to assist in removing food stains from clothing, but many people suffer allergic reactions to enzymes used in this way, and the practice was stopped. A major objective in enzyme technology is to develop an enzymatic process for the hydrolysis of cellulose to glucose (Section 20-7A). Some microorganisms do possess the requisite enzymes to catalyze the hydrolysis of the β-1,4 glucoside links in cellulose. If these enzymes could be harnessed for industrial production of glucose from cellulose, this could be an important supplementary food source. Technology already is available to convert glucose into ethanol and ethanoic acid, and from there to many chemicals now derived from petroleum.

A difficult problem in utilizing enzymes as catalysts for reactions in a noncellular environment is their instability. Most enzymes readily denature and become inactive on heating, exposure to air, or in organic solvents. An expensive catalyst that can be used only for one batch is not likely to be economical in an industrial process. Ideally, a catalyst, be it an enzyme or other, should be easily separable from the reaction mixtures and indefinitely reusable. A promising approach to the separation problem is to use the technique of **enzyme immobilization.** This means that the enzyme is modified by making it insoluble in the reaction medium. If the enzyme is insoluble and still able to manifest its catalytic activity, it can be separated from the reaction medium with minimum loss and reused. Immobilization can be achieved by linking the enzyme covalently to a polymer matrix in the same general manner as is used in solid-phase peptide synthesis (Section 25-7D).

Enzymes also have possible applications in organic synthesis. But there is another problem in addition to difficulties with enzyme stability. Enzymes that

achieve carbon–carbon bond formation, the **synthetases,** normally require cofactors such as ATP. How to supply ATP in a commercial process and regenerate it continuously from ADP or AMP is a technical problem that has to be solved if the synthetases are to be economically useful. This is a challenging field of biological engineering.

25-13 BIOSYNTHESIS OF PROTEINS

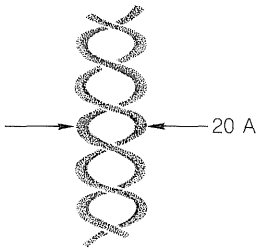
One of the most interesting and basic problems connected with the synthesis of proteins in living cells is how the component amino acids are induced to link together in the sequences that are specific for each type of protein. There also is the related problem of how the information as to the amino-acid sequences is perpetuated in each new generation of cells. We now know that the substances responsible for genetic control in plants and animals are present in and originate from the chromosomes of cell nuclei. Chemical analysis of the chromosomes has revealed them to be composed of giant molecules of deoxyribonucleoproteins, which are deoxyribonucleic acids (DNA) bonded to proteins. Since it is known that DNA rather than the protein component of a nucleoprotein contains the genetic information for the biosynthesis of enzymes and other proteins, we shall be interested mainly in DNA and will first discuss its structure. Part or perhaps all of a particular DNA is the chemical equivalent of the Mendelian gene—the unit of inheritance.

25-13A The Structure of DNA

The role of DNA in living cells is analogous to that of a punched tape used for controlling the operation of an automatic turret lathe—DNA supplies the information for the development of the cells, including synthesis of the necessary enzymes and such replicas of itself as are required for reproduction by cell division. Obviously, we would not expect the DNA of one kind of organism to be the same as DNA of another kind of organism. It is therefore impossible to be very specific about the structure of DNA without being specific about the organism from which it is derived. Nonetheless, the basic structural features of DNA are the same for many kinds of cells, and we mainly shall be concerned with these basic features in the following discussion.

In the first place, DNA molecules are quite large, sufficiently so to permit them to be seen individually in photographs taken with electron microscopes. The molecular weights vary considerably, but values of 1,000,000 to 4,000,000,000 are typical. X-ray diffraction indicates that DNA is made up of two long-chain molecules twisted around each other to form a double-stranded

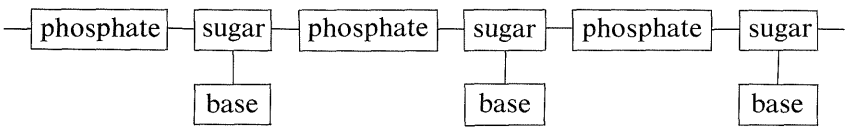
helix about 20 A in diameter. The arrangement is shown schematically in **12**:



12

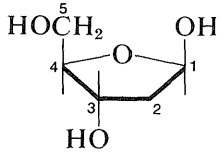
As we shall see, the components of the chains are such that the strands can be held together efficiently by hydrogen bonds. In agreement with this structure, it has been found that, when DNA is heated to about 80° under proper conditions, the strands of the helix unwind and dissociate into two randomly coiled fragments. Furthermore, when the dissociated material is allowed to cool slowly under the proper conditions, the fragments recombine and regenerate the helical structure.

Chemical studies show that the strands of DNA have the structure of a long-chain polymer made of alternating phosphate and sugar residues carrying nitrogen bases, **13**:



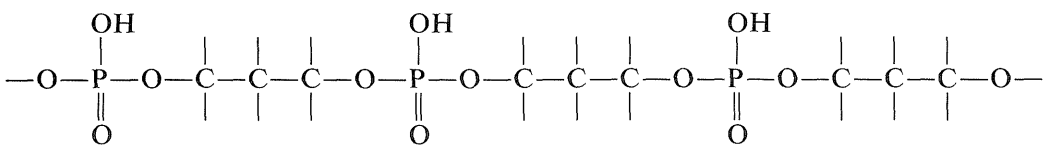
13

The sugar is D-2-deoxyribofuranose, **14**, and each sugar residue is bonded to two phosphate groups by way of ester links involving the 3- and 5-hydroxyl groups:



D-2-deoxyribofuranose, **14**

The backbone of DNA is thus a *polyphosphate ester of a 1,3-diol:*



With inclusion of the details of the sugar residue, the structure of DNA becomes as shown in Figure 25-21.

Each of the sugar residues of DNA is bonded at the 1-position to one of four bases: cytosine, **15**; thymine, **16**; adenine, **17**; and guanine, **18**. The four

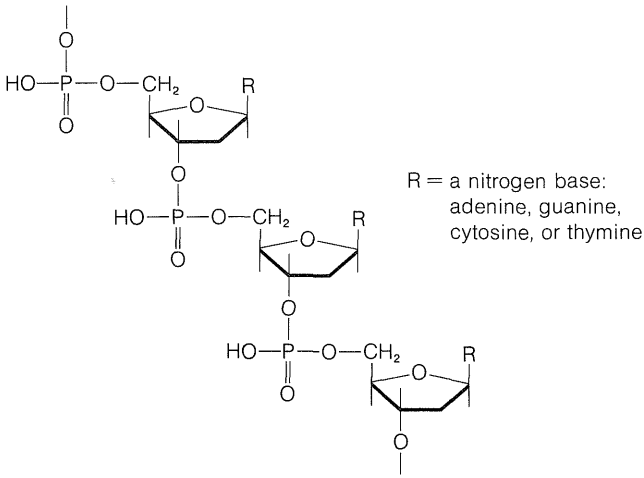
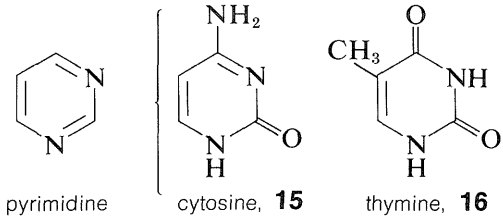
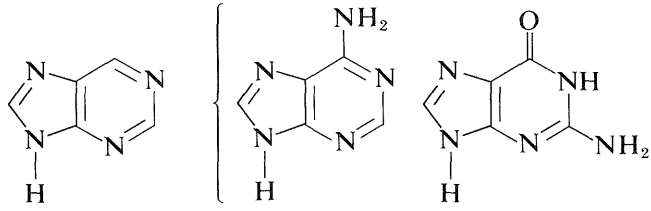


Figure 25-21 Structure of the strands of deoxyribonucleic acid (DNA)

bases are derivatives of either *pyrimidine* or *purine,* both of which are heterocyclic nitrogen bases:

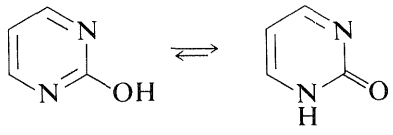


pyrimidine cytosine, **15** thymine, **16**



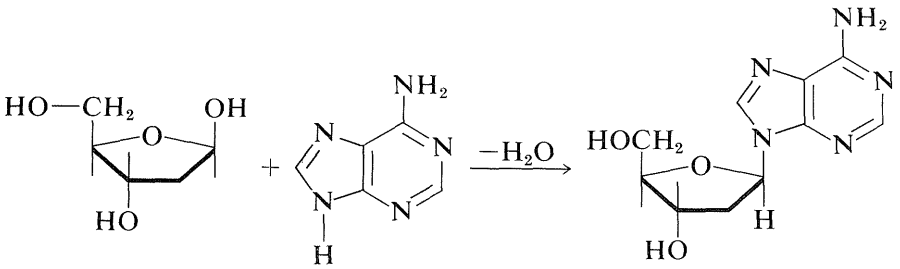
purine adenine, **17** guanine, **18**

Unlike phenols (Section 26-1), structural analysis of many of the hydroxy-substituted aza-aromatic compounds is complicated by isomerism of the keto-enol type, sometimes called **lactim-lactam** isomerism. For 2-hydroxypyrimidine, **19**, these isomers are **19a** and **19b**, and the lactam form is more stable, as also is true for cytosine, **15**, thymine, **16**, and the pyrimidine ring of guanine, **18**.

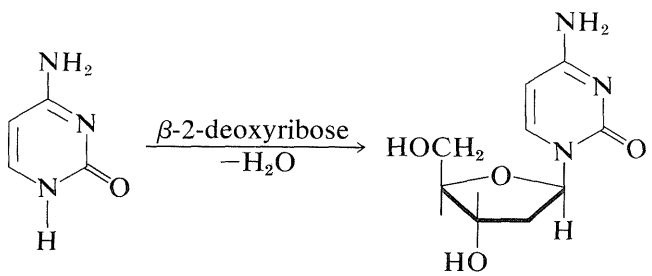


lactim, **19a** lactam, **19b** 2-hydroxypyrimidine, **19**

For the sake of simplicity in illustrating *N*-glycoside formation in DNA, we shall show the type of bonding involved for the sugar and base components only (i.e., the deoxyribose nucleoside structure). Attachment of 2-deoxyribose is through a NH group to form the β-*N*-deoxyribofuranoside (Section 20-5):

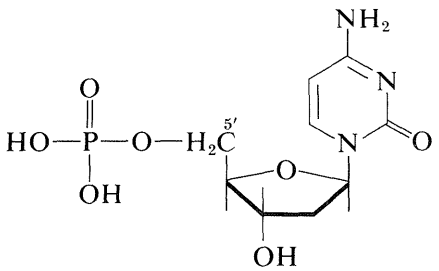


β-2-deoxyribofuranose adenine adenine β-2-deoxyriboside



cytosine cytosine β-2-deoxyriboside

Esterification of the 5′-hydroxyl group of deoxyribose *nucleosides*, such as cytosine deoxyriboside, with phosphoric acid gives the corresponding *nucleotides*:[11]

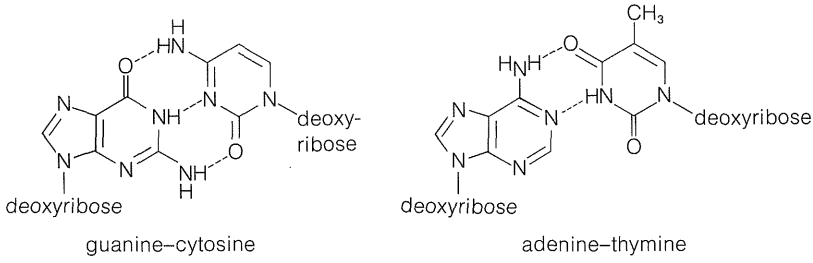


cytosine deoxyribonucleotide

Thus DNA may be considered to be built up from nucleotide monomers by esterification of the 3′-hydroxyl group of one nucleotide with the phosphate group of another (Figure 25-21).

The number of nucleotide units in a DNA chain varies from about 3,000 to 10,000,000. Although the sequence of the purine and pyrimidine bases in the chains are not known, there is a striking equivalence between the numbers of certain of the bases regardless of the origin of DNA. Thus the number of adenine (A) groups equals the number of thymine (T) groups, and the number of guanine (G) groups equals the number of cytosine (C) groups: A = T and

[11]The positions on the sugar ring are primed to differentiate them from the positions of the nitrogen base.



guanine–cytosine adenine–thymine

Figure 25-22 Hydrogen bonding postulated between DNA strands involving guanine–cytosine and adenine–thymine. In each case, the distance between the C1 of the two deoxyribose units is 11 A and the favored geometry has the rings coplanar.

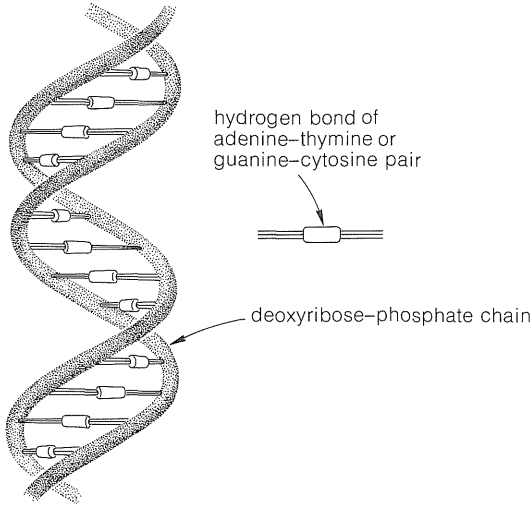


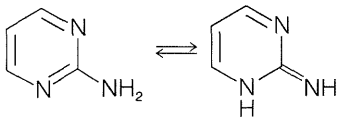
Figure 25-23 Schematic representation of configuration of DNA, showing the relationship between the axes of hydrogen-bonded purine and pyrimidine bases and the deoxyribosephosphate strands. There are 10 pairs of bases per complete 360° twist of the chain. The spacing between the strands is such that there is a wide and a narrow helical "groove" around the molecule. Proteins known as histones coordinate with DNA by winding around the helix, filling one of the other of the grooves. The histone–DNA combination is important in regulating the action of DNA.

G = C. The bases of DNA therefore are half purines and half pyrimidines. Furthermore, although the ratios of A to G and T to C are constant for a given species, they vary widely from one species to another.

The equivalence between the purine and pyrimidine bases in DNA was accounted for by J. D. Watson and F. Crick (1953) through the suggestion that the two strands are constructed so that, when twisted together in the helical structure, hydrogen bonds are formed involving adenine in one chain and thymine in the other, or cytosine in one chain and guanine in the other. Thus each adenine occurs paired with a thymine and each cytosine with a guanine and the strands are said to have complementary structures. The postulated hydrogen bonds are shown in Figure 25-22, and the relationship of the bases to the strands in Figure 25-23.

Exercise 25-33 a. The lactim-lactam equilibrium of 2-hydroxypyrimidine lies on the side of the lactam, yet the benzenol-cyclohexadienone equilibrium lies far on the side of benzenol (Section 26-1). Explain what factors make for the large difference in the positions of these two equilibria. (Bond energies and review of Section 24-1A may help in showing you the differences between these systems.)

b. Use bond energies to decide whether the following equilibrium with 2-amino-pyrimidine is likely to be more, or less, favorably to the right than the corresponding equilibrium **19a** $\rightleftharpoons$ **19b**. Give your reasoning.



c. Show how the resonance method could be used to predict whether cytosine or 2-hydroxypyrimidine would have the greater tendency to be more stable in the lactam rather than the lactim form.

Exercise 25-34 Write equations for the mechanistic steps involved in hydrolysis of adenine deoxyribonucleoside to deoxyribose and adenine. Would you expect the reaction to occur more readily in acidic, basic, or neutral solution? (Review Section 16-4C).

Exercise 25-35 The following steps have been used in the synthesis of 1-D-glucosylcytosine:

$$\text{D-glucose} \xrightarrow[(CH_3CO)_2O]{\text{HBr}} \text{1-bromoglucose tetraethanoate}$$

$$\xrightarrow{\text{2,4-diethoxypyrimidine}} \text{4-ethoxy-1-(tetraethanoyl-D-glucosyl)-2-pyrimidone}$$

$$\xrightarrow{NH_3\ (\text{excess})} \text{1-D-glucosylcytosine}$$

Write the structures of the various substances given, and as detailed a mechanism as you can for the reaction of the bromo compound with 2,4-diethoxypyrimidine. Would you expect the reaction of 1-bromoglucose tetraethanoate with 2,4-diethoxypyrimidine to yield significant amounts of 6-ethoxy-1-(tetraethanoyl-D-glucosyl)-2-pyrimidone? Give your reasoning.

25-13B Genetic Control and the Replication of DNA

It is now well established that DNA provides the genetic recipe that determines how cells reproduce. In the process of cell division, the DNA itself

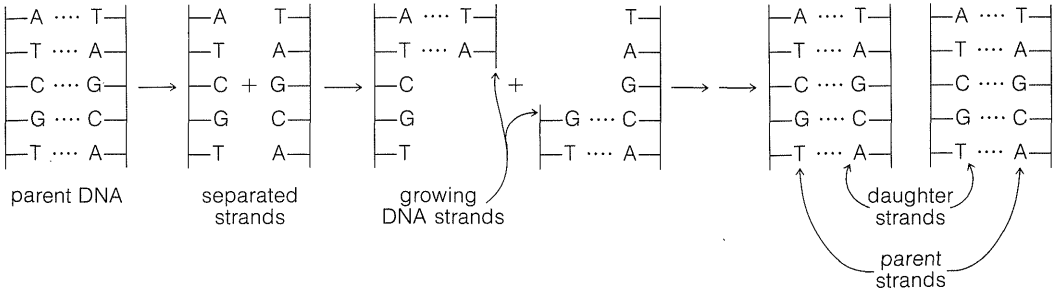


Figure 25-24 Representation of DNA replication

also is reproduced and thus perpetuates the information necessary to regulate the synthesis of specific enzymes and other proteins of the cell structure. In replicating itself prior to cell division, the DNA double helix evidently separates at least partly into two strands (see Figure 25-24). Each of the separated parts serves as a guide (template) for the assembly of a complementary sequence of nucleotides along its length. Ultimately, *two* new DNA double strands are formed, each of which contains one strand from the parent DNA.

The genetic information inherent in DNA depends on the arrangement of the bases (A, T, G, and C) along the phosphate-carbohydrate backbone—that is, on the arrangement of the four nucleotides specific to DNA. Thus the sequence A–G–C at a particular point conveys a different message than the sequence G–A–C.

It is quite certain that the code involves a particular sequence of *three* nucleotides for each amino acid. Thus the sequence A–A–A codes for lysine, and U–C–G codes for serine. The sequences or *codons* for all twenty amino acids are known.

25-13C Role of RNA in Synthesis of Proteins

It is clear that DNA does not play a direct role in the synthesis of proteins and enzymes because most of the protein synthesis takes place outside of the cell nucleus in the cellular cytoplasm, which does not contain DNA. Furthermore, it has been shown that protein synthesis can occur in the absence of a cell nucleus or, equally, in the absence of DNA. Therefore the genetic code in DNA must be passed on selectively to other substances that carry information from the nucleus to the sites of protein synthesis in the cytoplasm. These other substances are **ribonucleic acids** (RNA), which are polymeric molecules similar in structure to DNA, except that D-2-deoxyribofuranose is replaced by

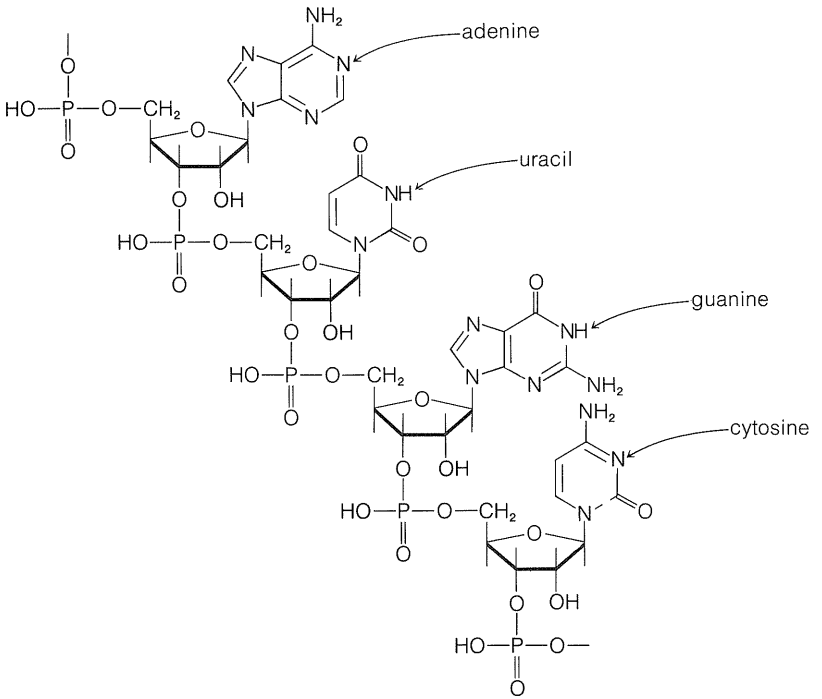
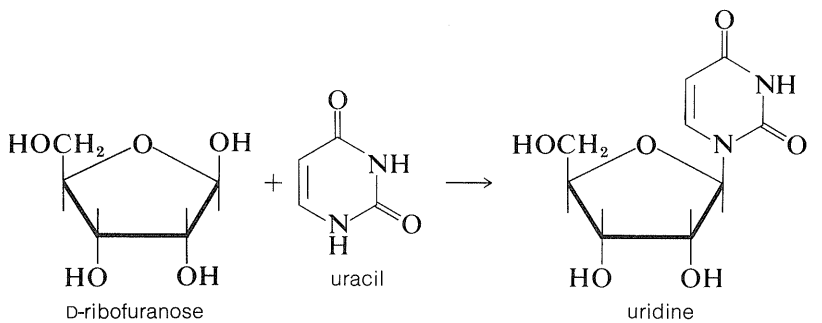


Figure 25-25 Structure of a ribonucleic acid (RNA) chain with the base sequence: adenine, uracil, guanine, cytosine

D-ribofuranose and the base thymine is replaced by uracil, as shown in Figure 25-25.



RNA also differs from DNA in that there are not the same regularities in the overall composition of its bases and it usually consists of a single polynucleotide chain. There are different types of RNA, which fulfill different functions. About 80% of the RNA in a cell is located in the cytoplasm in clusters closely associated with proteins. These ribonucleoprotein particles specifically are called ribosomes, and the ribosomes are the sites of most of the protein synthesis in the cell. In addition to the **ribosomal RNA** (rRNA),

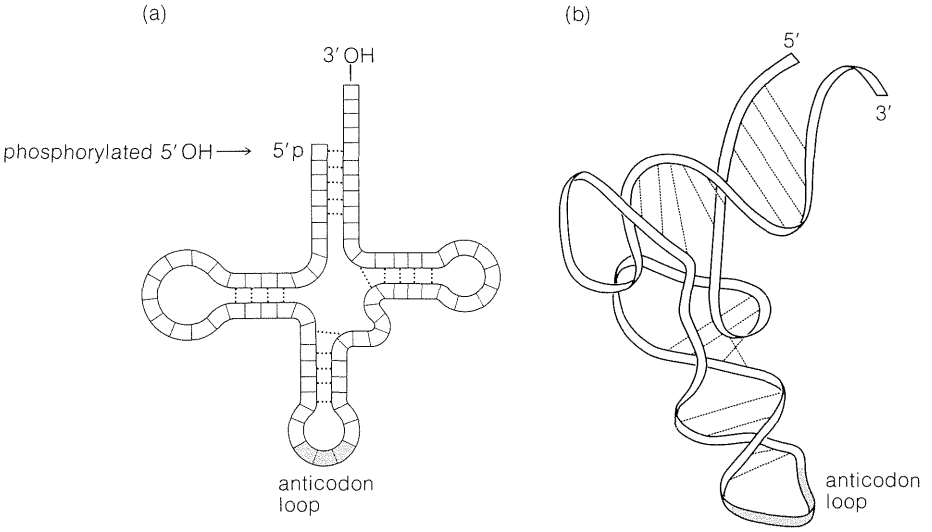


Figure 25-26 (a) Generalized representation of a tRNA molecule. Each segment represents a nucleotide; the actual number and sequence of nucleotides varies with the tRNA. There are regions of intrachain base-pairing (dashed lines). The nucleotide at the long end has a ribose with a free 3′-OH. The nucleotide at the short end is phosphorylated at 5′-OH. The three nucleotides of the anticodon loop pair with the appropriate bases in mRNA. (b) Three-dimensional picture of a tRNA to show the manner in which the chain is coiled. An excellent review article on the determination of the structure of phenylalanine tRNA by x-ray diffraction has been published, J. L. Sussman and S.-H. Kim, *Science* **192,** 853 (1976).

there are ribonucleic acids called **messenger RNA** (mRNA), which convey instructions as to what protein to make. In addition, there are ribonucleic acids called **transfer RNA** (tRNA), which actually guide the amino acids into place in protein synthesis. Much is now known about the structure and function of tRNA.

The principal structural features of tRNA molecules are shown schematically in Figure 25-26. Some of the important characteristics of tRNA molecules are summarized as follows.

1. There is at least one particular tRNA for each amino acid.
2. The tRNA molecules have single chains with between 73–93 ribonucleotides. Most of the tRNA bases are adenine (A), cytosine (C), guanine (G), and uracil (U). There also are a number of unusual bases that are methylated derivatives of A, C, G, and U.
3. The clover-leaf pattern of Figure 25-26 shows the general structure of tRNA. There are regions of the chain where the bases are complementary to one another, which causes it to fold into two double-helical regions. The chain has three bends or *loops* separating the helical regions.

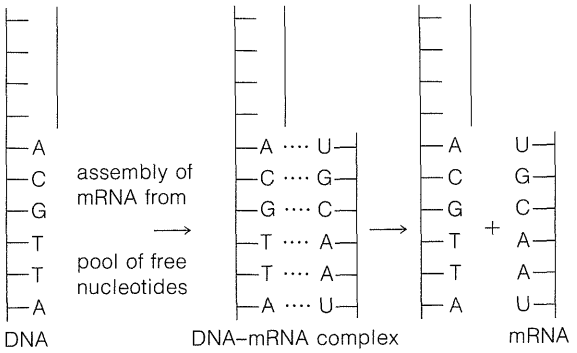
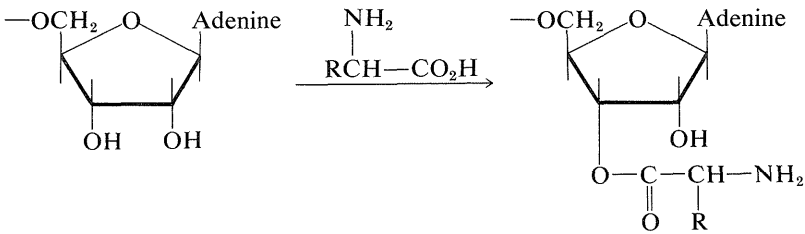


Figure 25-27 Representation of transmittal of information from DNA to RNA

4. The 5′-terminal residue usually is a guanine nucleotide; it is phosphorylated at the 5′-OH. The terminus at the 3′ end has the same sequence of three nucleotides in all tRNA's, namely, CCA. The 3′-OH of the adenosine in this grouping is the point of attachment of the tRNA to its specific amino acid:



(25-9)

5. The middle loop (Figure 25-26a) contains a different sequence of bases for each tRNA and it is this sequence that is recognized by the protein-synthesizing apparatus of the ribosomes. The loop is called the *anticodon loop.*

With this information on the structure of tRNA, we can proceed to a discussion of the essential features of biochemical protein synthesis.

The information that determines amino-acid sequence in a protein to be synthesized is contained in the DNA of a cell nucleus as a particular sequence of nucleotides derived from adenine, guanine, thymine, and cytosine. For each particular amino acid there is a sequence of *three* nucleotides called a **codon.**

The information on protein structure is transmitted from the DNA in the cell nucleus to the cytoplasm where the protein is assembled by messenger RNA. This messenger RNA, or at least part of it, is assembled in the nucleus with a base sequence that is complementary to the base sequence in the parent DNA. The assembly mechanism is similar to DNA replication except that thymine (T) is replaced by uracil (U). The uracil is complementary to adenine in the DNA chain. (See Figure 25-27.) After the mRNA is assembled, it is transported to the cytoplasm where it becomes attached to the ribosomes.

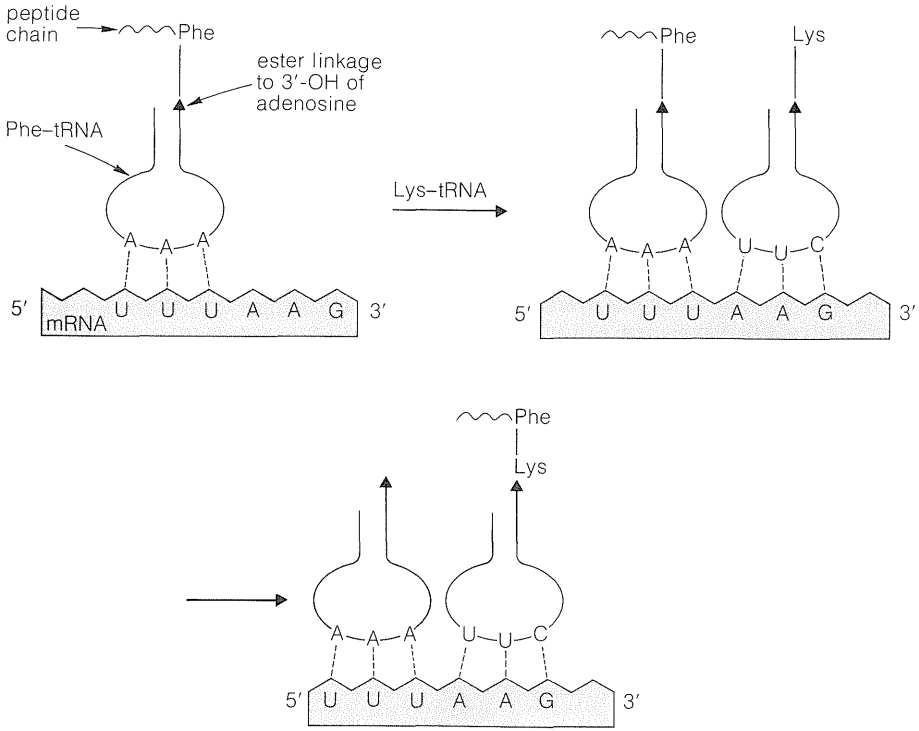
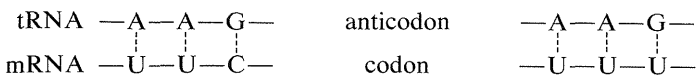


Figure 25-28 Peptide-bond formation in protein biosynthesis showing how the amino-acid sequence is determined by complementary base-pairing between messenger RNA and transfer RNA. The peptide chain is bound to tRNA, which is associated with mRNA through three bases in mRNA (codon) and three bases in tRNA (anticodon). In the diagram, the next codon A–A–G codes for lysine. Hence, Lys–tRNA associates with mRNA by codon-anticodon base-pairing and, under enzyme control, couples to the end of the peptide chain.

The amino acids in the cytoplasm will not form polypeptides unless activated by ester formation with appropriate tRNA molecules. The ester linkages are through the 3′-OH of the terminal adenosine nucleotide (Equation 25-9) and are formed only under the influence of a synthetase enzyme that is specific for the particular amino acid. The energy for ester formation comes from ATP hydrolysis (Sections 15-5F and 20-10). The product is called an **aminoacyl-tRNA.**

The aminoacyl-tRNA's form polypeptide chains in the order specified by codons of the mRNA bound to the ribosomes (see Figure 25-28). The order of incorporation of the amino acids depends on the recognition of a codon in mRNA by the corresponding anticodon in tRNA by a complementary base-pairing of the type A · · · U and C · · · G. The first two bases of the codon recognize only their complementary bases in the anticodon, but there is some

flexibility in the identity of the third base. Thus phenylalanine tRNA has the anticodon A–A–G and responds to the codons U–U–C and U–U–U, but not U–U–A or U–U–G:



The codons of the mRNA on the ribosomes are read from the 5′ to the 3′ end. Thus the synthetic polynucleotide (5′)A–A–A–(A–A–A)$_n$–A–A–C(3′) contains the code for lysine (A–A–A) and asparagine (A–A–C); the actual polypeptide obtained using this mRNA in a cell-free system was Lys–(Lys)$_n$–Asn, and not Asn–(Lys)$_n$–Lys.

The start of protein synthesis is signalled by specific codon-anticodon interactions. Termination is also signalled by a codon in the mRNA, although the stop signal is not recognized by tRNA, but by proteins that then trigger the hydrolysis of the completed polypeptide chain from the tRNA. Just how the secondary and tertiary structures of the proteins are achieved is not yet clear, but certainly the mechanism of protein synthesis, which we have outlined here, requires little modification to account for preferential formation of particular conformations.

25-14 CHEMICAL EVOLUTION

A problem of great interest to those curious about the evolution of life concerns the origins of biological molecules. When and how were the molecules of life, such as proteins, nucleic acids, and polysaccharides, first synthesized?

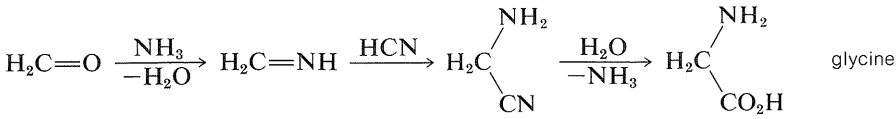
In the course of geological history, there must have been a prebiotic period when organic compounds were formed and converted to complex molecules similar to those we encounter in living systems. The composition of the earth's atmosphere in prebiotic times was almost certainly very different from what it is today. Probably it was a reducing atmosphere consisting primarily of methane, ammonia, and water, and because there was little or no free oxygen, there was no stratospheric ozone layer and little, if any, screening from the sun's ultraviolet radiation. Starting with CH_4, NH_3, and H_2O, it is plausible that photochemical processes would result in formation of hydrogen cyanide, HCN, and methanal, CH_2O, by reactions such as the following:

$$CH_4 + NH_3 \xrightarrow{h\nu} HCN + 3H_2$$

$$CH_4 + H_2O \xrightarrow{h\nu} CH_2{=}O + 2H_2$$

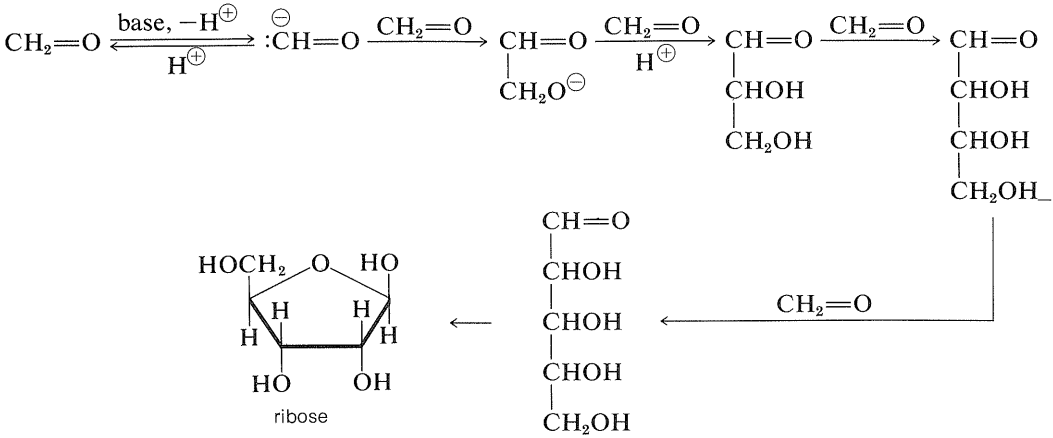
Hydrogen cyanide and methanal are especially reasonable starting materials for the prebiotic synthesis of amino acids, purine and pyrimidine bases, ribose and other sugars. Formation of glycine, for example, could have occurred by a Strecker synthesis (Section 25-6), whereby ammonia adds to methanal in the

presence of HCN. Subsequent hydrolysis of the intermediate aminoethanenitrile would produce glycine:

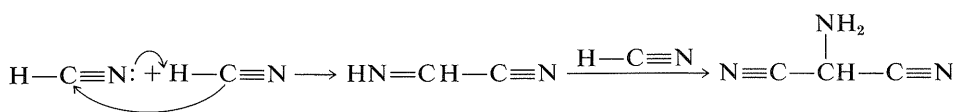


The plausibility of these reactions is strongly supported by the classic experiments of S. Miller (1953), who showed that a mixture of methane or ethane, ammonia, and water, on prolonged ultraviolet irradiation or exposure to an electric discharge, produced a wide range of compounds including racemic α- and β-amino acids.

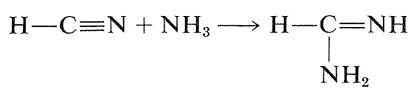
It is not difficult to visualize how sugars such as ribose may be formed. Methanal is known to be converted by bases through a series of aldol-type additions to a mixture of sugarlike molecules called "formose." Formation of racemic ribose along with its stereoisomers could occur as follows:



It is more difficult to conceive how the nucleic acid bases such as adenosine may be achieved in prebiotic syntheses. However, adenine, $C_5H_5N_5$, corresponds in composition to a pentamer of hydrogen cyanide and could result by way of a trimer of HCN, **20**, and the adduct of ammonia with HCN, **21**:

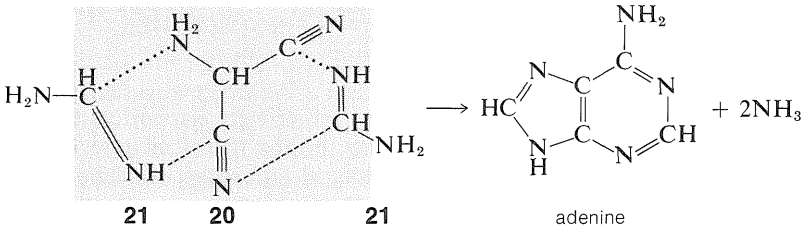


20



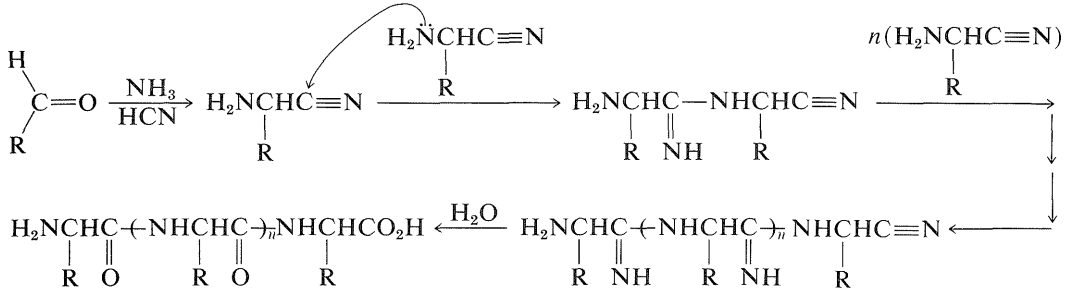
21

Without worrying about the mechanistic details of how adenine could be formed from **20** and **21**, you can see that the adenine ring system is equivalent to **20** plus two **21** minus two NH_3:



How did these small prebiotic organic molecules grow into large polymeric substances such as peptides, RNA, and so on? It is important to recognize that by whatever reactions polymerization occurred, they had to be reactions that would occur in an essentially aqueous environment. This presents difficulties because condensation of amino acids to form peptides, or of nucleotides to form RNA or DNA, is not thermodynamically favorable in aqueous solution.

It is quite possible that primitive polypeptides were formed by the polymerization of aminoethanenitriles produced by the addition of ammonia and HCN to methanal or other aldehydes. The resulting imino polymers certainly would hydrolyze to polypeptides:



There are many other questions regarding the origin of the molecules of life for which we have only partial answers, or no answers at all.

It is difficult to imagine just how these molecules, once formed, somehow evolved further into the extraordinarily complex systems afforded by even the simplest bacterium able to utilize energy from the sun to support and reproduce itself. Nonetheless, synthetic peptides do coil and aggregate rather like natural proteins, and some also have shown catalytic activities characteristic of natural enzymes. One would hope that some kind of life would be found elsewhere in the solar system, the analysis of which would help us to better understand how life began on earth.

Additional Reading

K. D. Kopple, *Peptides and Amino Acids,* W. A. Benjamin, Inc., Menlo Park, Calif., 1966.

W. A. Schroeder, *The Primary Structures of Proteins,* Harper and Row, New York, 1968.

M. Calvin and W. A. Pryor, Eds., *Organic Chemistry of Life,* W. H. Freeman and Co., San Francisco, 1973. This volume contains a selected group of readings from *Scientific American.*

R. E. Dickerson and I. Geis, *The Structure and Action of Proteins,* W. A. Benjamin, Inc., Menlo Park, Calif., 1969.

M. Goodman and F. Morehouse, *Organic Molecules in Action,* Gordon and Breach, New York, 1973. See, for example, the chapters on prebiotic synthesis, origin of the cell, nucleic acids, and protein synthesis.

M. Calvin, *American Scientist* **63,** 169 (1975). This is an article on chemical evolution with many key references.

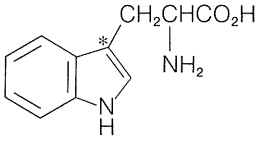
S. Fox, K. Harada, G. Krampitz, and G. Mueller, "Chemical Origins of Cells," *Chemical and Engineering News,* June 1970, p. 80.

Supplementary Exercises

25-36 The enzyme, *acetoacetate decarboxylase,* converts 3-oxobutanoic (acetoacetic) acid to 2-propanone and carbon dioxide. Investigation of the nature of the catalytically active site has been carried on by F. Westheimer and his coworkers with the following results. First, 3-oxobutanoic acid labeled with ^{18}O at the ketone group and decarboxylated in ordinary water in the presence of the enzyme gives 2-propanone containing no ^{18}O. Second, the enzyme-substrate complex combines with hydrogen cyanide and decarboxylation stops. However, if a solution of the hydrogen cyanide–deactivated enzyme–substrate complex is **dialyzed** (i.e., placed in a cellophane bag immersed in flowing water to permit separation of low-molecular-weight water-soluble materials from the enzyme by diffusion through the cellophane; peptides and proteins having molecular weights greater than 6000 to 10,000 do not diffuse through cellophane), then the enzyme is recovered fully active. Third, the mild reducing agent sodium borohydride, which reacts with C=N but not C—N or amide carbonyl groups, reduces the enzyme-substrate complex made from ^{14}C-labeled 3-oxobutanoic acid to give a product that is not enzymatically active and that retains essentially all the ^{14}C on dialysis without regenerating the enzyme. Sodium borohydride treatment of the enzyme–substrate–hydrogen cyanide complex followed by dialysis regenerates fully active enzyme. Borohydride reduction of the enzyme-substrate complex, prepared from 3-oxobutanoic acid labeled at the 2- and 4-positions with ^{14}C, followed by complete hydrolysis, gives 1 mole of *N*-(2-propyl-^{14}C-amino)-2-aminohexanoic acid.

Write a stepwise mechanism for the enzyme-induced decarboxylation, clearly indicating the nature of the bonding between the substrate and enzyme. Show how your mechanism can accommodate hydrogen cyanide inhibition and the results of the borohydride reactions. Utilize the results of the discussion of the ease of decarboxylation of various acids in Section 18-4 to deduce possible structural requirements for the active site so that decarboxylation of the enzyme-substrate complex can occur *more readily* than the uncatalyzed decarboxylation.

25-37 Figure 25-29 shows an unusually well-resolved ^{13}C nmr spectrum of the enzyme lysozyme (Table 25-3 and Figure 25-15) taken with proton decoupling. The closely spaced peaks on the left side of the spectrum are of the carbonyl groups. The peaks in the center are of unsaturated and aromatic carbons, while those on the right are of the aliphatic amino acid carbons. The five sharp resonances marked at about 110 ppm with * arise from tryptophan carbons marked with * in **22**:



22

a. How many tryptophan residues does the ^{13}C spectrum indicate to be present in lysozyme?

b. Lysozyme contains S–S bonds, and when these S–S bonds are cleaved by reduction, the resonances marked * in Figure 25-29 have much *smaller* chemical-shift differences. Explain why this might be so.

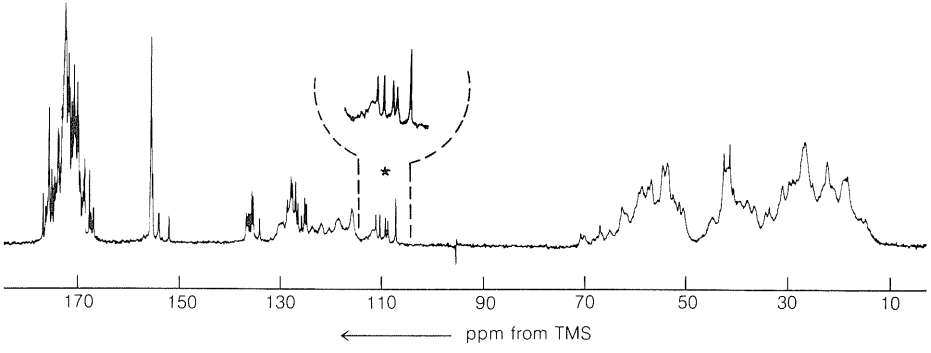


Figure 25-29 Carbon-13 nmr spectrum at 45.3 MHz of lysozyme, 0.015*M* in water solution, taken with proton decoupling (Section 9-10L)