Chapter 31 Heterocyclic Compounds

31.1 Heterocyclic systems

A **heterocyclic compound** is one that contains a ring made up of more than one kind of atom.

In most of the cyclic compounds that we have studied so far—benzene, naphthalene, cyclohexanol, cyclopentadiene—the rings are made up only of carbon atoms; such compounds are called *homocyclic* or *alicyclic* compounds. But there are also rings containing, in addition to carbon, other kinds of atoms, most commonly nitrogen, oxygen, or sulfur. For example:


Pyrrole


Furan


Thiophene


Imidazole


Oxazole


Thiazole

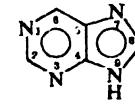

Pyrazole

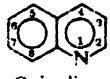

3-Pyrroline

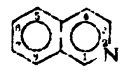

Pyrrolidine

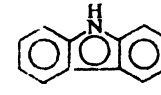

Pyridine


Pyrimidine


Purine


Quinoline


Isoquinoline


Carbazole

We notice that, in the numbering of ring positions, hetero atoms are generally given the lowest possible numbers.

Table 31.1 HETEROCYCLIC COMPOUNDS

Name	M.p., °C	B.p., °C	Name	M.p., °C	B.p., °C
Furan	− 30	32	Pyridine	− 42	115
Tetrahydrofuran	−108	66	α-Picoline	− 64	128
Furfuryl alcohol		171	β-Picoline		143
Furfural	− 36	162	γ-Picoline		144
Furoic acid	134		Piperidine	− 9	106
Pyrrole		130	Picolinic acid	137	
Pyrrolidine		88	Nicotinic acid	237	
Thiophene	− 40	84	Isonicotinic acid	317	
			Indole	53	254
			Quinoline	− 19	238
			Isoquinoline	23	243

Actually, of course, we have already encountered numerous heterocyclic compounds: *cyclic anhydrides* (Sec. 20.9) and *cyclic imides* (Sec. 20.14), for example: *lactones* (Sec. 20.15) and *lactams* (Problem 28.3, p. 891); *cyclic acetals* of dihydroxy alcohols (Problem 23, p. 651); the solvents *dioxane* and *tetrahydrofuran* (Sec. 17.9). In all these, the chemistry is essentially that of their open-chain analogs.

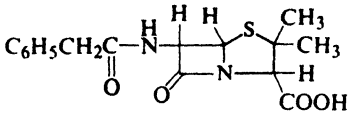
We have encountered three-membered heterocyclic rings which, because of ring strain, are highly reactive: *epoxides* (Secs. 17.10–17.15) and *aziridines* (Sec. 22.6); the fleeting but important intermediates, cyclic *halonium ions* (Secs. 7.12 and 28.10) and cyclic *sulfonium ions* (Sec. 28.11).

Heterocyclic intermediates are being used more and more in synthesis as *protecting groups*, readily generated and, when their job is done, readily removed. We have seen two examples of this: the temporary incorporation of the carboxyl group into a *2-oxazoline* ring (Sec. 26.6), and the temporary formation of *tetrahydropyranyl* (*THP*) *esters*, resistant toward alkali but extremely easily cleaved by acid (Problem 16, p. 692).

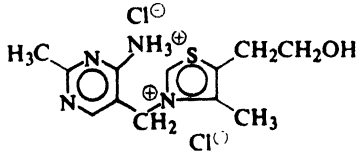
In the biological world, as we shall see in the final chapters of this book, heterocyclic compounds are everywhere. Carbohydrates are heterocyclic; so are chlorophyll and hemin, which make leaves green and blood red and bring life to plants and animals. Heterocycles form the sites of reaction in many enzymes and coenzymes. Heredity comes down, ultimately, to the particular sequence of attachment of a half-dozen heterocyclic rings to the long chains of nucleic acids.

In this chapter we can take up only a very few of the many different heterocyclic systems, and look only briefly at them. Among the most important and most interesting heterocycles are the ones that possess aromatic properties; we shall focus our attention on a few of these, and in particular upon their aromatic properties.

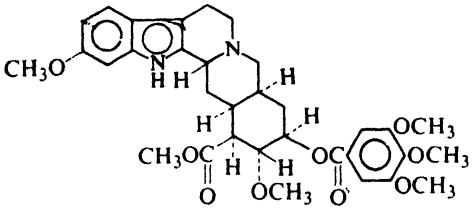
We can get some idea of the importance—as well as complexity—of heterocyclic systems from the following examples. Some others are *hemin* (p. 1152), *nicotinamide adenine dinucleotide* (p. 1153), and *oxytocin* (p. 1143).



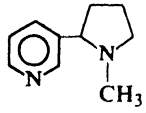
Penicillin G
Antibiotic



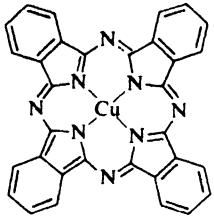
Thiamine
Vitamin B_1
Anti-beriberi factor



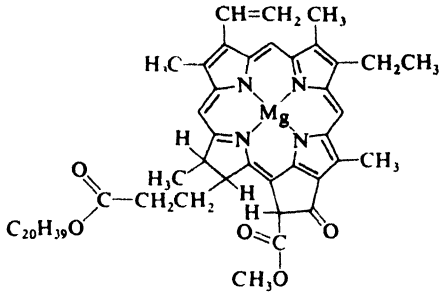
Reserpine
A tranquilizing drug



Nicotine
A tobacco alkaloid



Copper phthalocyanine
A blue pigment



Chlorophyll a
Green plant pigment: catalyst for photosynthesis

FIVE-MEMBERED RINGS

31.2 Structure of pyrrole, furan, and thiophene

The simplest of the five-membered heterocyclic compounds are **pyrrole, furan,** and **thiophene**, each of which contains a single hetero atom.

Judging from the commonly used structures I, II, and III, we might expect each of these compounds to have the properties of a conjugated diene and of an amine, an ether, or a sulfide (thioether). Except for a certain tendency to undergo addi-

I
Pyrrole

II
Furan

III
Thiophene

tion reactions, however, these heterocycles do not have the expected properties: thiophene does not undergo the oxidation typical of a sulfide, for example; pyrrole does not possess the basic properties typical of amines.

Instead, these heterocycles and their derivatives most commonly undergo electrophilic substitution: nitration, sulfonation, halogenation, Friedel-Crafts acylation, even the Reimer-Tiemann reaction and coupling with diazonium salts. Heats of combustion indicate resonance stabilization to the extent of 22–28 kcal/mole; somewhat less than the resonance energy of benzene (36 kcal/mole), but much greater than that of most conjugated dienes (about 3 kcal/mole). On the basis of these properties, pyrrole, furan, and thiophene must be considered *aromatic*. Clearly, formulas I, II, and III do not adequately represent the structures of these compounds.

Let us look at the orbital picture of one of these molecules, pyrrole. Each atom of the ring, whether carbon or nitrogen, is held by a σ bond to three other atoms. In forming these bonds, the atom uses three sp^2 orbitals, which lie in a plane and are 120° apart. After contributing one electron to each σ bond, each carbon atom of the ring has left *one* electron and the nitrogen atom has left *two* electrons; these electrons occupy p orbitals. Overlap of the p orbitals gives rise to π clouds, one above and one below the plane of the ring; the π clouds contain a total of six electrons, the *aromatic sextet* (Fig. 31.1).

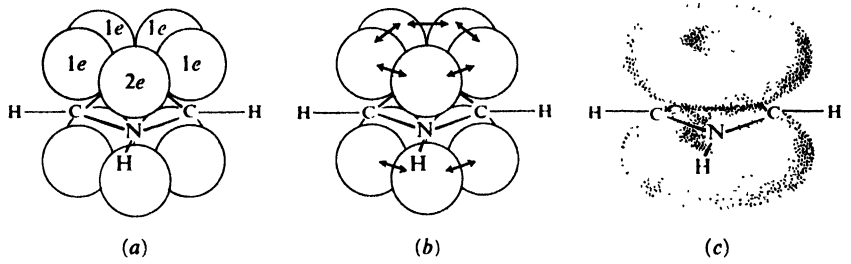
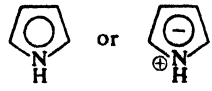


Figure 31.1. Pyrrole molecule. (*a*) Two electrons in p orbital of nitrogen; one electron in p orbital of each carbon. (*b*) Overlap of p orbitals to form π bonds. (*c*) Clouds above and below plane of ring; total of six π electrons, the aromatic sextet.

Delocalization of the π electrons stabilizes the ring. As a result, pyrrole has an abnormally low heat of combustion; it tends to undergo reactions in which the stabilized ring is retained, that is, to undergo substitution.

Nitrogen's extra pair of electrons, which is responsible for the usual basicity of nitrogen compounds, is involved in the π cloud, and is not available for sharing with acids. In contrast to most amines, therefore, pyrrole is an extremely weak base ($K_b \sim 2.5 \times 10^{-14}$). By the same token, there is a high electron density in the ring, which causes pyrrole to be extremely reactive toward electrophilic substitution: it undergoes reactions like nitrosation and coupling with diazonium salts which are characteristic of only the most reactive benzene derivatives, phenols and amines.

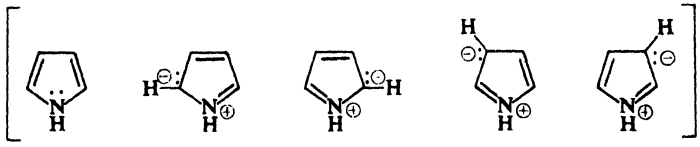
It thus appears that pyrrole is better represented by IV,



IV
Pyrrole

in which the circle represents the aromatic sextet.

What does IV mean in terms of conventional valence-bond structures? Pyrrole can be considered a hybrid of structures V–IX. Donation of electrons to the ring by nitrogen



V VI VII VIII IX
Pyrrole

is indicated by the ionic structures in which nitrogen bears a positive charge and the carbon atoms of the ring bear a negative charge.

Furan and thiophene have structures that are analogous to the structure of pyrrole. Where nitrogen in pyrrole carries a hydrogen atom, the oxygen or sulfur carries an unshared pair of electrons in an sp^2 orbital. Like nitrogen, the oxygen or

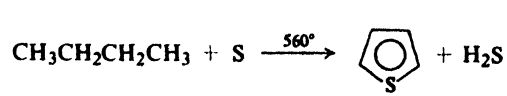
Furan Thiophene

sulfur atom provides two electrons for the π cloud; as a result these compounds, too, behave like extremely reactive benzene derivatives.

31.3 Source of pyrrole, furan, and thiophene

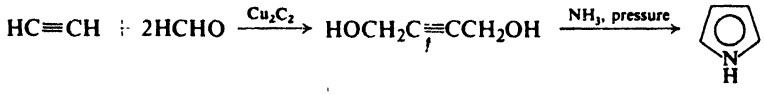
Pyrrole and thiophene are found in small amounts in coal tar. During the fractional distillation of coal tar, thiophene (b.p. 84°) is collected along with the benzene (b.p. 80°); as a result ordinary benzene contains about 0.5% of thiophene, and must be specially treated if *thiophene-free benzene* is desired.

Thiophene can be synthesized on an industrial scale by the high-temperature reaction between *n*-butane and sulfur.



n-Butane Thiophene

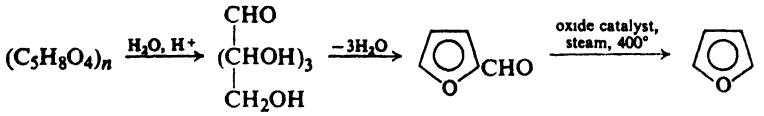
Pyrrole can be synthesized in a number of ways. For example:



1,4-Butynediol Pyrrole

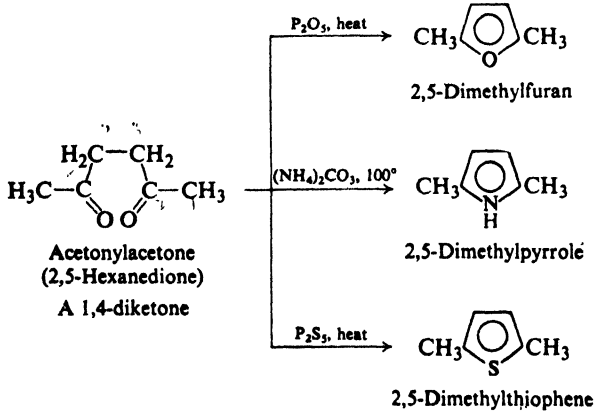
The pyrrole ring is the basic unit of the *porphyrin* system, which occurs, for example, in chlorophyll (p. 1004) and in hemoglobin (p. 1152).

Furan is most readily prepared by decarbonylation (elimination of carbon monoxide) of **furfural** (furfuraldehyde), which in turn is made by the treatment of oat hulls, corncobs, or rice hulls with hot hydrochloric acid. In the latter reaction pentosans (polypentosides) are hydrolyzed to pentoses, which then undergo dehydration and cyclization to form furfural.



Pentosan Pentose Furfural (2-Furancarboxyaldehyde) Furan

Certain substituted pyrroles, furans, and thiophenes can be prepared from the parent heterocycles by substitution (see Sec. 31.4); most, however, are prepared from open-chain compounds by ring closure. For example:



Acetonylacetone (2,5-Hexanedione) A 1,4-diketone

2,5-Dimethylfuran

2,5-Dimethylpyrrole

2,5-Dimethylthiophene

Problem 31.1 Give structural formulas for all intermediates in the following synthesis of acetonylacetone (2,5-hexanedione):

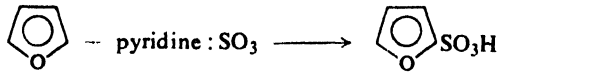
ethyl acetoacetate + $NaOC_2H_5$ ⟶ A ($C_6H_9O_3Na$)
A + I_2 ⟶ B ($C_{12}H_{18}O_6$) + NaI
B + dilute acid + heat ⟶ 2,5-hexanedione + carbon dioxide + ethanol

Problem 31.2 Outline a synthesis of 2,5-diphenylfuran, starting from ethyl benzoate and ethyl acetate.

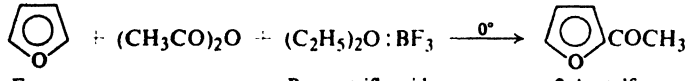
31.4 Electrophilic substitution in pyrrole, furan, and thiophene. Reactivity and orientation

Like other aromatic compounds, these five-membered heterocycles undergo nitration, halogenation, sulfonation, and Friedel-Crafts acylation. They are much more reactive than benzene, and resemble the most reactive benzene derivatives (amines and phenols) in undergoing such reactions as the Reimer-Tiemann reaction, nitrosation, and coupling with diazonium salts.

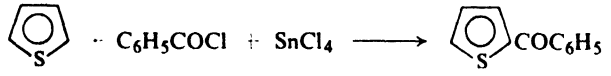
Reaction takes place predominantly at the 2-position. For example:



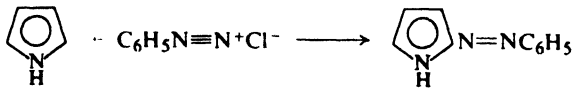
Furan — 2-Furansulfonic acid



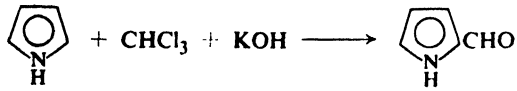
Furan — Boron trifluoride etherate — 2-Acetylfuran



Thiophene — 2-Benzoylthiophene



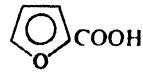
Pyrrole — 2-(Phenylazo)pyrrole



Pyrrole — 2-Pyrrolecarboxaldehyde (*Low yield*)

In some of the examples we notice modifications in the usual electrophilic reagents. The high reactivity of these rings makes it possible to use milder reagents in many cases, as, for example, the weak Lewis acid stannic chloride in the Friedel-Crafts acylation of thiophene. The sensitivity to protic acids of furan (which undergoes ring opening) and pyrrole (which undergoes polymerization) makes it necessary to modify the usual sulfonating agent

Problem 31.3 Furan undergoes ring opening upon treatment with sulfuric acid; it reacts almost explosively with halogens. Account for the fact that 2-furoic acid, however, can be sulfonated (in the 5-position) by treatment with fuming sulfuric acid, and brominated (in the 5-position) by treatment with bromine at 100°.

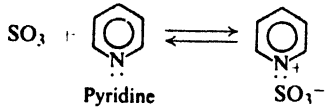


2-Furoic acid

Problem 31.4 Upon treatment with formaldehyde and acid, ethyl 2,4-dimethyl-3-pyrrolecarboxylate is converted into a compound of formula $C_{10}H_{26}O_4N_2$. What is the most likely structure for this product? How is it formed?

Problem 31.5 Predict the products from the treatment of furfural (2-furancarboxaldehyde) with concentrated aqueous NaOH.

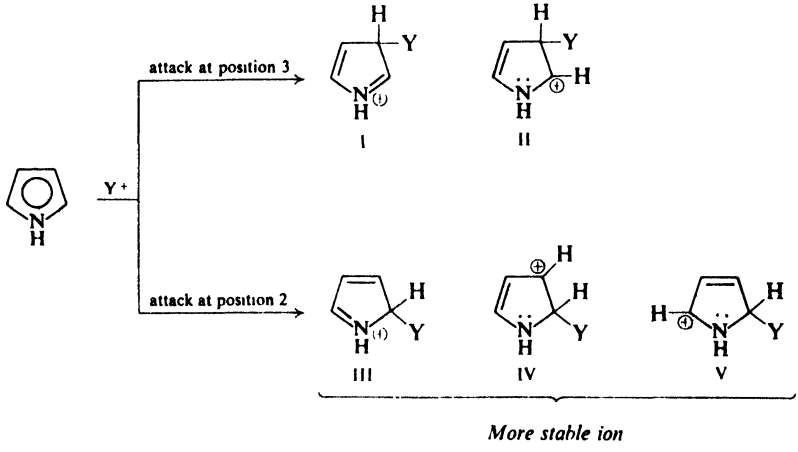
Problem 31.6 Sulfur trioxide dissolves in the tertiary amine pyridine to form a salt:



Show all steps in the most likely mechanism for the sulfonation of an aromatic compound by this reagent.

In our study of electrophilic aromatic substitution (Sec. 11.19 and Sec. 30.9), we found that we could account for orientation on the following basis: the controlling step is the attachment of the electrophilic reagent to the aromatic ring, which takes place in such a way as to yield the most stable intermediate carbonium ion. Let us apply this approach to the reactions of pyrrole.

Attack at position 3 yields a carbonium ion that is a hybrid of structures I and II. Attack at position 2 yields a carbonium ion that is a hybrid not only of structures III and IV (analogous to I and II) but also of structure V; the extra stabilization conferred by V makes this ion the more stable one.

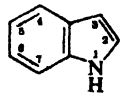


Viewed differently, attack at position 2 is faster because the developing positive charge is accommodated by *three* atoms of the ring instead of by only two.

Pyrrole is highly reactive, compared with benzene, because of contribution from the relatively stable structure III. In III *every atom has an octet of electrons*; nitrogen accommodates the positive charge simply by *sharing* four pairs of electrons. It is no accident that pyrrole resembles aniline in reactivity: both owe their high reactivity to the ability of nitrogen to share four pairs of electrons.

Orientation of substitution in furan and thiophene, as well as their high reactivity, can be accounted for in a similar way.

Problem 31.7 The heterocycle *indole*, commonly represented as formula VI, is found in coal tar and in orange blossoms.

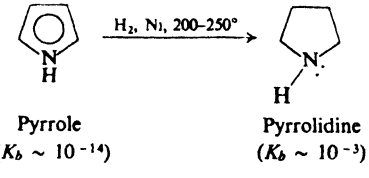


VI
Indole

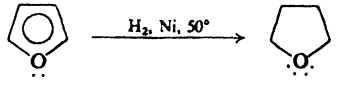
It undergoes electrophilic substitution, chiefly at position 3. Account (a) for the aromatic properties of indole, and (b) for the orientation in electrophilic substitution. (*Hint:* See Sec. 30.9.)

31.5 Saturated five-membered heterocycles

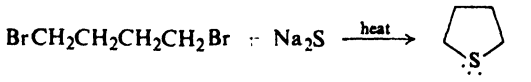
Catalytic hydrogenation converts pyrrole and furan into the corresponding saturated heterocycles, *pyrrolidine* and *tetrahydrofuran*. Since thiophene poisons most catalysts, *tetrahydrothiophene* is synthesized instead from open-chain compounds.



Pyrrole ($K_b \sim 10^{-14}$) Pyrrolidine ($K_b \sim 10^{-3}$)



Furan Tetrahydrofuran

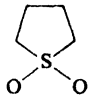


Tetrahydrothiophene

Saturation of these rings destroys the aromatic structure and, with it, the aromatic properties. Each of the saturated heterocycles has the properties we would expect of it: the properties of a secondary aliphatic amine, an aliphatic ether, or an aliphatic sulfide. With nitrogen's extra pair of electrons now available for sharing with acids, pyrrolidine ($K_b \sim 10^{-3}$) has the normal basicity of an aliphatic amine. Hydrogenation of pyrrole increases the base strength by a factor of 10^{11} (100 billion); clearly a fundamental change in structure has taken place.

Tetrahydrofuran is an important solvent, used, for example, in reductions with lithium aluminum hydride, in the preparation of arylmagnesium chlorides (Sec.

25.4), and in hydroborations. Oxidation of tetrahydrothiophene yields *tetramethylene sulfone* (or *sulfolane*), also used as a solvent (Sec. 1.21).



Tetramethylene sulfone
(Sulfolane)

We have encountered pyrrolidine as a secondary amine commonly used in making enamines (Sec. 26.8). The pyrrolidine ring occurs naturally in a number of alkaloids (Sec. 7.9), providing the basicity that gives these compounds their name (*alkali-like*).

Problem 31.8 An older process for the synthesis of both the adipic acid and the hexamethylenediamine needed in the manufacture of Nylon 66 (Sec. 32.7) started with tetrahydrofuran. Using only familiar chemical reactions, suggest possible steps in their synthesis.

Problem 31.9 Predict the products of the treatment of pyrrolidine with:

(a) aqueous HCl
(b) aqueous NaOH
(c) acetic anhydride
(d) benzenesulfonyl chloride + aqueous NaOH
(e) methyl iodide, followed by aqueous NaOH
(f) repeated treatment with methyl iodide, followed by Ag_2O and then strong heating

Problem 31.10 The alkaloid *hygrine* is found in the coca plant. Suggest a structure for it on the basis of the following evidence:

Hygrine ($C_8H_{15}ON$) is insoluble in aqueous NaOH but soluble in aqueous HCl. It does not react with benzenesulfonyl chloride. It reacts with phenylhydrazine to yield a phenylhydrazone. It reacts with NaOI to yield a yellow precipitate and a carboxylic acid ($C_7H_{13}O_2N$). Vigorous oxidation by CrO_3 converts hygrine into ***hygrinic acid*** ($C_6H_{11}O_2N$).

Hygrinic acid can be synthesized as follows:

$$BrCH_2CH_2CH_2Br + CH(COOC_2H_5)_2^-Na^+ \longrightarrow A\ (C_{10}H_{17}O_4Br)$$

$$A + Br_2 \longrightarrow B\ (C_{10}H_{16}O_4Br_2)$$

$$B + CH_3NH_2 \longrightarrow C\ (C_{11}H_{19}O_4N)$$

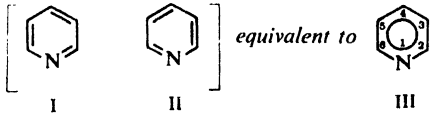
$$C + \text{aq. } Ba(OH)_2 + \text{heat} \longrightarrow D \xrightarrow{HCl} E \xrightarrow{\text{heat}} \text{hygrinic acid} + CO_2$$

SIX-MEMBERED RINGS

31.6 Structure of pyridine

Of the six-membered aromatic heterocycles, we shall take up only one, **pyridine**.

Pyridine is classified as aromatic on the basis of its properties. It is flat, with bond angles of 120°; the four carbon–carbon bonds are of the same length, and so are the two carbon–nitrogen bonds. It resists addition and undergoes electrophilic substitution. Its heat of combustion indicates a resonance energy of 23 kcal/mole.



I II III

Pyridine can be considered a hybrid of the Kekulé structures I and II. We shall represent it as structure III, in which the circle represents the aromatic sextet.

In electronic configuration, the nitrogen of pyridine is considerably different from the nitrogen of pyrrole. In pyridine the nitrogen atom, like each of the carbon atoms, is bonded to other members of the ring by the use of sp^2 orbitals, and provides one electron for the π cloud. The third sp^2 orbital of each carbon atom is used to form a bond to hydrogen; the third sp^2 orbital of nitrogen simply contains a pair of electrons, which are available for sharing with acids (Fig. 31.2).

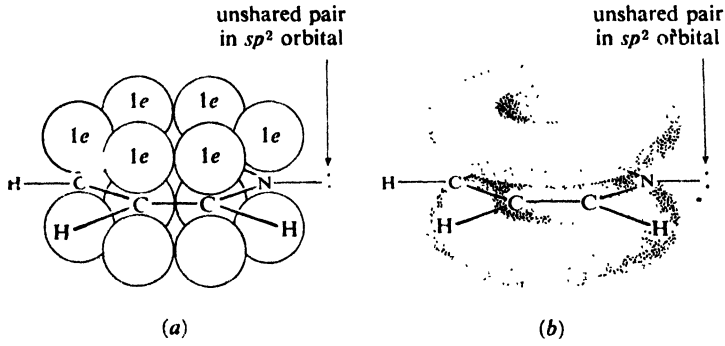


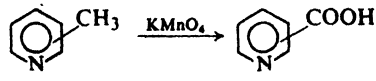
Figure 31.2. Pyridine molecule. (*a*) One electron in each *p* orbital; two electrons in sp^2 orbital of nitrogen. (*b*) The *p* orbitals overlap to form π clouds above and below plane of ring; two unshared electrons still in sp^2 orbital of nitrogen.

Because of this electronic configuration, the nitrogen atom makes pyridine a much stronger base than pyrrole, and affects the reactivity of the ring in a quite different way, as we shall see.

31.7 Source of pyridine compounds

Pyridine is found in coal tar. Along with it are found a number of methylpyridines, the most important of which are the monomethyl compounds, known as *picolines*.

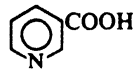
Oxidation of the picolines yields the pyridinecarboxylic acids.



Picoline (2-, 3-, or 4-)

Pyridinecarboxylic acid (2-, 3-, or 4-)

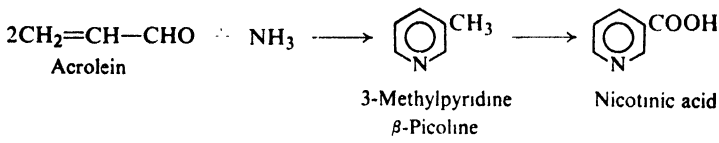
The 3-isomer (*nicotinic acid* or *niacin*) is a vitamin. The 4-isomer (*isonicotinic acid*) has been used, in the form of its hydrazide, in the treatment of tuberculosis.



Nicotinic acid
Niacin
3-Pyridinecarboxylic acid
Anti-pellagra factor

Isonicotinic acid hydrazide
(Isoniazid)

The increasing demand for certain pyridine derivatives has led to the development of syntheses involving ring closure. For example:



Acrolein

3-Methylpyridine
β-Picoline

Nicotinic acid

31.8 Reactions of pyridine

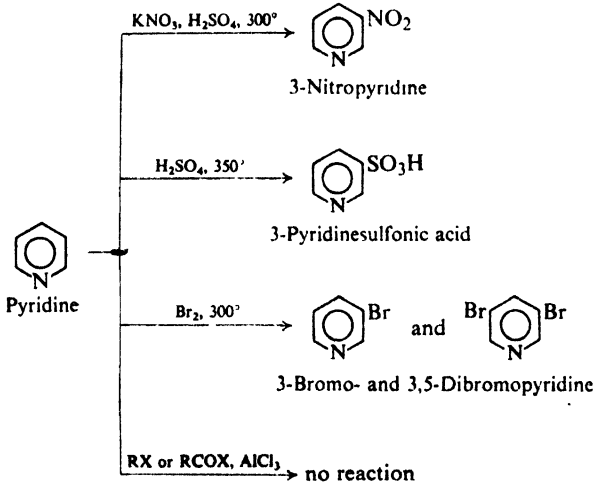
The chemical properties of pyridine are those we would expect on the basis of its structure. The ring undergoes the substitution, both electrophilic and nucleophilic, typical of aromatic rings; our interest will lie chiefly in the way the nitrogen atom affects these reactions.

There is another set of reactions in which pyridine acts as a base or nucleophile; these reactions involve nitrogen directly and are due to its unshared pair of electrons.

31.9 Electrophilic substitution in pyridine

Toward electrophilic substitution pyridine resembles a highly deactivated benzene derivative. It undergoes nitration, sulfonation, and halogenation only under very vigorous conditions, and does not undergo the Friedel-Crafts reaction at all.

Substitution occurs chiefly at the 3- (or β-) position.



Pyridine

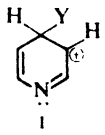
3-Nitropyridine

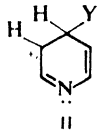
3-Pyridinesulfonic acid

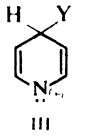
3-Bromo- and 3,5-Dibromopyridine

no reaction

Let us see if we can account for the reactivity and orientation on our usual basis of stability of the intermediate carbonium ion. Attack at the 4-position yields a carbonium ion that is a hybrid of structures I, II, and III;


I

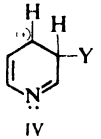

II

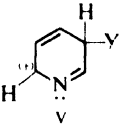

III

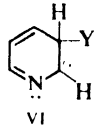
Especially unstable: nitrogen has sextet

Electrophilic attack at 4-position

Attack at the 3-position yields an ion that is a hybrid of structures IV, V, and VI.


IV


V


VI

Electrophilic attack at 3-position

(Attack at the 2-position resembles attack at the 4-position just as *ortho* attack resembles *para* attack in the benzene series.)

All these structures are less stable than the corresponding ones for attack on benzene, because of electron withdrawal by the nitrogen atom. As a result, pyridine undergoes substitution more slowly than benzene.

Of these structures, III is *especially* unstable, since in it the electronegative nitrogen atom has only a sextet of electrons. As a result, attack at the 4-position (or 2-position) is especially slow, and substitution occurs predominantly at the 3-position.

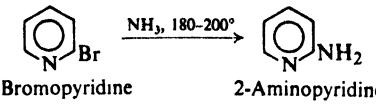
It is important to see the difference between substitution in pyridine and substitution in pyrrole. In the case of pyrrole, a structure in which nitrogen bears a positive charge (see Sec. 31.4) is especially stable since every atom has an octet of electrons; nitrogen accommodates the positive charge simply by sharing four pairs of electrons. In the case of pyridine, a structure in which nitrogen bears a positive charge (III) is especially unstable since nitrogen has only a sextet of electrons; nitrogen *shares* electrons readily, but as an electronegative atom it resists the *removal* of electrons.

Problem 31.11 2-Aminopyridine can be nitrated or sulfonated under much milder conditions than pyridine itself; substitution occurs chiefly at the 5-position. Account for these facts.

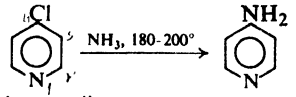
Problem 31.12 Because of the difficulty of nitrating pyridine, 3-aminopyridine is most conveniently made via nicotinic acid. Outline the synthesis of 3-aminopyridine from β-picoline.

31.10 Nucleophilic substitution in pyridine

Here, as in electrophilic substitution, the pyridine ring resembles a benzene ring that contains strongly electron-withdrawing groups. Nucleophilic substitution takes place readily, particularly at the 2- and 4-positions. For example:

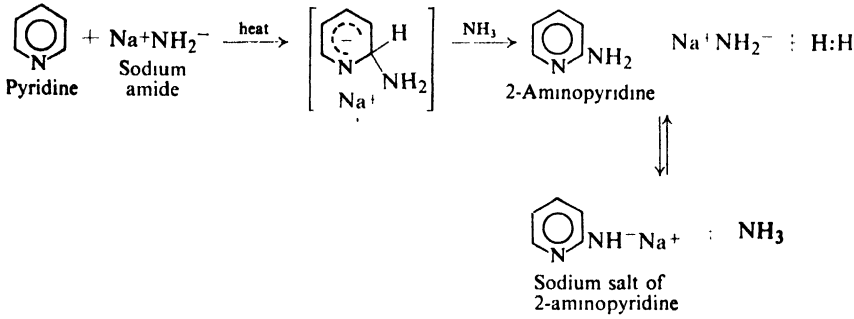


2-Bromopyridine 2-Aminopyridine

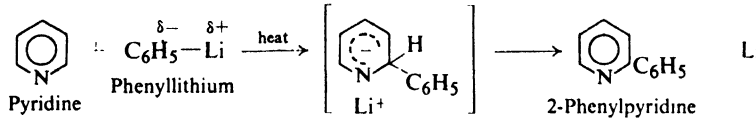


4-Chloropyridine 4-Aminopyridine

The reactivity of pyridine toward nucleophilic substitution is so great that even the powerfully basic hydride ion, $:H^-$, can be displaced. Two important examples of this reaction are amination by sodium amide (**Chichibabin reaction**), and alkylation or arylation by organolithium compounds.



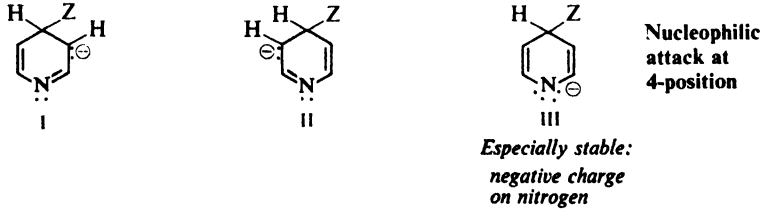
Pyridine Sodium amide 2-Aminopyridine Sodium salt of 2-aminopyridine



Pyridine Phenyllithium 2-Phenylpyridine

As we have seen (Sec. 25.8), nucleophilic aromatic substitution can take place by a mechanism that is quite analogous to the mechanism for electrophilic substitution. Reaction proceeds by two steps; the rate of the first step, formation of a charged particle, determines the rate of the overall reaction. In electrophilic substitution, the intermediate is positively charged; in nucleophilic substitution, the intermediate is negatively charged. The ability of the ring to accommodate the charge determines the stability of the intermediate and of the transition state leading to it, and hence determines the rate of the reaction.

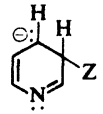
Nucleophilic attack at the 4-position yields a carbanion that is a hybrid of structures I, II, and III:

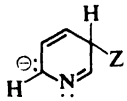


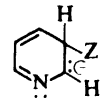
I II III

Especially stable: negative charge on nitrogen

Attack at the 3-position yields a carbanion that is a hybrid of structures\IV, V, and VI:


IV


V


VI

Nucleophilic attack at 3-position

(As before, attack at the 2-position resembles attack at the 4-position.)

All these structures are more stable than the corresponding ones for attack on a benzene derivative, because of electron withdrawal by the nitrogen atom. Structure III is *especially* stable, since the negative charge is located on the atom that can best accommodate it, the electronegative nitrogen atom. It is reasonable, therefore, that nucleophilic substitution occurs more rapidly on the pyridine ring than on the benzene ring, and more rapidly at the 2- and 4-positions than at the 3-position.

The same electronegativity of nitrogen that makes pyridine unreactive toward electrophilic substitution makes pyridine highly reactive toward nucleophilic substitution.

31.11 Basicity of pyridine

Pyridine is a base with $K_b = 2.3 \times 10^{-9}$. It is thus much stronger than pyrrole ($K_b \sim 2.5 \times 10^{-14}$) but much weaker than aliphatic amines ($K_b \sim 10^{-4}$).

Pyridine has a pair of electrons (in an sp^2 orbital) that is available for sharing with acids; pyrrole has not, and can accept an acid only at the expense of the aromatic character of the ring.

The fact that pyridine is a weaker base than aliphatic amines is more difficult to account for, but at least it fits into a pattern. Let us turn for a moment to the basicity of the carbon analogs of amines, the carbanions, and use the approach of Sec. 8.10.

Benzene is a stronger acid than an alkane, as shown by its ability to displace an alkane from its salts; this, of course, means that the phenyl anion, $C_6H_5^-$, is a weaker base than an alkyl anion, R^-.

$$\underset{\text{Stronger base}}{R{:}^-Na^+} + \underset{\text{Stronger acid}}{C_6H_5{:}H} \rightleftharpoons \underset{\text{Weaker acid}}{R{:}H} + \underset{\text{Weaker base}}{C_6H_5{:}^-Na^+}$$

In the same way, acetylene is a stronger acid than benzene, and the acetylide ion is a weaker base than the phenyl anion.

$$\underset{\text{Stronger base}}{C_6H_5{:}^-Na^+} + \underset{\text{Stronger acid}}{HC{\equiv}C{:}H} \rightleftharpoons \underset{\text{Weaker acid}}{C_6H_5{:}H} + \underset{\text{Weaker base}}{HC{\equiv}C{:}^-Na^+}$$

Thus we have the following sequences of acidity of hydrocarbons and basicity of their anions:

Relative acidity: $HC{\equiv}C{:}H > C_6H_5{:}H > R{:}H$

Relative basicity: $HC{\equiv}C{:}^- < C_6H_5{:}^- < R{:}^-$

A possible explanation for these sequences can be found in the electronic configuration of the carbanions. In the alkyl, phenyl, and acetylide anions, the unshared pair of electrons occupies respectively an sp^3, an sp^2, and an sp orbital. The availability of this pair for sharing with acids determines the basicity of the particular anion. As we proceed along the series sp^3, sp^2, sp, the p character of the orbital decreases and the s character increases. Now, an electron in a p orbital is at some distance from the nucleus and is held relatively loosely; an electron in an s orbital, on the other hand, is close to the nucleus and is held more tightly. Of the three anions, the alkyl ion is the strongest base since its pair of electrons is held most loosely, in an sp^3 orbital. The acetylide ion is the weakest base since its pair of electrons is held most tightly, in an sp orbital.

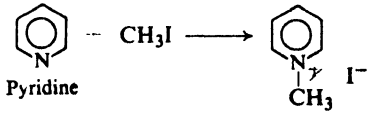
Pyridine bears the same relationship to an aliphatic amine as the phenyl anion bears to an alkyl anion. The pair of electrons that gives pyridine its basicity occupies an sp^2 orbital; it is held more tightly and is less available for sharing with acids than the pair of electrons of an aliphatic amine, which occupies an sp^3 orbital.

Problem 31.13 Predict the relative basicities of amines (RCH_2NH_2), imines (RCH=NH), and nitriles (RC≡N).

Pyridine is widely used in organic chemistry as a water-soluble base, as, for example, in the Schotten-Baumann acylation procedure (Sec. 20.8).

Problem 31.14 Ethyl bromosuccinate is converted into the unsaturated ester ethyl fumarate by the action of pyridine. What is the function of the pyridine? What advantage does it have here over the usual alcoholic KOH?

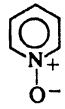
Like other amines, pyridine has nucleophilic properties, and reacts with alkyl halides to form quaternary ammonium salts.



Pyridine

N-Methylpyridinium iodide
(Pyridine methiodide)

Problem 31.15 Like any other tertiary amine, pyridine can be converted (by peroxybenzoic acid) into its N-oxide.



Pyridine N-oxide

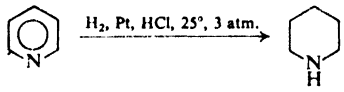
In contrast to pyridine itself, pyridine N-oxide readily undergoes nitration, chiefly in the 4-position. How do you account for this reactivity and orientation?

Problem 31.16 Pyridine N-oxides not only are reactive toward electrophilic substitution, but also seem to be reactive toward nucleophilic substitution, particularly at the 2- and 4-positions. For example, treatment of 4-nitropyridine N-oxide with hydrobromic acid gives 4-bromopyridine N-oxide. How do you account for this reactivity and orientation?

Problem 31.17 The oxygen of pyridine N-oxide is readily removed by treatment with PCl_3. Suggest a practical route to 4-nitropyridine. To 4-bromopyridine.

31.12 Reduction of pyridine

Catalytic hydrogenation of pyridine yields the aliphatic heterocyclic compound **piperidine**, $C_5H_{11}N$.



Pyridine ($K_b = 2.3 \times 10^{-9}$) Piperidine ($K_b = 2 \times 10^{-3}$)

Piperidine ($K_b = 2 \times 10^{-3}$) has the usual basicity of a secondary aliphatic amine. Like pyridine, it is often used as a basic catalyst in such reactions as the Knoevenagel reaction (Problem 21.22 (f), p. 714) or Michael addition (Sec. 27.7).

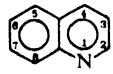
Like the pyrrolidine ring, the piperidine and pyridine rings are found in a number of alkaloids, including *nicotine*, *strychnine*, *cocaine*, and *reserpine* (see p. 1004).

Problem 31.18 Why can piperidine not be used in place of pyridine in the Schotten-Baumann procedure?

FUSED RINGS

31.13 Quinoline. The Skraup synthesis

Quinoline, C_9H_7N, contains a benzene ring and a pyridine ring fused as shown in I.



I

Quinoline ($K_b = 3 \times 10^{-10}$)

In general, its properties are the ones we would expect from what we have learned about pyridine and naphthalene.

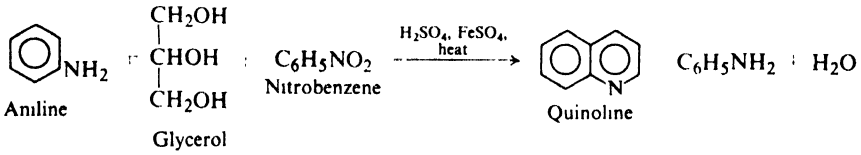
Problem 31.19 Account for the following properties of quinoline:

(a) Treatment with nitric and sulfuric acids gives 5- and 8-nitroquinolines; treatment with fuming sulfuric acid gives 5- and 8-quinolinesulfonic acids.
(b) Oxidation by $KMnO_4$ gives 2,3-pyridinedicarboxylic acid (*quinolinic acid*).
(c) Treatment with sodamide gives 2-aminoquinoline; treatment with alkyllithium compounds gives 2-alkylquinolines.

Problem 31.20 *8-Hydroxyquinoline* (8-quinolinol) is a reagent in inorganic analysis. Suggest a method of synthesizing it.

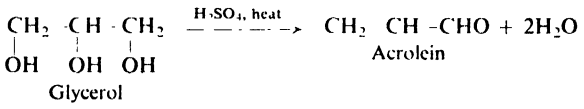
Quinoline is found in coal tar. Although certain derivatives of quinoline can be made from quinoline itself by substitution, most are prepared from benzene derivatives by ring closure.

Perhaps the most generally useful method for preparing substituted quinolines is the **Skraup synthesis**. In the simplest example, quinoline itself is obtained from the reaction of aniline with glycerol, concentrated sulfuric acid, nitrobenzene, and ferrous sulfate.

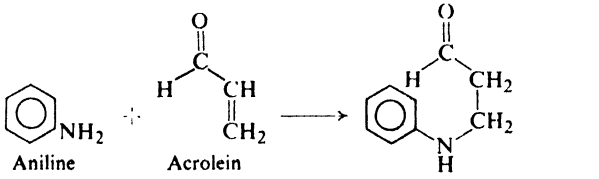


The following steps seem to be involved:

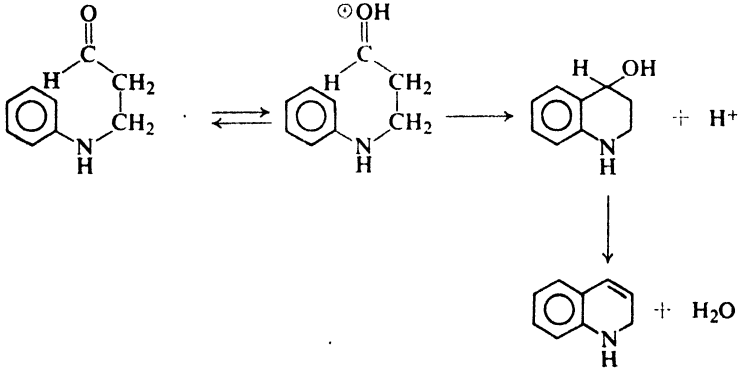
(1) Dehydration of glycerol by hot sulfuric acid to yield the unsaturated aldehyde acrolein:



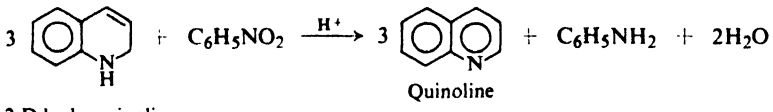
(2) Nucleophilic addition of aniline to acrolein to yield β-(phenylamino)propionaldehyde:



(3) Electrophilic attack on the aromatic ring by the electron-deficient carbonyl carbon of the protonated aldehyde (this is the actual ring-closing step):



(4) Oxidation by nitrobenzene resulting in the aromatization of the newly formed ring:



1,2-Dihydroquinoline

Quinoline

Ferrous sulfate in some way moderates the otherwise very vigorous reaction.

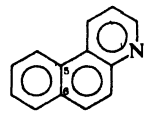
Thus we see that what at first appears to be a complicated reaction is actually a sequence of simple steps involving familiar, fundamental types of reactions: acid-catalyzed dehydration, nucleophilic addition to an α,β-unsaturated carbonyl compound, electrophilic aromatic substitution, and oxidation.

The components of the basic synthesis can be modified to yield a wide variety of quinoline derivatives. For example:

aniline + crotonaldehyde ⟶ 2-methylquinoline (quinaldine)

3-nitro-4-aminoanisole + glycerol ⟶ 6-methoxy-8-nitroquinoline

2-aminonaphthalene + glycerol ⟶



5,6-Benzoquinoline
(1-Azaphenanthrene)

Nitrobenzene is often replaced as oxidizing agent by arsenic acid, H_3AsO_4, which usually gives a less violent reaction; vanadium pentoxide is sometimes added as a catalyst. Sulfuric acid can be replaced by phosphoric acid or other acids.

Problem 31.21 Show all steps in the Skraup syntheses mentioned above.

Problem 31.22 The dehydration of glycerol to yield acrolein involves acid-catalyzed dehydration and keto–enol tautomerization. Outline the possible steps in the dehydration. (*Hint:* Which —OH is easier to eliminate, a primary or a secondary?)

Problem 31.23 What is the product of the application of the Skraup synthesis to (a) *o*-nitroaniline, (b) *o*-aminophenol, (c) *o*-phenylenediamine, (d) *m*-phenylenediamine, (e) *p*-toluidine?

Problem 31.24 Outline the synthesis of 6-bromoquinoline. Of 8-methylquinoline.

Problem 31.25 In the **Doebner-von Miller** modification of the Skraup synthesis, aldehydes, ketones, or mixtures of aldehydes and ketones replace the glycerol. If acetaldehyde is used, for example, the product from aniline is 2-methylquinoline (*quinaldine*). (a) Account for its formation. (b) Predict the product if methyl vinyl ketone were used. (c) If a mixture of benzaldehyde and pyruvic acid, $CH_3COCOOH$, were used.

Problem 31.26 Account for the formation of 2,4-dimethylquinoline from aniline and acetylacetone (2,4-pentanedione) by the Doebner-von Miller synthesis. (*Hint:* See Problem 21, p. 724.)

31.14 Isoquinoline. The Bischler-Napieralski synthesis

Isoquinoline, C_9H_7N, contains a benzene ring and a pyridine ring fused as shown in I:

I
Isoquinoline
($K_b = 1.1 \times 10^{-9}$)

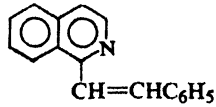
Isoquinoline, like quinoline, has the properties we would expect from what we know about pyridine and naphthalene.

Problem 31.27 Account for the following properties of isoquinoline. (*Hint:* Review orientation in β-substituted naphthalenes, Sec. 30.13.)

(a) Nitration gives 5-nitroisoquinoline.

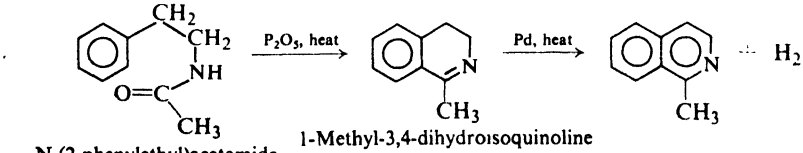
(b) Treatment with potassium amide, KNH_2, gives 1-aminoisoquinoline, and treatment with alkyllithium compounds gives 1-alkylisoquinoline; the 3-substituted products are not obtained.

(c) 1-Methylisoquinoline reacts with benzaldehyde to yield compound II, whereas 3-methylisoquinoline undergoes no reaction. (*Hint:* See Problem 21.22 (c), p. 714.)



II

An important method for making derivatives of isoquinoline is the **Bischler-Napieralski synthesis**. Acyl derivatives of β-phenylethylamine are cyclized by treatment with acids (often P_2O_5) to yield dihydroisoquinolines, which can then be aromatized.



N-(2-phenylethyl)acetamide

1-Methyl-3,4-dihydroisoquinoline

1-Methylisoquinoline

Problem 31.28 To what general class of reactions does the ring closure belong? What is the function of the acid? (Check your answers in Sec. 32.7.)

Problem 31.29 Outline the synthesis of N-(2-phenylethyl)acetamide from toluene and aliphatic and inorganic reagents.

PROBLEMS

1. Give structures and names of the principal products from the reaction (if any) of pyridine with:

(a) Br_2, 300°
(b) H_2SO_4, 350°
(c) acetyl chloride, $AlCl_3$
(d) KNO_3, H_2SO_4, 300°
(e) $NaNH_2$, heat
(f) C_6H_5Li
(g) dilute HCl
(h) dilute NaOH
(i) acetic anhydride
(j) benzenesulfonyl chloride
(k) ethyl bromide
(l) benzyl chloride
(m) peroxybenzoic acid
(n) peroxybenzoic acid, then HNO_3, H_2SO_4
(o) H_2, Pt

2. Give structures and names of the principal products from each of the following reactions:

(a) thiophene + conc. H_2SO_4
(b) thiophene + acetic anhydride, $ZnCl_2$
(c) thiophene + acetyl chloride, $TiCl_4$
(d) thiophene + fuming nitric acid in acetic anhydride
(e) product of (d) + Sn, HCl
(f) thiophene + one mole Br_2
(g) product of (f) + Mg; then CO_2; then H^+
(h) pyrrole + pyridine:SO_3
(i) pyrrole + diazotized sulfanilic acid
(j) product of (i) + $SnCl_2$
(k) pyrrole + H_2, Ni $\longrightarrow$ C_4H_9N
(l) furfural + acetone + base
(m) quinoline + HNO_3/H_2SO_4
(n) quinoline N-oxide + HNO_3/H_2SO_4
(o) isoquinoline + *n*-butyllithium

3. Pyrrole can be reduced by zinc and acetic acid to a *pyrroline*, C_4H_7N. (a) What structures are possible for this pyrroline?

(b) On the basis of the following evidence which structure must the pyrroline have?

pyrroline + O_3; then H_2O; then H_2O_2 $\longrightarrow$ A ($C_4H_7O_4N$)
chloroacetic acid + NH_3 $\longrightarrow$ B ($C_2H_5O_2N$)
B + chloroacetic acid $\longrightarrow$ A

4. Furan and its derivatives are sensitive to protic acids. The following reactions illustrate what happens.

2,5-dimethylfuran + dilute H_2SO_4 $\longrightarrow$ C ($C_6H_{10}O_2$)
C + NaOI $\longrightarrow$ succinic acid

(a) What is C? (b) Outline a likely series of steps for its formation from 2,5-dimethylfuran.

5. Pyrrole reacts with formaldehyde in hot pyridine to yield a mixture of products from which there can be isolated a small amount of a compound of formula $(C_5H_5N)_4$. Suggest a possible structure for this compound. (*Hint:* See Sec. 32.7 and p. 1004.)

6. There are three isomeric pyridinecarboxylic acids, $(C_5H_4N)COOH$: D, m.p. 137°; E, m.p. 234–7°; and F, m.p. 317°. Their structures were proved as follows:

quinoline + $KMnO_4$, OH^- $\longrightarrow$ a diacid ($C_7H_5O_4N$) $\xrightarrow{\text{heat}}$ E, m.p. 234–7°
isoquinoline + $KMnO_4$, OH^- $\longrightarrow$ a diacid ($C_7H_5O_4N$) $\xrightarrow{\text{heat}}$ E, m.p. 234–7° and F, m.p. 317°

What structures should be assigned to D, E, and F?

7. (a) What structures are possible for G?

m-toluidine + glycerol $\xrightarrow{\text{Skraup}}$ G ($C_{10}H_9N$)

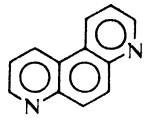
(b) On the basis of the following evidence which structure must G actually have?

2,3-diaminotoluene + glycerol $\xrightarrow{\text{Skraup}}$ H ($C_{10}H_{10}N_2$)

H + $NaNO_2$, HCl; then H_3PO_2 $\longrightarrow$ G

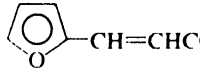
8. Outline all steps in a possible synthesis of each of the following from benzene, toluene, and any needed aliphatic and inorganic reagents:

(a) 1-phenylisoquinoline
(b) 1-benzylisoquinoline
(c) 1,5-dimethylisoquinoline
(d) 6-nitroquinoline
(e) 2-methyl-6-quinolinecarboxylic acid
(f) 1,8-diazaphenanthrene (*Hint.* Use the Skraup synthesis twice.)



1,8-Diazaphenanthrene

9. Outline all steps in each of the following syntheses, using any other needed reagents:

(a) β-cyanopyridine from β-picoline
(b) 2-methylpiperidine from pyridine
(c) 5-aminoquinoline from quinoline
(d) ethyl 5-nitro-2-furoate from furfural
(e) furylacrylic acid,  , from furfural
(f) 1,2,5-trichloropentane from furfural
(g) 3-indolecarboxaldehyde from indole

10. Give the structures of compounds I through JJ formed in the following syntheses of heterocyclic systems.

(a) ethyl malonate + urea, base, heat $\longrightarrow$ I ($C_4H_4O_3N_2$), a *pyrimidine* (1,3-diazine)
(b) 2,5-hexanedione + H_2N NH_2 $\longrightarrow$ J ($C_6H_{10}N_2$)
J + air $\longrightarrow$ K ($C_6H_8N_2$), a *pyridazine* (1,2-diazine)
(c) 2,4-pentanedione + H_2N NH_2 $\longrightarrow$ L ($C_5H_8N_2$), a *pyrazole*
(d) 2,3-butanedione + *o*-$C_6H_4(NH_2)_2$ $\longrightarrow$ M ($C_{10}H_{10}N_2$), a *quinoxaline*
(e) ethylene glycol + phosgene $\longrightarrow$ N ($C_3H_4O_3$), a *1,3-dioxolanone*
(f) *o*-aminobenzoic acid + chloroacetic acid $\longrightarrow$ O ($C_9H_9O_4N$)
O + base, strong heat $\longrightarrow$ P (C_8H_7ON), *indoxyl*, an intermediate in the synthesis of indigo
(g) aminoacetone $\longrightarrow$ Q ($C_6H_{10}N_2$)
Q + air $\longrightarrow$ R ($C_6H_8N_2$), a *pyrazine* (1,4-diazine)
(h) ethylenediamine + ethyl carbonate $\longrightarrow$ S ($C_3H_6ON_2$), an *imidazolidone*
(i) *o*-$C_6H_4(NH_2)_2$ + acetic acid, strong heat $\longrightarrow$ T ($C_8H_8N_2$), a *benzimidazole*
(j) ethyl *o*-aminobenzoate + malonic ester $\longrightarrow$ U ($C_{14}H_{17}O_5N$), insoluble in dilute acid
U $\xrightarrow{NaOC_2H_5}$ V ($C_{12}H_{11}O_4N$)
V + acid, warm $\longrightarrow$ W ($C_9H_7O_2N$), a *quinoline*
(k) repeat (j) starting with ethyl 3-amino-2-pyridinecarboxylate $\longrightarrow$ a *1,5-diazanaphthalene*
(l) benzalacetophenone + KCN + acetic acid $\longrightarrow$ X ($C_{16}H_{13}ON$)
X + CH_3OH, H^+, H_2O $\longrightarrow$ Y ($C_{17}H_{16}O_3$) + NH_4^+
Y + phenylhydrazine $\longrightarrow$ Z ($C_{22}H_{18}ON_2$), a *dihydro-1,2-diazine*

(m) acrylic acid + H_2N-NH_2 ⟶ AA ($C_3H_8O_2N_2$) ⟶ BB ($C_3H_6ON_2$), a *pyrazolidone*

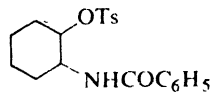
(n) *o*-$C_6H_4(NH_2)_2$ + glycerol $\xrightarrow{\text{Skraup}}$ CC ($C_{12}H_8N_2$), a *4,5-diazaphenanthrene*

(o) di(*o*-nitrophenyl)acetylene + Br_2 ⟶ DD ($C_{14}H_8O_4N_2Br_2$)
DD + Sn, HCl ⟶ EE ($C_{14}H_{12}N_2Br_2$)
EE $\xrightarrow{\text{warm}}$ [FF ($C_{14}H_{11}N_2Br$)] ⟶ GG ($C_{14}H_{10}N_2$), which contains four fused aromatic rings

(p) *m*-$ClC_6H_4CH_2CH_2CH_2NHCH_3$ + C_6H_5Li ⟶ HH ($C_{10}H_{13}N$), a *tetrahydroquinoline*

(q) *o*-$ClC_6H_4NHCOC_6H_5$ + KNH_2/NH_3 ⟶ II ($C_{13}H_9ON$), a *benzoxazole*

(r) *trans*-I + base ⟶ JJ ($C_{13}H_{15}ON$), an *oxazoline*



I

(s) How do you account for the fact that *cis*-I undergoes reaction (r) much more slowly than *trans*-I?

11. The structure of *papaverine*, $C_{20}H_{21}O_4N$, one of the opium alkaloids, has been established by the following synthesis:

3,4-dimethoxybenzyl chloride + KCN ⟶ KK ($C_{10}H_{11}O_2N$)
KK + hydrogen, Ni ⟶ LL ($C_{10}H_{15}O_2N$)
KK + aqueous acid, heat ⟶ MM $\xrightarrow{PCl_5}$ NN ($C_{10}H_{11}O_3Cl$)
LL + NN ⟶ OO ($C_{20}H_{25}O_3N$)
OO + P_2O_5, heat ⟶ PP ($C_{20}H_{23}O_4N$)
PP + Pd, 200° ⟶ papaverine

12. *Plasmochin* (also called *Pamaquine*), a drug effective against malaria, has been synthesized as follows:

ethylene oxide + diethylamine ⟶ QQ ($C_6H_{15}ON$)
QQ + $SOCl_2$ ⟶ RR ($C_6H_{14}NCl$)
RR + sodioacetoacetic ester ⟶ SS ($C_{12}H_{23}O_3N$)
SS + dilute H_2SO_4, warm ⟶ TT ($C_9H_{19}ON$) + CO_2 + C_2H_5OH
TT + H_2, Ni ⟶ UU ($C_9H_{21}ON$)
UU + conc. HBr ⟶ VV ($C_9H_{20}NBr$)
4-amino-3-nitroanisole + glycerol $\xrightarrow{\text{Skraup}}$ WW ($C_{10}H_8O_3N_2$)
WW + Sn + HCl ⟶ XX ($C_{10}H_{10}ON_2$)
VV + XX ⟶ Plasmochin ($C_{19}H_{29}ON_3$)

What is the most likely structure of Plasmochin?

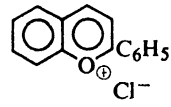
13. (−)-*Nicotine*, the alkaloid in tobacco, can be synthesized in the following way:

nicotinic acid + $SOCl_2$, heat ⟶ nicotinoyl chloride (C_6H_4ONCl)
nicotinoyl chloride + $C_2H_5OCH_2CH_2CH_2CdCl$ ⟶ YY ($C_{11}H_{15}O_2N$)
YY + NH_3, H_2, catalyst ⟶ ZZ ($C_{11}H_{18}ON_2$)
ZZ + HBr + strong heat ⟶ AAA ($C_9H_{12}N_2$) + ethyl bromide
AAA + CH_3I, NaOH ⟶ (±)-nicotine ($C_{10}H_{14}N_2$)
(±)-nicotine + (+)-tartaric acid ⟶ BBB and CCC (both $C_{14}H_{20}O_6N_2$)
BBB + NaOH ⟶ (−)-nicotine + sodium tartrate

What is the structure of (±)-nicotine? Write equations for all the above reactions.

14. The red and blue colors of many flowers and fruits are due to the *anthocyanins*, glycosides of pyrylium salts. The parent structure of the pyrylium salts is *flavylium chloride*, which can be synthesized as follows:

salicylaldehyde + acetophenone $\xrightarrow{\text{aldol}}$ DDD ($C_{15}H_{12}O_2$)
DDD + HCl $\longrightarrow$ flavylium chloride, a salt containing three aromatic rings



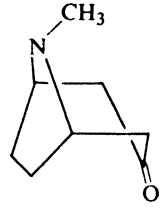
Flavylium chloride

(a) What is the structure of DDD? (b) Outline a likely series of steps leading from DDD to flavylium chloride. (c) Account for the aromatic character of the fused ring system.

15. *Tropinic acid*, $C_8H_{13}O_4N$, is a degradation product of atropine, an alkaloid of the deadly nightshade, *Atropa belladonna*. It has a neutralization equivalent of 94 ± 1. It does not react with benzenesulfonyl chloride, cold dilute $KMnO_4$, or Br_2/CCl_4. Exhaustive methylation gives the following results:

tropinic acid + CH_3I $\longrightarrow$ EEE ($C_9H_{16}O_4NI$)
EEE + Ag_2O, then strong heat $\longrightarrow$ FFF ($C_9H_{15}O_4N$)
FFF + CH_3I $\longrightarrow$ GGG ($C_{10}H_{18}O_4NI$)
GGG + Ag_2O, then strong heat $\longrightarrow$ HHH ($C_7H_8O_4$) + $(CH_3)_3N$ + H_2O
HHH + H_2, Ni $\longrightarrow$ heptanedioic acid (pimelic acid)

(a) What structures are likely for tropinic acid?
(b) Tropinic acid is formed by oxidation with CrO_3 of *tropinone*, whose structure has been shown by synthesis to be



Tropinone

Now what is the most likely structure for tropinic acid?

16. *Tropilidene*, 1,3,5-cycloheptatriene, has been made from tropinone (Problem 15). Show how this might have been done. (*Hint:* See Problem 23, p. 782.)

17. Reduction of tropinone (Problem 15) gives *tropine* and *pseudotropine*, both $C_8H_{15}ON$. When heated with base, tropine is converted into pseudotropine. Give likely structures for tropine and pseudotropine, and explain your answer.

18. *Arecaidine*, $C_7H_{11}O_2N$, an alkaloid of betel nut, has been synthesized in the following way:

ethyl acrylate + NH_3 $\xrightarrow{\text{Michael}}$ III ($C_5H_{11}O_2N$)
III + ethyl acrylate $\xrightarrow{\text{Michael}}$ JJJ ($C_{10}H_{19}O_4N$)
JJJ + sodium ethoxide $\xrightarrow{\text{Dieckmann}}$ KKK ($C_8H_{13}O_3N$)
KKK + benzoyl chloride $\longrightarrow$ LLL ($C_{15}H_{17}O_4N$)
LLL + H_2, Ni $\longrightarrow$ MMM ($C_{15}H_{19}O_4N$)
MMM + acid, heat $\longrightarrow$ NNN ($C_6H_9O_2N$), *guvacine*, another betel nut alkaloid + C_6H_5COOH + C_2H_5OH
NNN + CH_3I $\longrightarrow$ arecaidine ($C_7H_{11}O_2N$)

(a) What is the most likely structure of arecaidine? Of guvacine?
(b) What will guvacine give upon dehydrogenation?

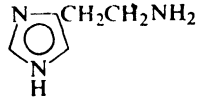
19. Give the structures of compounds OOO through UUU.

thiophene + 3-hexanone + H_2SO_4 ⟶ OOO ($C_{14}H_{18}S_2$)
OOO + $(CH_3CO)_2O$ + $HClO_4$ ⟶ PPP ($C_{16}H_{20}OS_2$)
PPP + N_2H_4 + KOH + heat ⟶ QQQ ($C_{16}H_{22}S_2$)
QQQ + $C_6H_5N(CH_3)CHO$ ⟶ RRR ($C_{17}H_{22}OS_2$), an aldehyde
RRR + Ag_2O ⟶ SSS ($C_{17}H_{22}O_2S_2$)
SSS *was resolved*
(+)-SSS + Cu, quinoline, heat ⟶ CO_2 + (+)-TTT ($C_{16}H_{22}S_2$)
(+)-TTT + H_2/Ni ⟶ UUU ($C_{16}H_{34}$), *optically inactive*

What is the significance of the optical inactivity of UUU?

20. (a) Account for the aromatic properties of the imidazole ring.

(b) Arrange the nitrogen atoms of *histamine* (the substance responsible for many allergenic reactions) in order of their expected basicity, and account for your answer.

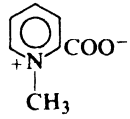


Histamine

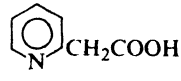
21. When heated in solution, 2-pyridinecarboxylic acid (II) loses carbon dioxide and forms pyridine. The rate of this decarboxylation is slowed down by addition of either acid or base. When decarboxylation is carried out in the presence of a ketone, R_2CO, there is obtained not only pyridine but also the tertiary alcohol III. The N-methyl derivative (IV) is decarboxylated much faster than II.

II

III



IV



V

(a) Show all steps in the most likely mechanism for decarboxylation of II. Show how this mechanism is consistent with each of the above facts.

(b) In the decarboxylation of the isomeric pyridinecarboxylic acids (II and its isomers), the order of reactivity is:

$$2 > 3 > 4$$

In the decarboxylation of the isomeric pyridineacetic acids (V and its isomers), on the other hand, the order of reactivity is:

$$2 \text{ or } 4 > 3$$

How do you account for each order of reactivity? Why is there a difference between the two sets of acids? (The same mechanism seems to be involved in both cases.)

Chapter 32 Macromolecules. Polymers and Polymerization

32.1 Macromolecules

So far, our study of organic chemistry has dealt mainly with rather small molecules, containing perhaps as many as 50 to 75 atoms. But there also exist enormous molecules called *macromolecules*, which contain hundreds of thousands of atoms. Some of these are naturally occurring, and make up classes of compounds that are, quite literally, vital: the *polysaccharides* starch and cellulose, which provide us with food, clothing, and shelter (Chap. 35); *proteins*, which constitute much of the animal body, hold it together, and run it (Chap. 36); and *nucleic acids*, which control heredity on the molecular level (Chap. 37).

Macromolecules can be man-made, too. The first syntheses were aimed at making substitutes for the natural macromolecules, rubber and silk; but a vast technology has grown up that now produces hundreds of substances that have no natural counterparts. Synthetic macromolecular compounds include: **elastomers**, which have the particular kind of elasticity characteristic of rubber; **fibers**, long, thin, and threadlike, with the great strength *along the fiber* that characterizes cotton, wool, and silk; and **plastics**, which can be extruded as sheets or pipes, painted on surfaces, or molded to form countless objects. We wear these man-made materials, eat and drink from them, sleep between them, sit and stand on them; turn knobs, pull switches, and grasp handles made of them; with their help we hear sounds and see sights remote from us in time and space; we live in houses and move about in vehicles that are increasingly made of them.

We sometimes deplore the resistance to the elements of these seemingly all too immortal materials, and fear that civilization may some day be buried beneath a pile of plastic debris—plastic cigar tips have been found floating in the Sargasso Sea—but with them we can do things never before possible. By use of plastics, blind people can be made to see, and cripples to walk; heart valves can be repaired and arteries patched; damaged tracheas, larynxes, and ureters can be replaced, and some day, perhaps, entire hearts. These materials protect us against heat and

cold, electric shock and fire, rust and decay. As tailor-made solvents, they may soon be used to extract fresh water from the sea. Surely the ingenuity that has produced these substances can devise ways of disposing of the waste they create: the problem is not one of technology, but of sociology and, ultimately, of politics.

In this chapter, we shall be first—and chiefly—concerned with the chemical reactions by which macromolecules are formed, and the structures that these reactions produce. Then, we shall see how these structures lead to the properties on which the use of the macromolecules depend: why rubber is elastic, for example, and why nylon is a strong fiber. In later chapters, we shall take up the natural macromolecules—polysaccharides, proteins, and nucleic acids—and study them in much the same way.

In all this, we must remember that what makes macromolecules special is, of course, their great size. This great size permits a certain complexity of structure, not just on the molecular level, but on a *secondary* level that involves the disposition of molecules with respect to each other. Are the molecules stretched out neatly alongside one another, or coiled up independently? What forces act between different molecules? What happens to a collection of giant molecules when it is heated, or cooled, or stretched? As we shall see, the answers to questions like these are found ultimately in structure as we have known it: the nature of functional groups and substituents, their sequence in the molecule, and their arrangement in space.

32.2 Polymers and polymerization

Macromolecules, both natural and man-made, owe their great size to the fact that they are *polymers* (Greek: many parts); that is, each one is made up of a great many simpler units—identical to each other or at least chemically similar—joined together in a regular way. They are formed by a process we touched on earlier: **polymerization**, the *joining together of many small molecules to form very large molecules.* The simple compounds from which polymers are made are called *monomers.*

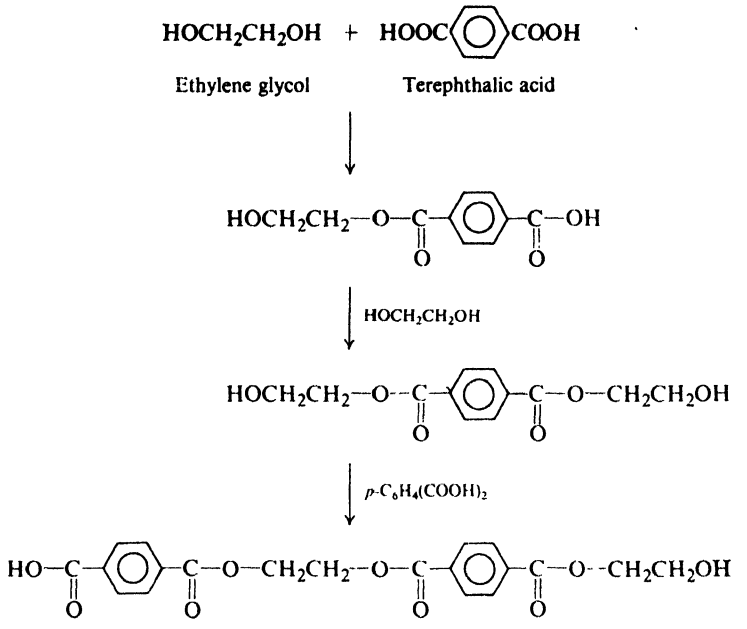
Polymers are formed in two general ways.

(a) In **chain-reaction polymerization**, there is a series of reactions each of which consumes a reactive particle and produces another, similar particle; each individual reaction thus depends upon the previous one. The reactive particles can be free radicals, cations, or anions. A typical example is the polymerization of ethylene

$$\text{Rad}\cdot + \text{CH}_2{=}\text{CH}_2 \longrightarrow \text{RadCH}_2\text{CH}_2\cdot \xrightarrow{\text{CH}_2=\text{CH}_2} \text{RadCH}_2\text{CH}_2\text{CH}_2\text{CH}_2\cdot \longrightarrow \textit{etc.}$$

(Sec. 6.19). Here the chain-carrying particles are free radicals, each of which adds to a monomer molecule to form a new, bigger free radical.

(b) In **step-reaction polymerization**, there is a series of reactions each of which is essentially independent of the preceding one; a polymer is formed simply because the monomer happens to undergo reaction at more than one functional group. A glycol, for example, reacts with a dicarboxylic acid to form an ester; but each moiety of the simple ester still contains a group that can react to generate another ester linkage and hence a larger molecule, which itself can react further, and so on.



There is an alternative, somewhat less meaningful system of classification: *addition polymerization*, in which molecules of monomer are simply added together; and *condensation polymerization*, in which monomer molecules combine with loss of some simple molecules like water. As it happens, the two systems almost exactly coincide; nearly all cases of chain-reaction polymerization involve addition polymerization; nearly all cases of step-reaction polymerization involve condensation polymerization. Indeed, some chemists use the term "addition polymerization" to *mean* polymerization via chain reactions.

Let us look first at chain-reaction polymerization, starting with the kind that involves free radicals.

Problem 32.1 Examine the structure of each of the following synthetic polymers. Tell what class of compound it belongs to and give structures of the most likely monomers.

(a) nylon 6,6 (fibers), $\sim\!\!-\overset{\overset{O}{\|}}{C}(CH_2)_4\overset{\overset{O}{\|}}{C}NH(CH_2)_6NH\overset{\overset{O}{\|}}{C}(CH_2)_4\overset{\overset{O}{\|}}{C}NH(CH_2)_6NH\sim\!\!-$

(b) nylon 6 (fibers), $\sim\!\!-\overset{\overset{O}{\|}}{C}CH_2(CH_2)_4NH\overset{\overset{O}{\|}}{C}CH_2(CH_2)_4NH\sim\!\!-$

(c) Carbowax (water-soluble wax), $\sim\!\!-OCH_2CH_2OCH_2CH_2OCH_2CH_2\sim\!\!-$

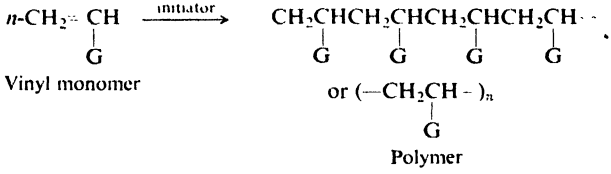
(d) Neoprene (oil-resistant elastomer), $\sim\!\!-CH_2\underset{\underset{Cl}{|}}{C}=CHCH_2CH_2\underset{\underset{Cl}{|}}{C}=CHCH_2\sim\!\!-$

(e) Saran (packaging film, seat covers), $\sim\!\!-CH_2CCl_2CH_2CCl_2\sim\!\!-$

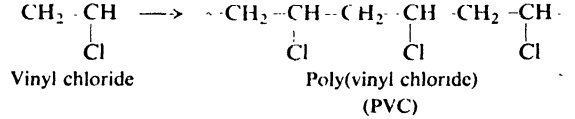
Problem 32.2 Answer the questions of Problem 32.1 for each of the following kinds of natural macromolecules: (a) a protein, p. 1151; (b) a nucleic acid, p. 1178; (c) starch (amylose), p. 1121; (d) cellulose, p. 1126.

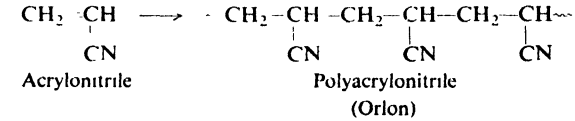
32.3 Free-radical vinyl polymerization

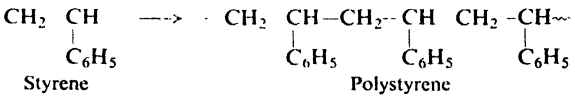
In Sec. 6.19 we discussed briefly the polymerization of ethylene and sub i-tuted ethylenes under conditions where free radicals are generated—typically in the presence of small amounts of an initiator, such as a peroxide. Reaction

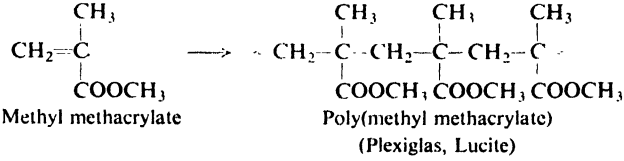


occurs at the doubly-bonded carbons of the vinyl groups—and is called *vinyl polymerization*. A wide variety of unsaturated monomers may be used, to yield polymers with different *pendant groups* (G) attached to the polymer backbone. For example:

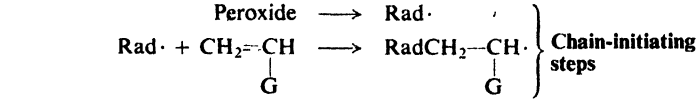








Polymerization involves addition of free radicals to the double bond of the monomer: addition, first, of the free radical generated from the initiator, and then of the growing polymer molecule. This is, of course, an example of chain-reaction polymerization.



(3) $\mathrm{RadCH_2{-}\underset{\displaystyle G}{\underset{|}{CH}}\cdot + CH_2{=}\underset{\displaystyle G}{\underset{|}{CH}} \longrightarrow RadCH_2{-}\underset{\displaystyle G}{\underset{|}{CH}}{-}CH_2{-}\underset{\displaystyle G}{\underset{|}{CH}}\cdot}$ **Chain-propagating steps**

then steps like (3) *repeated, until finally:*

(4) $\mathrm{2Rad(CH_2\underset{\displaystyle G}{\underset{|}{CH}})_nCH_2\underset{\displaystyle G}{\underset{|}{CH}}\cdot}$

combination

or $\mathrm{Rad(CH_2\underset{\displaystyle G}{\underset{|}{CH}})_nCH_2\underset{\displaystyle G}{\underset{|}{CH}}\text{-}\underset{\displaystyle G}{\underset{|}{CH}}CH_2(\underset{\displaystyle G}{\underset{|}{CH}}CH_2)_nRad}$

Chain-terminating steps

(5) $\mathrm{2Rad(CH_2\underset{\displaystyle G}{\underset{|}{CH}})_nCH_2\underset{\displaystyle G}{\underset{|}{CH}}\cdot}$

disproportionation

$\mathrm{Rad(CH_2\underset{\displaystyle G}{\underset{|}{CH}})_nCH_2\underset{\displaystyle G}{\underset{|}{CH_2}} + Rad(CH_2\underset{\displaystyle G}{\underset{|}{CH}})_nCH{=}\underset{\displaystyle G}{\underset{|}{CH}}}$

In each step the consumption of a free radical is accompanied by the formation of a new, bigger free radical. Eventually, the reaction chain is terminated by steps that consume but do not form free radicals: *combination* or *disproportionation* of two free radicals.

Problem 32.3 Free-radical polymerization of 1,3-butadiene gives molecules containing both the following units,

$$\sim CH_2CH{-}\underset{}{CHCH_2}\sim \quad \text{and} \quad \sim CH_2\underset{\displaystyle CH{=}CH_2}{\underset{|}{CH}}\sim$$

the exact proportions depending on the temperature. Account in detail for the formation of the two different units.

Problem 32.4 Polystyrene formed with isotopically labeled AIBN as initiator

$$\mathrm{(CH_3)_2\underset{\displaystyle CN}{\underset{|}{C}}{-}N{=}N{-}\underset{\displaystyle CN}{\underset{|}{C}}(CH_3)_2 \longrightarrow N_2 + 2(CH_3)_2\underset{\displaystyle CN}{\underset{|}{C}}\cdot}$$

Azoisobutyronitrile (AIBN)

was found to contain *two* initiator fragments per molecule. What termination reaction is indicated by this finding?

Added compounds can modify the polymerization process drastically. For example, in the presence of carbon tetrachloride, styrene undergoes polymerization at the same rate as in its absence, but the polystyrene obtained has a lower average molecular weight; furthermore, it contains small amounts of chlorine.

This is an example of **chain-transfer**, the termination of one polymerization chain (7) with the simultaneous initiation of another (8).

(6) $$\sim CH_2{-}\underset{\underset{Ph}{|}}{CH}\cdot + CH_2{=}\underset{\underset{Ph}{|}}{CH} \xrightarrow{\text{polymerization}} \sim CH_2{-}\underset{\underset{Ph}{|}}{CH}{-}CH_2{-}\underset{\underset{Ph}{|}}{CH}\cdot$$

(7) $$\sim CH_2{-}\underset{\underset{Ph}{|}}{CH}\cdot + CCl_4 \xrightarrow{\text{chain-transfer}} \sim CH_2{-}\underset{\underset{Ph}{|}}{CH}{-}Cl + \cdot CCl_3$$

(8) $$Cl_3C\cdot + CH_2{=}\underset{\underset{Ph}{|}}{CH} \longrightarrow Cl_3C{-}CH_2{-}\underset{\underset{Ph}{|}}{CH}\cdot \xrightarrow{\text{styrene}} \text{polymer}$$

Ordinarily a growing polystyrene radical adds (6) to styrene monomer to continue the reaction chain. Every so often, however, it abstracts an atom from the chain-transfer agent (7) to end the original polymerization chain and generate a new particle ($CCl_3\cdot$ in this case) that initiates a new polymerization chain (8). Since one reaction chain is replaced by another, the rate of polymerization is unaffected. Since the average number of chain-propagating steps in each reaction chain is reduced, the average molecular weight of the polymer is lowered. A transfer agent thus competes with the monomer for the growing radicals. The ratio of rate constants for (7) and (6), $k_{\text{transfer}}/k_{\text{polymerization}}$, is called the *transfer constant*; it is a measure of how effective the transfer agent is at lowering the molecular weight of the polymer.

Problem 32.5 For polymerization of styrene at 60°, the following chain-transfer constants have been measured. Account for the relative effectiveness of the members of each sequence.

(a) benzene 0.018, *tert*-butylbenzene 0.04, toluene 0.125, ethylbenzene 0.67, isopropylbenzene 0.86;
(b) *n*-heptane 0.42, 2-heptene 2.7;
(c) CCl_4 90, CBr_4 13,600.

An added compound may react with the growing free radical to generate a new free radical that is not reactive enough to add to monomer; a reaction chain is terminated but no new one is begun. Such a compound is, of course, an **inhibitor** (Sec. 2.14). Many amines, phenols, and quinones act as inhibitors. Although their exact mode of action is not understood, it seems clear that they are converted into free radicals that do not add to monomer; instead, they may combine or disproportionate, or combine with another growing radical to halt a second reaction chain.

$$\sim CH_2\cdot + \text{Inhibitor} \longrightarrow \sim CH_2CH_2 + \text{Inhibitor}\cdot$$

(G on the first CH_2; G on the second CH_2)

Unreactive: cannot initiate new chain

Since even traces of certain impurities, acting as chain-transfer agents or inhibitors, can interfere with the polymerization process, the monomers used are among the purest organic chemicals produced.

In an extreme case—if the alkene is of low reactivity and the transfer agent of high reactivity—chain transfer is so effective that there is *no* polymerization. Then

we observe simply addition of the "transfer agent" to the double bond, a reaction we encountered in Sec. 6.18. For example:

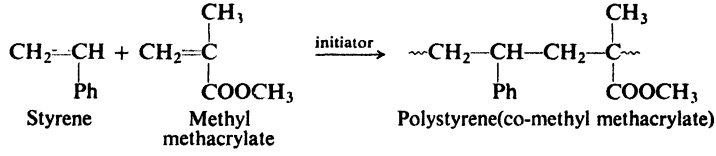
$$n\text{-}C_6H_{13}CH{=}CH_2 + CBr_4 \xrightarrow{\text{peroxides}} n\text{-}C_6H_{13}\underset{\displaystyle Br}{\underset{|}{C}}HCH_2CBr_3$$

Problem 32.6 (a) Chain transfer can cause *branching* of a polymer molecule. Show how this could happen. What is the chain-transfer agent? (b) Rather short branches (4 or 5 carbons) are attributed to "back-biting." What do you think is meant by this term? Show the chemical reactions probably involved.

32.4 Copolymerization

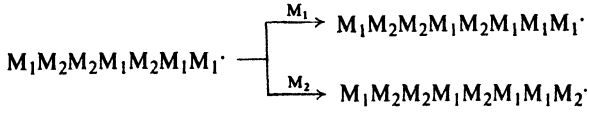
So far, we have discussed only polymerization of a single monomeric compound to form a *homopolymer*, a polymer made up—except, of course, at the two ends of the long molecule—of identical units.

Now, if a mixture of two (or more) monomers is allowed to undergo polymerization, there is obtained a **copolymer**: a polymer that contains two (or more) kinds of monomeric units in the same molecule. For example:



Through copolymerization there can be made materials with different properties than those of either homopolymer, and thus another dimension is added to the technology. Consider, for example, styrene. Polymerized alone, it gives a good electric insulator that is molded into parts for radios, television sets, and automobiles. Copolymerization with butadiene (30%) adds toughness; with acrylonitrile (20–30%) increases resistance to impact and to hydrocarbons; with maleic anhydride yields a material that, on hydrolysis, is water-soluble, and is used as a dispersant and sizing agent. The copolymer in which butadiene predominates (75% butadiene, 25% styrene) is an elastomer, and since World War II has been the principal rubber substitute manufactured in the United States.

Let us look more closely at the copolymerization process. Consider free radical vinyl polymerization of two monomers, M_1 and M_2. In each step the growing free

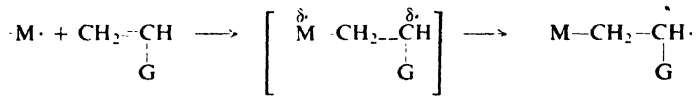


radical can react with either monomer to continue the reaction chain. What are the factors that determine *which* monomer it preferentially reacts with?

First, of course, there are the relative *concentrations* of the two monomers; the higher the concentration of a particular monomer, the greater its chance of

being incorporated into the chain, and the more abundant its units are in the final product.

Next, there are the relative *reactivities* of the monomers toward free radical addition; in general, the more reactive the monomer, the greater its chance of being incorporated into the polymer. We know that the reactivity of a carbon–carbon double bond toward free radical addition is affected by the stability of the new free radical being formed; factors that tend to stabilize the free radical product tend to stabilize the incipient free radical in the transition state, so that the more stable free radical tends to be formed faster. Now, stability of a free radical depends upon accommodation of the odd electron. The group G stabilizes the radical



by delocalization: the phenyl group in styrene, through formation of a benzylic radical; the vinyl group of 1,3-butadiene, through formation of an allylic radical; the $-COOCH_3$ group of methyl methacrylate, through formation of a radical in which acyl oxygen helps carry the odd electron. (*Problem:* Draw resonance structures to show how this last effect could arise.)

(We notice that the above discussion does not take into account the nature of the attacking radical, and hence would predict the same relative reactivities for a pair of alkenes toward all free radicals. We shall return to this point later.)

Now, let us see what kind of copolymer we would expect to get on the basis of what we have said so far. In the copolymerization of styrene (M_1) and butadiene (M_2), for example, reaction can proceed via either of two growing radicals: one ending in a styrene unit ($\cdots M_1\cdot$), or one ending in a butadiene unit ($\cdots M_2\cdot$). Either radical can add to either monomer, to form a copolymer with styrene and butadiene units distributed *randomly* along the molecule:

$$\cdots M_1M_2M_2M_1M_2M_1M_1\cdots$$ **Random copolymer**

With these particular monomers, copolymerization is in fact random. Now, toward either free radical type, it happens, butadiene is about 1.4 times as reactive as styrene, so that, if monomer concentrations were equal, butadiene units would tend to predominate in the product. Furthermore, since butadiene is consumed faster, the relative concentrations of monomers would change as reaction goes on, and so would the composition of the polymer being produced. These effects can be compensated for by adjusting the ratio of monomers fed into the reaction vessel; indeed, by control of the feed ratio, copolymers of any desired composition can be made.

Random copolymerization, of the kind observed for styrene and butadiene, is actually rather rare. In general, copolymerization shows, to a greater or lesser extent, a tendency to *alternation* of monomer units. An extreme case is that of

$$M_1M_2M_1M_2M_1M_2M_1M_2\cdots$$ **Alternating copolymer**

stilbene (1,2-diphenylethene) and maleic anhydride, which copolymerize with absolutely regular alternation of units; regardless of the feed ratio, a 50:50 copolymer is obtained.

How are we to account for this tendency toward alternation? It must mean that a growing radical ending in one unit tends to add to the *opposite* monomer.

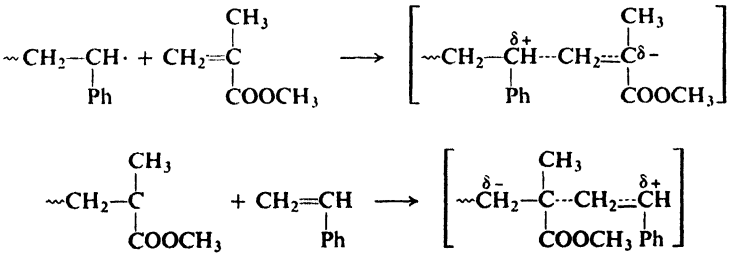
$$\sim M_1\cdot \xrightarrow{M_2} \sim M_1M_2\cdot \xrightarrow{M_1} \sim M_1M_2M_1\cdot \xrightarrow{M_2} \textit{etc.}$$

Clearly, the relative reactivity of a monomer *does* depend upon the nature of the radical that is attacking it. Maleic anhydride is much more reactive than stilbene toward radicals ending in a stilbene unit, and stilbene is much more reactive than maleic anhydride toward the other kind of radical. (Indeed, these two compounds, individually, undergo self-polymerization only with extreme difficulty.) A more modest—and more typical– tendency toward alternation is shown by styrene and methyl methacrylate. Here, toward either radical ($\sim M_1\cdot$) the "opposite" monomer (M_2) is about twice as reactive as the "same" monomer (M_1).

The alternating tendency in copolymerization was established on a quantitative basis by Frank R. Mayo (of the Stanford Research Institute) and Cheves Walling (of the University of Utah) while working in the laboratories of the U.S. Rubber Company. Their work was fundamental to the development of free radical chemistry: it showed clearly for the first time the dependence of reactivity on the nature of the attacking free radical, and led directly to the concept of *polar factors*, working not only in copolymerization and other additions of free radicals, but in free radical reactions of all kinds.

Basically, Mayo and Walling's interpretation was the following. Although free radicals are neutral, they have certain tendencies to gain or lose electrons, and hence they partake of the character of electrophilic or nucleophilic reagents. The transition states for their reactions can be polar, with the radical moiety acquiring a partial negative or positive charge at the expense of the substrate—the alkene, in the case of addition. In copolymerization, a substituent generally exerts the same polar effect—electron-withdrawing or electron-releasing—on a free radical as on the alkene (monomer) from which the free radical was derived. Electron-withdrawal makes a free radical electrophilic, but makes an alkene less able to supply the electrons which that radical is seeking. An electrophilic radical will, then, preferentially add to a monomer containing an electron-releasing group. In a similar way, a nucleophilic radical, containing an electron-releasing substituent, will seek out a monomer containing an electron-withdrawing substituent.

Styrene and methyl methacrylate tend to alternate because their substituents are of opposite polarity: in methacrylate the $COOCH_3$ group tends to withdraw electrons; in styrene the phenyl group tends (via resonance) to release electrons. The transition states for addition to the opposite monomers are thus stabilized:



Perhaps the most convincing evidence for the play of polar forces comes from copolymerization of a series of ring-substituted styrenes; here relative reactivities toward a variety of monomers not only fall into a pattern consistent with the familiar electronic effects of the substituents, but show the same quantitative relationships (the Hammett sigma-rho relationship, Sec. 18.11) as do *ionic* reactions: dissociation of carboxylic acids, for example, or hydrolysis of esters.

The concept of polar transition states in free-radical reactions has recently been questioned, at least for reactions in which hydrogen is abstracted—halogenation, for example. Here, it has been suggested, electron-withdrawing or electron-releasing groups affect reactivity simply by strengthening or weakening the bonds holding hydrogen in the substrate.

If we define polar effects on free-radical reactions as effects due to electron-withdrawal or electron-release—rather than to accommodation of the odd electron—then there is no doubt about their existence; it is the *interpretation* of such effects that is open to question.

We must realize that polar effects are *superimposed on* effects due to delocalization of the odd electron. Styrene and butadiene, for example, are highly reactive toward any radical since the transition state contains an incipient benzylic or allylic free radical. This high reactivity is modified—enhanced or lowered—by the demands of the particular attacking radical.

Problem 32.7 (a) Draw structures to account for the strong alternating tendency in copolymerization of butadiene (M_1) and acrylonitrile (M_2). (b) Toward $\sim M_1\cdot$ acrylonitrile is 2.5 times as reactive as butadiene, but toward $\sim M_2\cdot$ butadiene is *20 times as reactive* as acrylonitrile. How do you account for this contrast?

Copolymers can be made not just from two different monomers but from three, four, or even more. They can be made not only by free-radical chain reactions, but by any of the polymerization methods we shall take up: ionic, coordination, or step-reaction. The monomer units may be distributed in various ways, depending on the technique used. As we have seen, they may alternate along a chain, either randomly or with varying degrees of regularity. In *block copolymers*, sections made up of one monomer alternate with sections of another:

$$\sim M_1M_1M_1M_1M_1M_2M_2M_2M_2\sim$$ **Block copolymer**

In *graft copolymers*, a branch of one kind is grafted to a chain of another kind:

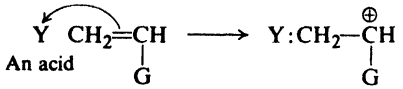
$$\begin{array}{l}\sim M_1M_1M_1M_1M_1M_1M_1\sim\\ \qquad\quad |\\ \qquad\quad M_2M_2M_2M_2\end{array}$$ **Graft copolymer**

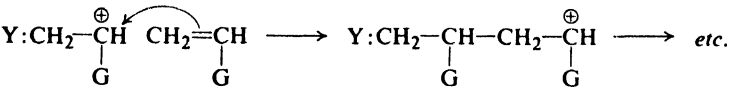
Problem 32.8 Graft copolymers can be made by each of the following processes. Show the chemistry most likely involved, and the structure of the product. (a) Polybutadiene is treated with styrene in the presence of a free-radical initiator. (b) Poly(vinyl chloride) is treated with methyl methacrylate in the presence of benzoyl peroxide, $(C_6H_5COO)_2$.

32.5 Ionic polymerization. Living polymers

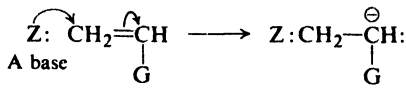
Chain-reaction polymerization can proceed with ions instead of free radicals as the chain-carrying particles: either cations or anions, depending on the kind of initiator that is used.

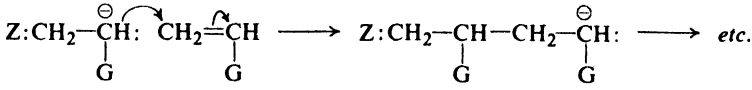
Cationic polymerization



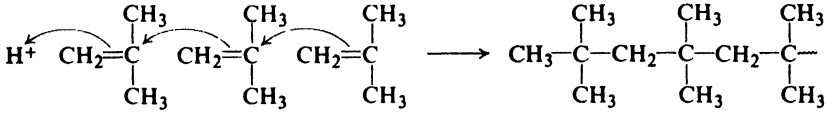


Anionic polymerization

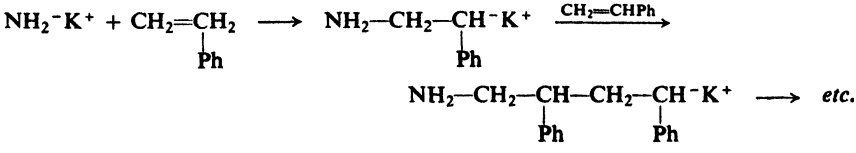


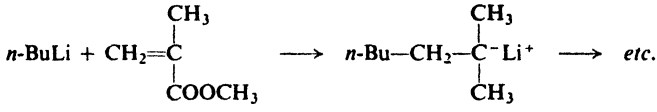


Cationic polymerization is initiated by *acids*. Isobutylene, for example, undergoes cationic polymerization to a tacky material used in adhesives. Copolymerization with a little isoprene gives *butyl rubber*, used to make automobile innertubes and tire liners. A variety of acids can be used; sulfuric acid; $AlCl_3$ or BF_3 plus a trace of water. We recognize this process as an extension of the dimerization discussed in Sec. 6.15.

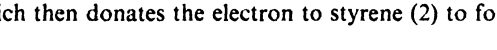


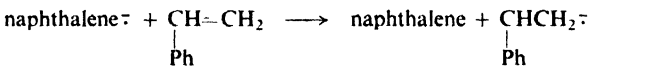
Anionic polymerization, as we might expect, is initiated by *bases*: $Li^+NH_2^-$, for example, or organometallic compounds like *n*-butyllithium. For example:

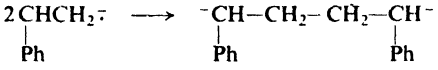




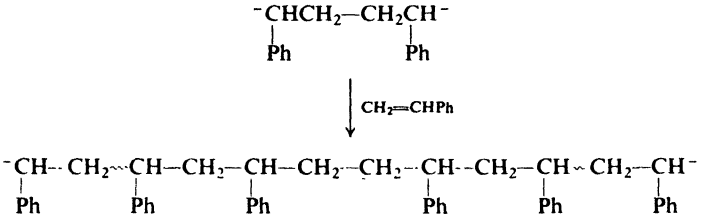
Active metals like Na or Li can be used; here the initiation becomes a little more complicated, as in the polymerization of styrene by the action of sodium metal and naphthalene. A sodium atom transfers an electron (1) to naphthalene to form a *radical-anion*, which then donates the electron to styrene (2) to form the styrene

(1) 

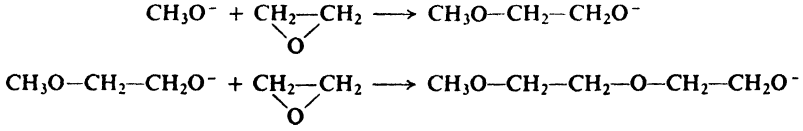
(2) 

(3) 

radical-anion. Like many other free-radicals, these dimerize (3). The resulting dianion is the true initiator, and begins to grow *at both ends*:



Anionic polymerization is not limited to the vinyl kind, involving addition to carbon–carbon double bonds. Ethylene oxide, for example, is converted by a small amount of base into a high-molecular-weight polyether.



Problem 32.9 The presence of methanol during the polymerization of ethylene oxide by sodium methoxide tends to lower the molecular weight of the product. (a) How do you think it does this? What process is this an example of ? (b) What product will be obtained in the presence of *much* methanol?

In 1956, Michael Szwarc (of the State University of New York at Syracuse)

reported the following observations. When a sample of styrene was treated with a little sodium naphthalene initiator, rapid polymerization took place and was complete within a few seconds. When more styrene was added, it, too, underwent polymerization; viscosity measurements showed that the molecular weight of the polystyrene was now *higher than before*. If, instead of the second batch of styrene, there was added butadiene, polymerization again took place; the product was found to contain no (homo)polystyrene—all styrene units were now part of a (block) copolymer with butadiene. How are these results to be interpreted?

So far, we have not mentioned chain-terminating steps for ionic polymerization. Such steps do exist. In cationic polymerization, for example, the growing carbonium ion can undergo either of two familiar reactions: ejection of a proton to form an alkene, or combination with an anion. But, particularly in the case of anionic polymerization, termination often involves reaction with an impurity or some other molecule not a necessary part of the polymerization system. Under carefully controlled conditions, there are *no* termination steps—or at least none that happen very fast. Reaction stops when all monomer is consumed, but the reaction mixture contains what Szwarc has named *living* polymer molecules. When these are "fed" additional monomer—either styrene or butadiene, for example—they continue to grow. They are not immortal, however, but can be "killed" by addition of a compound that reacts with carbanions: water, say. The generation of living polymers is of immense practical importance; it provides the best route to block copolymers, and permits the introduction of a variety of terminal groups.

Problem 32.10 Draw the structure of the product expected from the killing of living polystyrene by each of the following reagents: (a) water; (b) carbon dioxide, then water; (c) a small amount of ethylene oxide, then water; (d) a large amount of ethylene oxide, then water.

32.6 Coordination polymerization

When we speak of organic ions as chain-carriers, we realize, of course, that each of these must be balanced by an ion of opposite charge, a *counterion*. A growing carbanion, for example, has more or less closely associated with it a metallic cation like Li^+ or Na^+. Ion pairs—or even higher aggregates—can play important parts in polymerization. If the bonding between the reactive center and the metal is appreciably covalent, the process is called *coordination polymerization*. The growing organic chain is not a full-fledged anion, but its reactivity is due to its *anion-like character*.

Until 1953, almost all vinyl polymerization of commercial importance was of the free-radical type. Since that time, however, ionic polymerization, chiefly in the form of coordination polymerization, has revolutionized the field. Following discoveries by Karl Ziegler (of the Max Planck Institute for Coal Research) and by Guilio Natta (of the Polytechnic Institute of Milan)—who jointly received the Nobel Prize in 1963 for this work—catalysts have been developed that permit control of the polymerization process to a degree never before possible.

These Ziegler-Natta catalysts are complexes of transition metal halides with organometallic compounds: typically, triethylaluminum–titanium trichloride.

Reaction involves nucleophilic addition to the carbon–carbon double bond in the monomer, with the carbanion-like organic group of the growing organometallic compound as nucleophile. The transition metal may play a further role in complexing with the π electrons of the monomer and thus holding it at the reaction site. Polymerization thus amounts to insertion of alkene molecules into the bond between metal and the growing alkyl group. For example, in the formation of polyethylene:

$$M{-}CH_2CH_3 \xrightarrow[CH_2{=}CH_2]{} M{-}CH_2CH_2CH_2CH_3 \xrightarrow[CH_2{=}CH_2]{} M{-}CH_2CH_2CH_2CH_2CH_2CH_3 \xrightarrow[CH_2{=}CH_2]{} \textit{etc.}$$

Polymerization with Ziegler-Natta catalysts has two important advantages over free-radical polymerization: (a) it gives *linear* polymer molecules; and (b) it permits *stereochemical control.*

Polyethylene made by the free-radical process has a highly branched structure due to chain-transfer of a special kind, in which the transfer agent *is a polymer molecule*: a hydrogen atom is abstracted from somewhere along the polymer chain,

$$\sim CH_2\overset{\overset{H}{|}}{C}HCH_2CH_2\sim \xrightarrow{\sim CH_2CH_2\cdot} \sim CH_2CH_2{-}H + \sim CH_2\dot{C}HCH_2CH_2\sim \xrightarrow{CH_2{=}CH_2}$$

$$\sim CH_2\overset{\overset{CH_2}{|}\\ \overset{|}{CH_2}}{C}HCH_2CH_2\sim \longrightarrow \textit{etc.}$$

and a branch grows at the point of attack. In contrast, polyethylene made by the coordination process is virtually unbranched. These unbranched molecules fit together well, and the polymer is said to have a *high degree of crystallinity*; as a result, it has a higher melting point and higher density than the older (*low density*) polyethylene, and is mechanically much stronger. (We shall look at the crystallinity of polymers and its effect on their properties in Sec. 32.8.)

A second, far-reaching development in coordination polymerization is *stereochemical control.* Propylene, for example, could polymerize to any of three different arrangements (Fig. 32.1): *isotactic*, with all methyl groups on one side of an extended chain; *syndiotactic*, with methyl groups alternating regularly from side to side; and *atactic*, with methyl groups distributed at random. By proper choice of experimental conditions—catalyst, temperature, solvent—each of these stereoisomeric polymers has been made. Atactic polypropylene is a soft, elastic, rubbery material. Both isotactic and syndiotactic polypropylenes are highly crystalline: regularity of structure permits their molecules to fit together well. Over a billion pounds of isotactic polypropylene is produced every year, to be molded or extruded as sheets, pipes, and filaments; it is on its way to becoming one of the principal synthetic fibers.

Coordination catalysts also permit stereochemical control about the carbon–carbon double bond. By their use, isoprene has been polymerized to a material virtually identical with natural rubber: *cis*-1,4-polyisoprene. (See Sec. 8.25.)

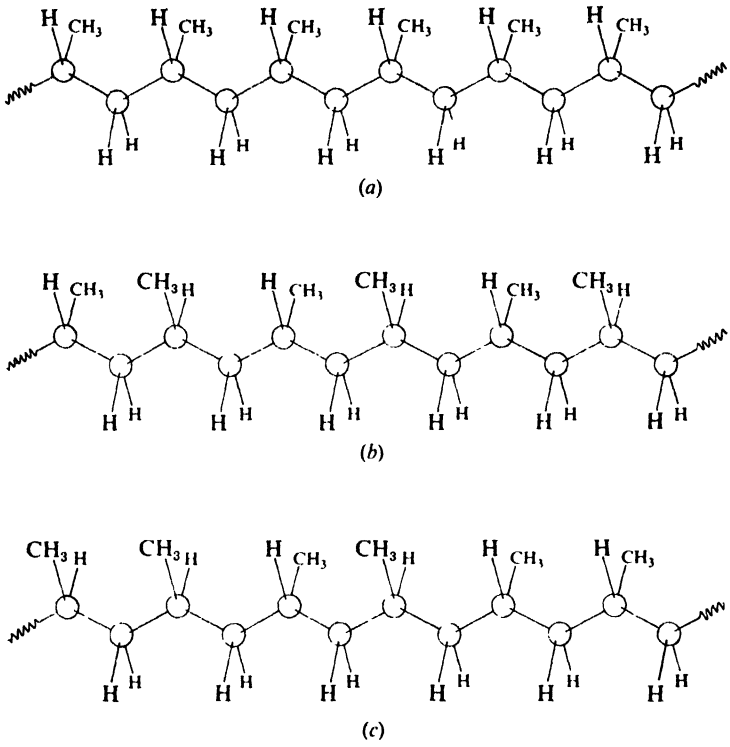


Figure 32.1. Polypropylene. (*a*) Isotactic. (*b*) Syndiotactic. (*c*) Atactic.

The Ziegler-Natta polymerization of ethylene can be adapted to make molecules of only modest size (C_6–C_{20}) and containing certain functional groups. If, for example, the metal-alkyls initially obtained are heated (in the presence of ethylene and a nickel catalyst), the hydrocarbon groups are displaced as straight-chain 1-alkenes of even carbon number. Large quantities of such alkenes in the C_{12}–C_{20} range are

$$M{-}(CH_2CH_2)_nCH_2CH_3 \xrightarrow[\text{heat}]{CH_2{=}CH_2,\ Ni} CH_2{=}CH{-}(CH_2CH_2)_{n-1}CH_2CH_3$$

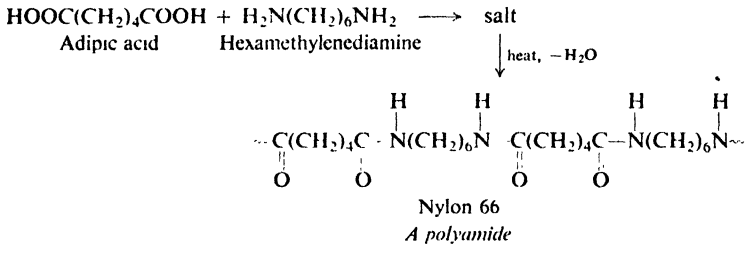
consumed in the manufacture of detergents (Sec. 33.5). Alternatively, the metal-alkyls can be oxidized by air to give straight-chain primary alcohols:

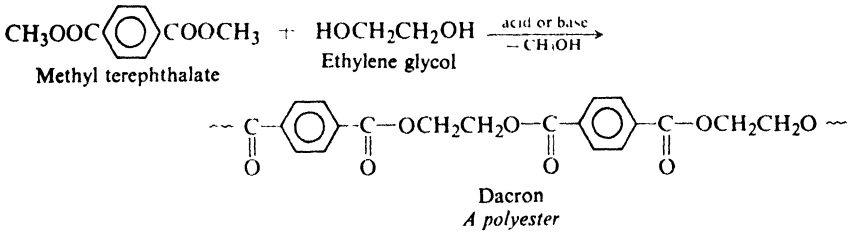
$$M{-}(CH_2CH_2)_nCH_2CH_3 \xrightarrow[30\text{–}95^\circ]{\text{air}} M{-}O(CH_2CH_2)_nCH_2CH_3 \xrightarrow[40^\circ]{H_2O,\ H_2SO_4} HO(CH_2CH_2)_nCH_2CH_3$$

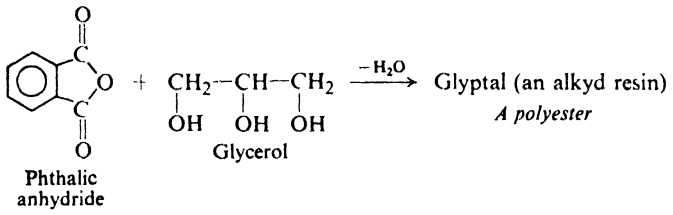
"A chemist setting out to build a giant molecule is in the same position as an architect designing a building. He has a number of building blocks of certain shapes and sizes, and his task is to put them together in a structure to serve a particular purpose. . . . What makes high polymer chemistry still more exciting just now is that almost overnight, within the last few years, there have come discoveries of new ways to put the building blocks together—discoveries which promise a great harvest of materials that have never existed on the earth." (Giulio Natta, *Scientific American*, September, 1957, p. 98.)

32.7 Step-reaction polymerization

Carboxylic acids react with amines to yield amides, and with alcohols to form esters. When an acid that contains more than one —COOH group reacts with an amine that contains more than one $-NH_2$ group, or with an alcohol that contains more than one —OH group, then the products are *polyamides* and *polyesters*. For example:







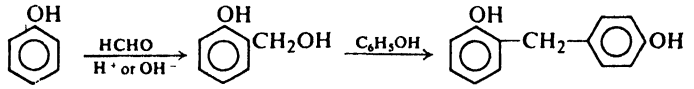
These are examples of step-reaction polymerization (Sec. 32.2). Here, reaction does not depend on chain-carrying free radicals or ions. Instead, the steps are essentially independent of each other; they just happen to involve more than one functional group in a monomer molecule.

If each monomer molecule contains just two functional groups, growth can occur in only two directions, and a *linear* polymer is obtained, as in nylon 66 or Dacron. But if reaction can occur at more than two positions in a monomer, there is formed a highly cross-linked *space network* polymer, as in Glyptal, an *alkyd resin*. Dacron and Glyptal are both polyesters, but their structures are quite different and, as we shall see, so are their uses.

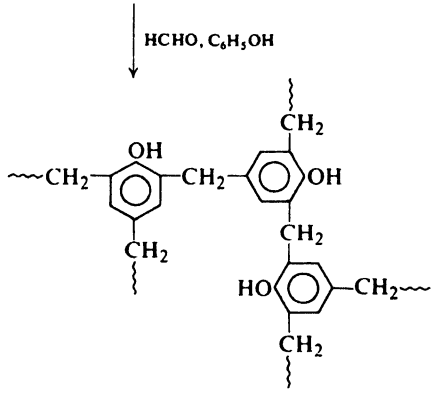
Problem 32.11 Work out a possible structure for an alkyd resin formed from phthalic anhydride and glycerol, considering the following points: (a) In the first stage

a linear polyester is formed. (Which hydroxyl groups are esterified more rapidly, primary or secondary?) (b) In the second stage these linear molecules are cross-linked to give a rather rigid network.

Step-reaction polymerization can involve a wide variety of functional groups and a wide variety of reaction types. Among the oldest of the synthetic polymers, and still extremely important, are those resulting from reaction between phenols and formaldehyde: the *phenol-formaldehyde resins* (Bakelite and related polymers). When phenol is treated with formaldehyde in the presence of alkali or acid, there is obtained a high molecular weight substance in which many phenol rings are held together by $-CH_2-$ groups:

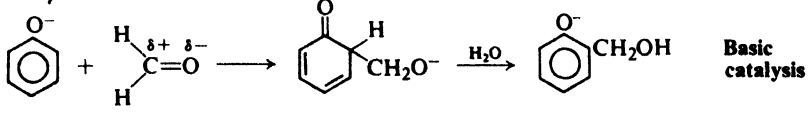


o-Hydroxymethylphenol

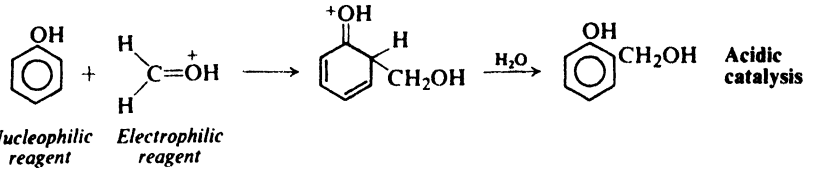


The stages involved in the formation of the polymer seem to be the following. First, phenol reacts with formaldehyde to form *o*- or *p*-hydroxymethylphenol. Hydroxymethylphenol then reacts with another molecule of phenol, with the loss of water, to form a compound in which two rings are joined by a $-CH_2-$ link. This process then continues, to yield a product of high molecular weight. Since three positions in each phenol molecule are susceptible to attack, the final product contains many cross-links and hence has a rigid structure.

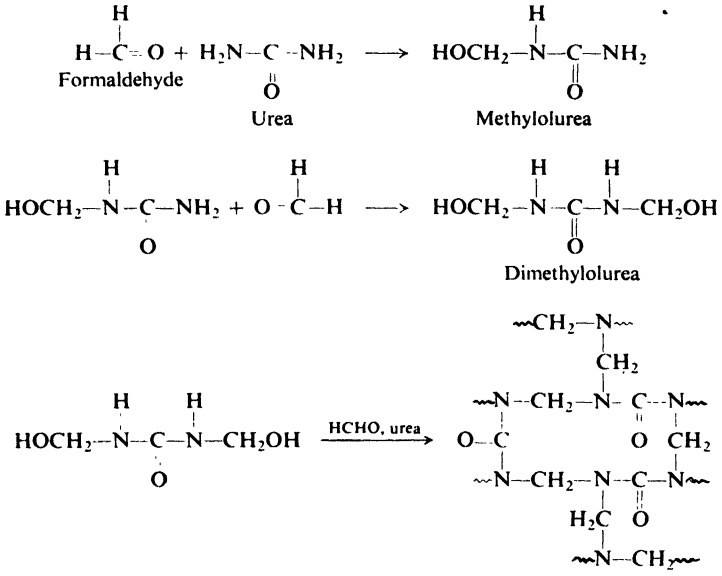
The first stage can be viewed as both electrophilic substitution on the ring by the electron-deficient carbon of formaldehyde, and nucleophilic addition of the aromatic ring to the carbonyl group. Base catalyzes reaction by converting phenol into the more reactive (more nucleophilic) phenoxide ion. Acid catalyzes reaction by protonating formaldehyde and increasing the electron deficiency of the carbonyl carbon.



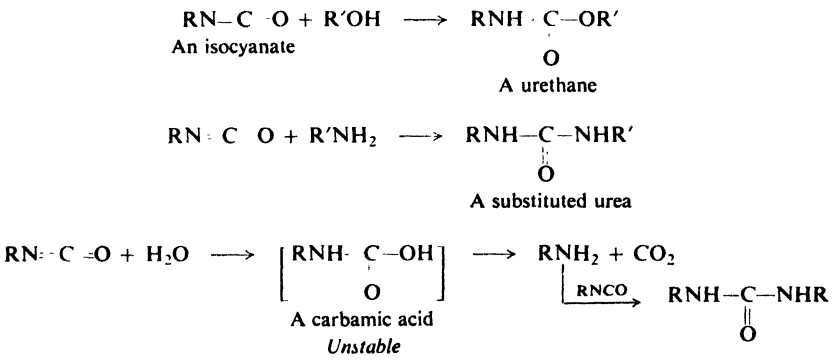
Nucleophilic reagent *Electrophilic reagent*



Urea reacts with formaldehyde to form the *urea-formaldehyde resins*, highly important in molded plastics. Here, too, a space-network polymer is formed.



Organic *isocyanates*, RNCO, undergo reactions of the following kinds (compare Sec. 20.23), all of which are used, in one way or another, in the synthesis of



polymers. Reaction of *di*hydroxy alcohols with *di*isocyanates gives the important polyurethanes.

Problem 32.12 Give the structure of the polymer expected from the reaction of ethylene glycol and 2,4-tolylene diisocyanate, $2,4\text{-}(OCN)_2C_6H_3CH_3$.

32.8 Structure and properties of macromolecules

The characteristic thing about macromolecules, we have said, is their great size. This size has little effect on chemical properties. A functional group reacts much as we would expect, whether it is in a big or little molecule: an ester is hydrolyzed, an epoxide undergoes ring-opening, an allylic hydrogen is susceptible to abstraction by free radicals.

Problem 32.13 Describe reagents and conditions—if any—that would be expected to cleave the natural polymers of Problem 32.2 (p. 1030) into monomers.

Problem 32.14 When poly(vinyl acetate) is treated with methanol (b.p. 65°) in the presence of a little sulfuric acid, a substance of b.p. 57° distills from the mixture, and a new polymer is left behind. (a) What reaction has taken place? What is the structure of the new polymer? Why must it be prepared in this indirect manner? (b) When this new polymer is treated with *n*-butyraldehyde in the presence of a little phosphoric acid, a third polymer is formed, Butvar, which is used in making safety glass. What reaction has taken place here, and what is the structure of Butvar?

It is in their physical properties that macromolecules differ from ordinary molecules, and it is on these that their special functions depend. To begin with, let us look at the property of *crystallinity*. In a crystalline solid, we know, the structural units—molecules, in the case of a non-ionic compound—are arranged in a very regular, symmetrical way, with a geometric pattern repeated over and over. If a long molecule is to fit into such a pattern, it cannot be looped and coiled into a random conformation, but must be extended in a regular zig-zag (see Fig. 32.2). This lack of randomness corresponds to an unfavorable entropy for the system (Sec. 18.11). On the other hand, the regularity and close fitting of the molecules in a crystal permits operation of strong intermolecular forces—hydrogen bonding, dipole–dipole attractions, van der Waals forces—which result in a favorable enthalpy (heat content). As we shall see, this tug-of-war between entropy and enthalpy is a key factor in determining the use to which a macromolecule can be put.

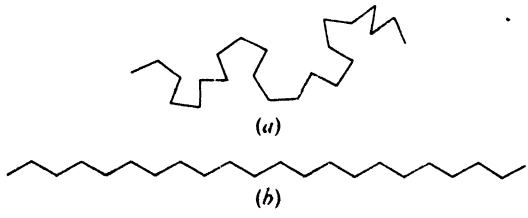


Figure 32.2. Long chain (*a*) in a random conformation, and (*b*) extended.

Now, in general, a high polymer does not exist entirely in crystalline form—not even a polymer whose regularity of molecular structure might be expected to permit this. The problem is the size of the molecule. As solidification begins, the viscosity of the material rises and the polymer molecules find it difficult to move about and arrange their long chains in the regular pattern needed for crystal formation. Chains become entangled; a change in shape of a chain must involve rotation about single bonds, and this becomes difficult because of hindrance to the swinging about of pendant groups. Polymers, then, form solids made up of regions of crystallinity, called *crystallites*, embedded in amorphous material. We speak of the *degree of crystallinity* of a polymer to mean the extent to which it is composed of crystallites.

Problem 32.15 Although both polymers are prepared by free-radical processes, poly(vinyl chloride) is amorphous and poly(vinylidene chloride) (Saran) is highly crystalline. How do you account for the difference? (Vinylidene chloride is 1,1-dichloroethene.)

Let us examine the various uses of polymers, and see how these depend on their structure—molecular and intermolecular.

Fibers are long, thin, threadlike bits of material that are characterized by great tensile (pulling) strength *in the direction of the fiber.* The natural fibers—cotton, wool, silk—are typical. Fibers are twisted into threads, which can then be woven into cloth, or embedded in plastic material to impart strength. The tensile strength can be enormous, some synthetic fibers rivalling—on a weight basis—steel.

The gross characteristics of fibers are reflected on the molecular level—the molecules, too, are long, thin, and threadlike. Furthermore, and most essential, they lie stretched out alongside each other, *lined up in the direction of the fiber.* The strength of the fiber resides, ultimately, in the strength of the chemical bonds of the polymer chains. The lining-up is brought about by *drawing*—stretching—the polymeric material. Once lined up, the molecules stay that way; the tendency to return to random looping and coiling is overcome by strong intermolecular attractions. In a fiber, enthalpy wins out over entropy. This high degree of molecular orientation is usually—although not always—accompanied by appreciable crystallinity.

The key requirements of a fiber are, then, a molecular shape—linear—that permits side-by-side alignment, and strong intermolecular forces to maintain this alignment. In addition, the intermolecular forces prevent "slipping" of one molecule past another. Now, what are these intermolecular forces?

The principal synthetic fibers are polyamides (the nylons), polyesters (Dacron, Terylene, Vycron), polyacrylonitrile ("acrylic fibers," Orlon, Acrilan), polyurethanes (Spandex, Vycra), and isotactic polypropylene. In nylon and polyurethanes, molecular chains are held to each other by hydrogen bonds (Fig. 32.3). In polyesters and polyacrylonitrile, the polar carbonyl and cyano groups lead to powerful dipole–dipole attractions. The stereoregular chains of isotactic polypropylene fit together so well that van der Waals forces are strong enough to maintain alignment.

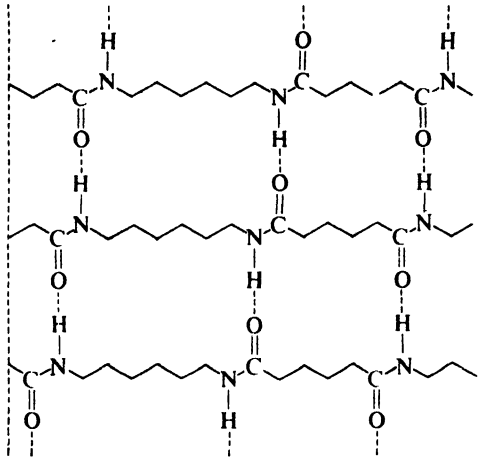
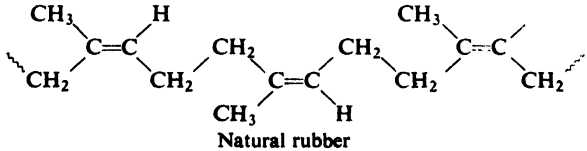


Figure 32.3. Hydrogen bonding in crystallites of nylon 66.

An **elastomer** possesses the high degree of elasticity that is characteristic of rubber: it can be greatly deformed—stretched to eight times its original length, for example—and yet return to its original shape. Here, as in fibers, the molecules are long and thin; as in fibers, they become lined up when the material is stretched. The big difference is this: when the stretching force is removed, the molecular chains of an elastomer do not remain extended and aligned, but return to their original random conformations favored by entropy. They do not remain aligned because the intermolecular forces necessary to hold them that way are weaker than in a fiber. In general, elastomers do not contain highly polar groups or sites for hydrogen bonding; the extended chains do not fit together well enough for van der Waals forces to do the job. In an elastomer entropy beats enthalpy.

One further requirement: the long chains of an elastomer must be connected to each other by occasional cross-links: enough of them to prevent slipping of molecules past one another; not so many as to deprive the chains of the flexibility that is needed for ready extension and return to randomness.

Natural rubber illustrates these structural requirements of an elastomer: long, flexible chains; weak intermolecular forces; and occasional cross-linking. Rubber is *cis*-1,4-polyisoprene. With no highly polar substituents, intermolecular attraction is largely limited to van der Waals forces. But these are weak because of



Natural rubber
All cis-configurations

the all-*cis* configuration about the double bond. Figure 32.4 compares the extended chains of rubber with those of its *trans* stereoisomer. As we can see, the *trans* configuration permits highly regular zig-zags that fit together well; the *cis* configuration does not. The all-*trans* stereoisomer occurs naturally as *gutta percha*; it is highly crystalline and non-elastic.

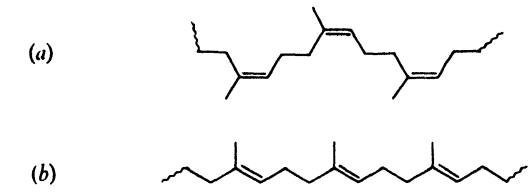


Figure 32.4. Extended chains of (*a*) natural rubber, *cis*-1,4-polyisoprene, and of (*b*) gutta percha, its *trans* stereoisomer.

Cross-linking in rubber, as we have seen (Sec. 8.25), is brought about by *vulcanizing*—heating with sulfur—which causes formation of sulfur bridges between molecules. This reaction involves reactive allylic positions, and thus depends on the double bond in the polymer.

Chief among the synthetic elastomers is SBR, a copolymer of butadiene (75%) and styrene (25%) produced under free-radical conditions; it competes with natural rubber in the main use of elastomers, the making of automobile tires. All-*cis* polybutadiene and polyisoprene can be made by Ziegler-Natta polymerization.

An elastomer that is entirely or mostly polydiene is, of course, highly unsaturated. All that is required of an elastomer, however, is enough unsaturation to permit cross-linking. In making butyl rubber (Sec. 32.5), for example, only 5% of isoprene is copolymerized with isobutylene.

Problem 32.16 (a) A versatile elastomer is obtained by Ziegler-Natta copolymerization of ethylene and propylene in the presence of a little diene, followed by vulcanization. How does the use of ethylene *and* propylene—instead of just one or the other—help to give the polymer elasticity?

(b) A similar copolymer can be made without the diene. This is cured by heating, not with sulfur, but with benzoyl peroxide. Why is this? What is the nature of the cross-links generated here?

Although enormous quantities of man-made fibers and elastomers are produced each year, the major consumption of synthetic polymers is as **plastics**, materials used in the form of sheets, pipes, films, and, most important of all, molded objects: toys and bottles; knobs, handles, and switches; dishes, fountain pens, toothbrushes; valves, gears, bearings; cases for radios and television sets; boats, automobile bodies, and even houses.

The molecular structure of plastics is of two general kinds: long molecules, either linear or branched; and space-network molecules.

The *linear* and *branched* polymers may be more or less crystalline, and include some of the materials also used as fibers: nylon, for example. They include the various polyalkenes we have mentioned: polyethylene, poly(vinyl chloride), polystyrene, etc. On heating, these polymers soften, and for this reason are called *thermoplastic*. It is in this softened state that they can be molded or extruded.

Space-network polymers (or *resins*) are highly cross-linked to form a rigid but irregular three-dimensional structure, as in phenol-formaldehyde or urea-formaldehyde resins. A sample of such material is essentially one gigantic molecule; heating does not soften it, since softening would require breaking of covalent bonds. Indeed, heating may cause formation of additional cross-links and thus make the material harder; for this reason, these polymers are called *thermosetting* polymers. This continuation of the polymerization process through heating is often coupled with the shaping of the product.

Certain linear, thermoplastic polymers are, like the space-network polymers, amorphous—and for basically the same reason. On cooling, their molecules form a rigid but irregular three-dimensional structure; they are held there, not by covalent cross-links, but by powerful dipole–dipole forces which lock the molecules into position before they can shake down into the regular arrangement required of a crystal. These materials are called *glasses*; poly(methyl methacrylate)—Plexiglas, Lucite—is the commonest one. Like ordinary (inorganic) glass, they lack crystalline planes for reflecting light, and are transparent. Like ordinary glass—and like the space-network polymers—they are brittle; when struck, these molecules cannot "give" with the blow through the sliding of crystalline planes over one another; they either resist—or break.

The rest of this book is devoted to organic compounds of biological importance. Many of these are macromolecules. We shall find that, just as the technological function of a macromolecule—fiber, elastomer, plastic—depends on its structure, so does the biological function: to hold the organism together, to nourish it, to control it, to allow it to reproduce itself.

PROBLEMS

1. Account for the fact that, whatever the mechanism—free-radical, cationic, anionic—vinyl polymerization gives products with almost exclusively "head-to-tail" arrangement of units.

2. Like other oxygen-containing compounds, alcohols dissolve in cold concentrated H_2SO_4 (Sec. 6.30). In the case of some secondary and tertiary alcohols, dissolution is followed by the gradual separation of an insoluble liquid of high boiling point. How do you account for this behavior?

3. Isobutylene does not give the kinds of stereoisomeric polymers (isotactic, etc.) that propylene does. Why not? What can you say about 1-butene?

4. Formaldehyde is polymerized by the action of a strong base like sodium methoxide. Suggest a mechanism for the process, and a structure for the polymer. To what general class of organic reactions does this polymerization belong?

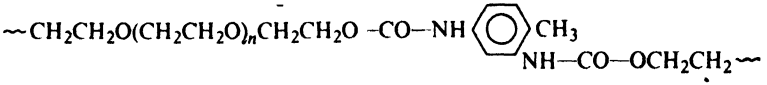
5. A simple process for recycling polyurethanes has been developed by the Ford Motor Company. Can you suggest a way to accomplish this? What products would you expect to obtain?

6. Suggest an explanation for the following order of reactivity toward the addition of $BrCCl_3$ in the presence of peroxides: $C_6H_5CH{=}CH_2$ over 100, 1-octene 1.0, $C_6H_5CH_2CH{=}CH_2$ 0.7, $ClCH_2CH{=}CH_2$ 0.5, $Cl_3CCH_2CH{=}CH_2$ 0.3.

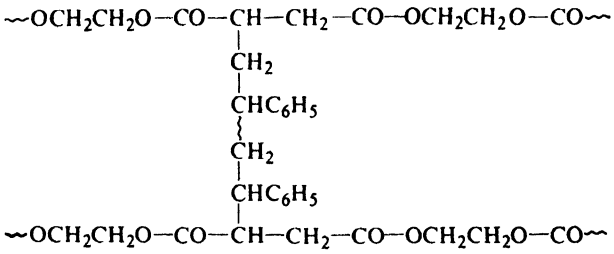
7. Account for each of the following observations. (a) In the presence of peroxides, CCl_4 reacts with 1-octene, $RCH{=}CH_2$, to give not only the 1:1 adduct, $RCHClCH_2CCl_3$, but also the 2:1 adduct, $RCHClCH_2CH(R)CH_2CCl_3$. (b) In contrast, CBr_4 adds to the 1-octene, to give *only* the 1:1 product. (c) Styrene reacts with peroxides in the presence of CCl_4 to give only polymer.

8. Outline all steps in a possible synthesis from non-polymeric starting materials of each of the following polymers.

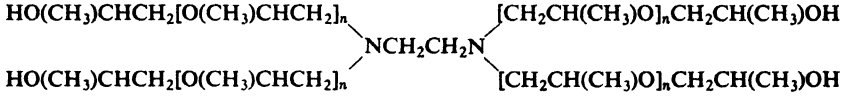
(a) Elastic fibers, used in girdles and bathing suits (Spandex, Lycra).



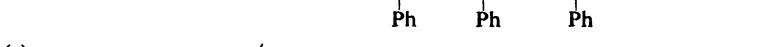
(b) A polyester resin, used in making pipe, boats, automobile bodies, etc.



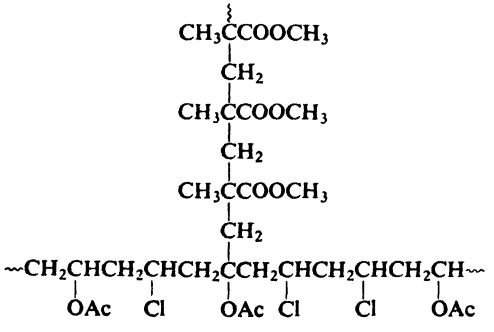
(c) A surface-active polymer.



(d)



(e)

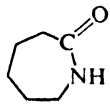


9. Treatment of β-propiolactone with base gives a polymer. Give a likely structure for this polymer, and show a likely mechanism for the process. Is this an example of chain-reaction or step-reaction polymerization?

10. When styrene is treated with KNH_2 in liquid ammonia, the product is a dead polymer that contains one $-NH_2$ group per molecule and no unsaturation. Suggest a termination step for the process.

11. When poly(vinyl acetate) was hydrolyzed, and the product treated with periodic acid and then re-acetylated, there was obtained poly(vinyl acetate) of lower molecular weight than the starting material. What does this indicate about the structure of the original polymer? About the polymerization process?

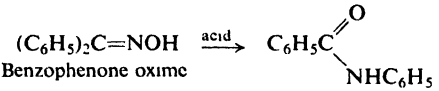
12. (a) What is the structure of nylon 6, made by alkaline polymerization of caprolactam?



Caprolactam

(b) Suggest a mechanism for the process. Is polymerization of the chain-reaction or step-reaction type?

13. In the *Beckmann rearrangement* (Problem 6, p. 919) oximes are converted into amides by the action of acids. For example:



Benzophenone oxime — Benzanilide

Caprolactam (preceding problem) can be made by the Beckmann rearrangement. With what ketone must the process start?

14. Fibers of very high tensile strength ("high-modulus fibers") have been made by reactions like the one between terephthalic acid and *p*-phenylenediamine, $p\text{-}C_6H_4(NH_2)_2$. Of key importance is the isomer composition of the monomers: the more exclusively *para*, the higher the melting point and the lower the solubility of the polymer, and the stronger the fibers. How do you account for this effect?

15. Evidence of many kinds shows that the metal–carbon bond in compounds like *n*-butyllithium is covalent, although highly polar. Yet living polystyrene solutions, which are colored, have virtually identical spectra whether the metal involved is sodium, potassium, cesium, or lithium. Can you suggest an explanation for this?

16. (a) When the alkane 2,4,6,8-tetramethylnonane was synthesized by an unambiguous method (Problem 13 (l), p. 544), there was obtained a product which was separated by gas chromatography into two components, A and B. The two components had identical mol. wt. and elemental composition, but different m.p., b.p., and infrared and nmr spectra. Looking at the structure of the expected product, what are these two components?

(b) When the same synthesis was carried out starting with an optically active reactant, compound B was obtained in optically active form, but A was still inactive. What is the structure of A? Of B?

(c) The nmr and infrared spectra of A and B were compared with the spectra of isotactic and syndiotactic polypropylenes (Fig. 32.1, p. 1041). With regard to their spectra, A showed a marked resemblance to one of the polymers, and B showed a marked resemblance to the other. It was concluded that the results "confirm the structures originally assigned [by Natta, p. 1039] for the two crystalline polymers of propylene." Which polymer did A resemble? Which polymer did B resemble?

17. Material similar to foam rubber can be made by the following sequence:

adipic acid + excess ethylene glycol $\longrightarrow$ C

C + excess $p\text{-OCN}\text{---}C_6H_4\text{---}C_6H_4\text{---}\text{NCO-}p$ $\longrightarrow$ D

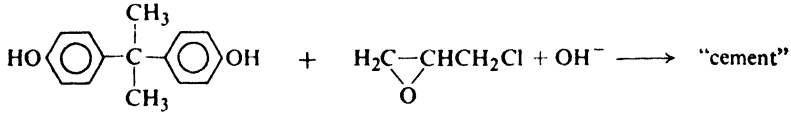
D + limited H_2O $\longrightarrow$ E

Write equations for all steps, and show structures for C, D, and E. Be sure to account for the cross-linking in the final polymer, and its *foamy* character. (***Remember:*** A foam is a dispersion of a gas in a solid.)

18. In the presence of benzoyl peroxide, allyl acetate gives poor yields of polymer of low molecular weight. The deuterium-labeled ester, $CH_2{=}CHCD_2OAc$, polymerizes 2 to 3 times as fast as the ordinary ester, and gives polymer of about twice the molecular weight. How do you account for these facts?

19. Linseed oil and tung oil, important constituents of paints, are esters (Sec. 33.6) derived from acids that contain two or three double bonds per molecule: 9,12-octadecadienoic acid, for example. On exposure to air, paint forms a tough protective film; oddly enough, after the initial rapid evaporation of solvent, this "drying" of paint is accompanied by a *gain* in weight. What kind of process do you think is involved? Be as specific as you can be.

20. To use an epoxy cement, one mixes the fluid "cement" with the "hardener," applies the mixture to the surfaces being glued together, brings them into contact, and waits for hardening to occur. The fluid cement is a low-molecular-weight polymer prepared by the following reaction:



2,2-Bis(*p*-hydroxyphenyl)propane ("Bisphenol A") | Epichlorohydrin *Excess* | *Contains no chlorine*

The hardener can be any of a number of things: $NH_2CH_2CH_2NHCH_2CH_2NH_2$, diethylenetriamine, for example.

(a) What is the structure of the fluid cement, and how is it formed? What is the purpose of using *excess* epichlorohydrin? (b) What happens during hardening? What is the structure of the final epoxy resin? (c) Suggest a method of making bisphenol A, starting from phenol.

21. Poly(methyl methacrylate) was prepared in two different ways: polymer F, with initiation by benzoyl peroxide at 100°; polymer G, with initiation by *n*-butyllithium at −62°. Their nmr spectra were, with considerable simplification, as follows:

F *a* singlet, δ 1.10
b singlet, δ 2.0
c singlet, δ 3.58
approximate area ratios, $a:b:c = 3:2:3$

G *a* singlet, δ 1.33
b doublet, δ 1.7
c doublet, δ 2.4
d singlet, δ 3.58
approximate area ratios, $a:b:c:d = 3:1:1:3$

Account in detail for the difference in spectra. What, essentially, is polymer F? Polymer G?

PART III

Biomolecules

Chapter 33 Fats

33.1 The organic chemistry of biomolecules

The study of biology at the molecular level is called biochemistry. It is a branch of biology, but it is equally a branch of organic chemistry. Most of the molecules involved, the *biomolecules*, are bigger and more complicated than the ones we have so far studied, and their environment—a living organism—is a far cry from the stark simplicity of the reaction mixture of the organic chemist. But the physical and chemical properties of these compounds depend on molecular structure in exactly the same way as do the properties of other organic compounds.

The detailed chemistry of biological processes is vast and complicated, and is beyond the scope of this book; indeed, the study of biochemistry must be *built upon* a study of the fundamentals of organic chemistry. We can, however, attempt to close the gap between the subject "organic chemistry" and the subject "biochemistry."

In the remaining chapters of this book, we shall take up the principal classes of biomolecules: fats, carbohydrates, proteins, and nucleic acids. Our chief concern will be with their structures—since structure is fundamental to everything else—and with the methods used to determine these structures. Because biomolecules are big ones, we shall encounter structure on several levels: first, of course, the *sequence of functional groups* and the *configuration* at any chiral centers or double bonds; then, *conformation*, with loops, coils, and zig-zags on a grander scale than anything we have seen yet; finally, the arrangement of *collections of molecules*, and even of collections of these collections. We shall see remarkable effects due to our familiar intermolecular forces: operating between biomolecules; between biomolecules—or *parts* of them—and the solvent; between different parts of the same biomolecule.

We shall study the chemical properties of these compounds observed in the test tube, since these properties must lie behind the reactions they undergo in living organisms. In doing this, we shall reinforce our knowledge of basic organic

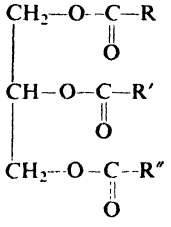
chemistry by applying it to these more complex substances. Finally, mostly in the last chapter, we shall look—very briefly—at a few biochemical processes, just to catch a glimpse of the ways in which molecular structure determines biological behavior.

33.2 Occurrence and composition of fats

Biochemists have found it convenient to define one set of biomolecules, the *lipids*, as substances, insoluble in water, that can be extracted from cells by organic solvents of low polarity like ether or chloroform. This is a catch-all sort of definition, and lipids include compounds of many different kinds: steroids (Sec. 15.16), for example, and terpenes (Sec. 8.26). Of the lipids, we shall take up only the *fats* and certain closely related compounds. These are not the only important lipids—indeed, every compound in an organism seems to play an important role, if only as an unavoidable waste product of metabolism—but they are the most abundant.

Fats are the main constituents of the storage fat cells in animals and plants, and are one of the important food reserves of the organism. We can extract these animal and vegetable fats—liquid fats are often referred to as *oils*—and obtain such substances as corn oil, coconut oil, cottonseed oil, palm oil, tallow, bacon grease, and butter.

Chemically, fats are carboxylic esters derived from the single alcohol, glycerol, $HOCH_2CHOHCH_2OH$, and are known as *glycerides*. More specifically, they are *triacylglycerols*. As Table 33.1 shows, each fat is made up of glycerides derived



A triacylglycerol
(A glyceride)

from many different carboxylic acids. The proportions of the various acids vary from fat to fat; each fat has its characteristic composition, which does not differ very much from sample to sample.

With only a few exceptions, the fatty acids are all straight-chain compounds, ranging from three to eighteen carbons; except for the C_3 and C_5 compounds, only acids containing an even number of carbons are present in substantial amounts. As we shall see in Sec. 37.6, these even numbers are a natural result of the biosynthesis of fats: the molecules are built up two carbons at a time from acetate units, in steps that closely resemble the malonic ester synthesis of the organic chemist (Sec. 26.2).

Problem 33.1 *n*-Heptadecane is the principal *n*-alkane found both in a 50 million-year-old shale and in the blue-green algae, primitive organisms still existing. When blue-green algae were grown on a medium containing stearic-18-^{14}C acid, essentially all the radioactivity that was not left in unconsumed stearic acid was found in *n*-hepta-

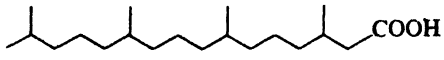
Table 33.1 FATTY ACID COMPOSITION OF FATS AND OILS

Fat or Oil	Saturated Acids, %							Unsaturated Acids, %					
								Enoic				Dienoic	Trienoic
	C_8	C_{10}	C_{12}	C_{14}	C_{16}	C_{18}	$>C_{18}$	$<C_{16}$	C_{16}	C_{18}	$>C_{18}$	C_{18}	C_{18}
Beef tallow			0.2	2–3	25–30	21–26	0.4–1	0.5	2–3	39–42	0.3	2	
Butter	1–2[a]	2–3	1–4	8–13	25–32	8–13	0.4–2	1–2	2–5	22–29	0.2–1.5	3	
Coconut	5–9	4–10	44–51	13–18	7–10	1–4				5–8	0–1	1–3	
Corn				0–2	8–10	1–4			1–2	30–50	0–2	34–56	
Cottonseed				0–3	17–23	1–3				23–44	0–1	34–55	
Lard				1	25–30	12–16		0.2	2–5	41–51	2–3	3–8	
Olive			0–1	0–2	7–20	1–3	0–1		1–3	53–86	0–3	4–22	
Palm				1–6	32–47	1–6				40–52		2–11	
Palm kernel	2–4	3–7	45–52	14–19	6–9	1–3	1–2		0–1	10–18		1–2	
Peanut				0.5	6–11	3–6	5–10		1–2	39–66		17–38	
Soybean				0.3	7–11	2–5	1–3		0–1	22–34		50–60	2–10
											C_{20} $>C_{20}$		
Cod liver				2–6	7–14	0–1		0–2	10–20	25–31	25–32 10–20		
Linseed				0.2	5–9	4–7	0.5–1			9–29		8–29[b]	45–67[c]
Tung										4–13		8–15	78–82[d]

[a] 3–4% C_4, 1–2% C_6.
[b] Linoleic acid, *cis,cis*-9,12-octadecadienoic acid.
[c] Linolenic acid, *cis,cis,cis*-9,12,15-octadecatrienoic acid.
[d] Eleostearic acid, *cis,trans,trans*-9,11,13-octadecatrienoic acid, and 3–6% saturated acids.

decane. By what kind of chemical reaction is the hydrocarbon evidently produced? Of what geological significance is this finding?

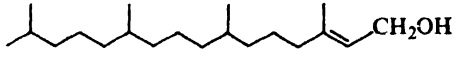
Problem 33.2 (a) Acetate is not the only building block for the long chains of lipids. From a 50 million-year-old shale (see Problem 33.1)—as well as from modern organisms—there has been isolated 3,7,11,15-tetramethylhexadecanoic acid,



3,7,11,15-Tetramethylhexadecanoic acid

What familiar structural unit occurs here?

(b) The long side chain of chlorophyll (p. 1004) is derived from the alcohol *phytol*, which is cis-7(R),11(R)-3,7,11,15-tetramethyl-2-hexadecen-1-ol. The acid in (a)



cis-7(R)-11(R)-3,7,11,15-Tetramethyl-2-hexadecen-1-ol
Phytol

was found to be a mixture of two diastereomers: the 3(S),7(R),11(R) and 3(R),7(R),-11(R). Of what biogenetic significance is this finding?

Besides saturated acids, there are unsaturated acids containing one or more double bonds per molecule. The most common of these acids are:

$CH_3(CH_2)_7CH{=}CH(CH_2)_7COOH$
Oleic acid
(*cis*-isomer)

$CH_3(CH_2)_4CH{=}CHCH_2CH{=}CH(CH_2)_7COOH$
Linoleic acid
(*cis,cis*-isomer)

$CH_3CH_2CH{=}CHCH_2CH{=}CHCH_2CH{=}CH(CH_2)_7COOH$
Linolenic acid
(*cis,cis,cis*-isomer)

The configuration about these double bonds is almost invariably *cis*, rather than the more stable *trans*.

Unsaturation *with this particular stereochemistry* has an effect that is seemingly trivial but is actually (Sec. 33.8) of vital biological significance: it lowers the melting point. In the solid phase, the molecules of a fat fit together as best they can; the closer they fit, the stronger the intermolecular forces, and the higher the melting point. Saturated acid chains are extended in a linear fashion—with, of course, the zig-zag due to the tetrahedral bond angles—and fit together rather well. *trans*-Unsaturated acid chains can be similarly extended to linear conformations that match saturated chains rather well (Fig. 33.1) But *cis*-unsaturated acid chains have a *bend* at the double bond, and fit each other—and saturated chains—badly. The net result is that *cis* unsaturation lowers the melting point of fat.

While we synthesize fats in our own bodies, we also eat fats synthesized in plants and other animals; they are one of the three main classes of foods, the others being carbohydrates (Chap. 35) and proteins (Chap. 36). Fats are used in enormous amounts as raw materials for many industrial processes; let us look at some of these before we turn our attention to some close relatives of the fats.

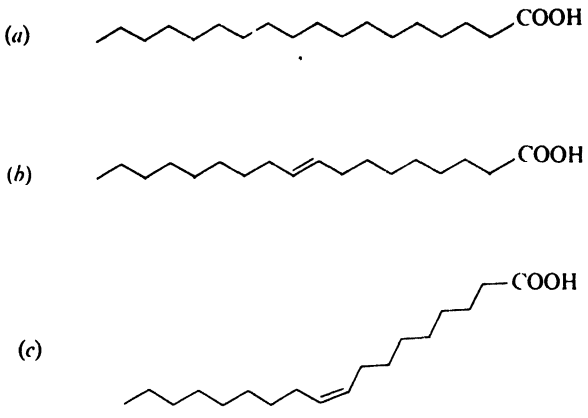
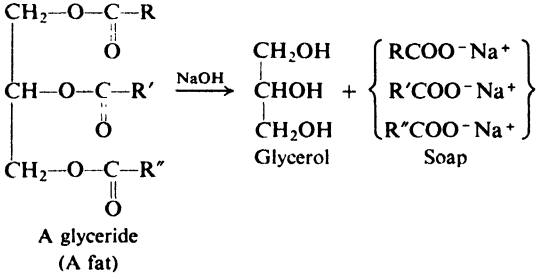


Figure 33.1. Extended chains of fatty acids: (*a*) saturated, (*b*) *trans*-unsaturated, (*c*) *cis*-unsaturated. Note bend in (*c*).

33.3 Hydrolysis of fats. Soap. Micelles

The making of soap is one of the oldest of chemical syntheses. (It is not nearly so old, of course, as the production of ethyl alcohol; man's desire for cleanliness is much newer than his desire for intoxication.) When the German tribesmen of Caesar's time boiled goat tallow with potash leached from the ashes of wood fires, they were carrying out the same chemical reaction as the one carried out on a tremendous scale by modern soap manufacturers: *hydrolysis of glycerides*. Hydrolysis yields salts of the carboxylic acids, and glycerol, $CH_2OHCHOHCH_2OH$.



Ordinary soap today is simply a mixture of sodium salts of long-chain fatty acids. It is a mixture because the fat from which it is made is a mixture, and for washing our hands or our clothes a mixture is just as good as a single pure salt. Soap may vary in composition and method of processing: if made from olive oil, it is *Castile soap*; alcohol can be added to make it transparent; air can be beaten in to make it float; perfumes, dyes, and germicides can be added; if a potassium salt (instead of a sodium salt), it is *soft soap*. Chemically, however, soap remains pretty much the same, and does its job in the same way.

We might at first expect these salts to be water-soluble—and, indeed, one can prepare what are called "soap solutions." But these are not true solutions, in which solute molecules swim about, separately and on their own. Instead, soap is

dispersed in spherical clusters called **micelles**, each of which may contain hundreds of soap molecules. A soap molecule has a polar end, COO^-Na^+, and a non-polar end, the long carbon chain of 12 to 18 carbons. The polar end is water-soluble, and is said to be *hydrophilic* (water-loving). The non-polar end is water-insoluble, and is said to be *hydrophobic* (water-fearing); it is, of course, soluble in non-polar solvents. Molecules like these are called *amphipathic*: they have both polar and non-polar ends and, in addition, are big enough for each end to display its own solubility behavior. In line with the rule of "like dissolves like," each non-polar end seeks a non-polar environment; in this situation, the only such environment about is the non-polar ends of other soap molecules, which therefore huddle together in the center of the micelle (Fig. 33.2). The polar ends project outward into the polar

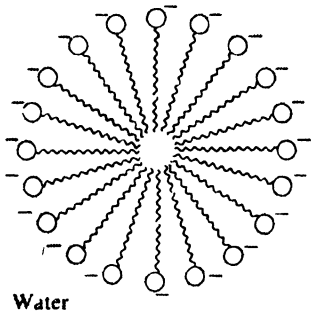


Figure 33.2. Soap micelle. Non-polar hydrocarbon chains "dissolve" in each other. Polar $-COO^-$ groups dissolve in water. Similarly charged micelles repel each other.

solvent, water. Negatively charged carboxylate groups stud the surface of the micelle, and it is surrounded by an ionic atmosphere. Repulsion between similar charges keeps the micelles dispersed.

Now, how does a soap clean? The problem in cleansing is the fat and grease that make up and contain the dirt. Water alone cannot dissolve these hydrophobic substances; oil droplets in contact with water tend to coalesce so that there is a water layer and an oil layer. But the presence of soap changes this. The non-polar ends of soap molecules dissolve in the oil droplet, leaving the carboxylate ends projecting into the surrounding water layer. Repulsion between similar charges keeps the oil droplets from coalescing; a stable emulsion of oil and water forms, and can be removed from the surface being cleaned. As we shall see, this emulsifying, and hence cleansing, property is not limited to carboxylate salts, but is possessed by other amphipathic molecules (Sec. 33.5).

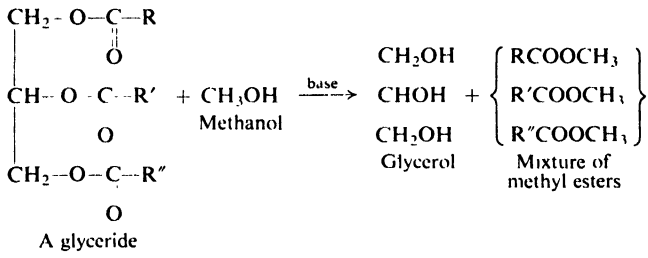
Hard water contains calcium and magnesium salts, which react with soap to form insoluble calcium and magnesium carboxylates (the "ring" in the bathtub).

33.4 Fats as sources of pure acids and alcohols

Treatment of the sodium soaps with mineral acid (or hydrolysis of fats under acidic conditions) liberates a mixture of the free carboxylic acids. In recent years,

fractional distillation of these mixtures has been developed on a commercial scale to furnish individual carboxylic acids of over 90% purity.

Fats are sometimes converted by transesterification into the methyl esters of carboxylic acids; the glycerides are allowed to react with methanol in the presence of a basic or acidic catalyst. The mixture of methyl esters can be separated by



fractional distillation into individual esters, which can then be hydrolyzed to individual carboxylic acids of high purity. Fats are thus the source of straight-chain acids of even carbon number ranging from six to eighteen carbons.

Alternatively, these methyl esters, either pure or as mixtures, can be catalytically reduced to straight-chain primary alcohols of even carbon number, and from these can be derived a host of compounds (as in Problem 18.10, p. 604). Fats thus provide us with long straight-chain units to use in organic synthesis.

33.5 Detergents

Of the straight-chain primary alcohols obtained from fats—or in other ways (Sec. 32.6)—the C_8 and C_{10} members are used in the production of high-boiling esters used as *plasticizers* (e.g., octyl phthalate). The C_{12} to C_{18} alcohols are used in enormous quantities in the manufacture of *detergents* (cleansing agents).

Although the synthetic detergents vary considerably in their chemical structure, the molecules of all of them have one common feature, a feature they share with ordinary soap: they are amphipathic, and have a large non-polar hydrocarbon end that is oil-soluble, and a polar end that is water-soluble. The C_{12} to C_{18} alcohols are converted into the salts of alkyl hydrogen sulfates. For example:

$$\underset{\text{Lauryl alcohol}}{n\text{-}C_{11}H_{23}CH_2OH} \xrightarrow{H_2SO_4} \underset{\text{Lauryl hydrogen sulfate}}{n\text{-}C_{11}H_{23}CH_2OSO_3H} \xrightarrow{NaOH} \underset{\text{Sodium lauryl sulfate}}{n\text{-}C_{11}H_{23}CH_2OSO_3^-\ Na^+}$$

For these, the non-polar end is the long chain, and the polar end is the $-OSO_3^-Na^+$.

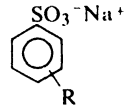
Treatment of alcohols with ethylene oxide (Sec. 17.13) yields a *non-ionic* detergent:

$$\underset{\text{Lauryl alcohol}}{CH_3(CH_2)_{10}CH_2OH} + 8\underset{\text{Ethylene oxide}}{CH_2\text{—}CH_2\ (\text{O})} \xrightarrow{base} \underset{\text{An ethoxylate}}{CH_3(CH_2)_{10}CH_2(OCH_2CH_2)_8OH}$$

Hydrogen-bonding to the numerous oxygen atoms makes the polyether end of the

molecule water-soluble. Alternatively, the ethoxylates can be converted into sulfates and used in the form of the sodium salts.

Perhaps the most widely used detergents are sodium salts of alkylbenzenesulfonic acids. A long-chain alkyl group is attached to a benzene ring by the action



of a Friedel-Crafts catalyst and an alkyl halide, an alkene, or an alcohol. Sulfonation and neutralization yields the detergent.

Formerly, polypropylene was commonly used in the synthesis of these alkylbenzenesulfonates; but the highly-branched side chain it yields blocks the rapid biological degradation of the detergent residues in sewage discharge and septic tanks. Since about 1965 in this country, such "hard" detergents have been replaced by "soft" (biodegradable) detergents: alkyl sulfates, ethoxylates and their sulfates; and alkylbenzenesulfonates in which the phenyl group is randomly attached to the various secondary positions of a long straight chain (C_{12}–C_{18} range). (See Problem 17, p. 403.) The side chains of these "linear" alkylbenzenesulfonates are derived from straight-chain 1-alkenes (Sec. 32.6), or chlorinated straight-chain alkanes separated (by use of molecular sieves) from kerosene.

These detergents act in essentially the same way as soap does. They are used because they have certain advantages. For example, the sulfates and sulfonates retain their efficiency in hard water, since the corresponding calcium and magnesium salts are soluble. Being salts of strong acids, they yield neutral solutions, in contrast to the soaps, which, being salts of weak acids, yield slightly alkaline solutions (Sec. 18.10).

33.6 Unsaturated fats. Hardening of oils. Drying oils

We have seen that fats contain, in varying proportions, glycerides of unsaturated carboxylic acids. We have also seen that, other things being equal, unsaturation in a fat tends to lower its melting point and thus tends to make it a liquid at room temperature. In the United States the long-established use of lard and butter for cooking purposes has led to a prejudice against the use of the cheaper, equally nutritious oils. Hydrogenation of some of the double bonds in such cheap fats as cottonseed oil, corn oil, and soybean oil converts these liquids into solids having a consistency comparable to that of lard or butter. This *hardening* of oils is the basis of an important industry that produces cooking fats (for example, Crisco, Spry) and oleomargarine. Hydrogenation of the carbon–carbon double bonds takes place under such mild conditions (Ni catalyst, 175–190°, 20–40 lb/in.^{2}) that hydrogenolysis of the ester linkage does not occur.

Hydrogenation not only changes the physical properties of a fat, but also—and this is even more important—changes the chemical properties: a hydrogenated fat becomes *rancid* much less readily than does a non-hydrogenated fat. Rancidity is due to the presence of volatile, bad-smelling acids and aldehydes. These compounds result (in part, at least) from attack by oxygen at reactive allylic positions in

the fat molecules; hydrogenation slows down the development of rancidity presumably by decreasing the number of double bonds and hence the number of allylic positions.

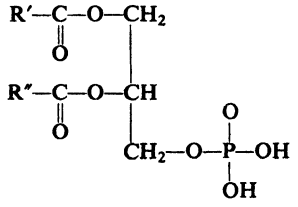
(In the presence of hydrogenation catalysts, unsaturated compounds undergo not only hydrogenation but also isomerization—shift of double bonds, or stereochemical transformations—which also affects physical and chemical properties.)

Linseed oil and tung oil have special importance because of their high content of glycerides derived from acids that contain two or three double bonds. They are known as **drying oils** and are important constituents of paints and varnishes. The "drying" of paint does not involve merely evaporation of a solvent (turpentine, etc.), but rather a chemical reaction in which a tough organic film is formed. Aside from the color due to the pigments present, protection of a surface by this film is the chief purpose of paint. The film is formed by a polymerization of the unsaturated oils that is brought about by oxygen. The polymerization process and the structure of the polymer are extremely complicated and are not well understood. The process seems to involve, in part, free-radical attack at reactive allylic hydrogens, free-radical chain-reaction polymerization similar to that previously described (Secs. 6.19 and 32.3), and cross-linking by oxygen analogous to that by sulfur in vulcanized rubber (Sec. 8.25).

Problem 33.3 In paints, tung oil "dries" faster than linseed oil. Suggest a reason why. (See Table 33.1.)

33.7 Phosphoglycerides. Phosphate esters

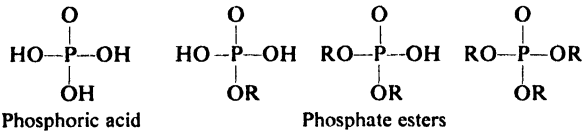
So far, we have talked only about glycerides in which all three ester linkages are to acyl groups, that is, triacylglycerols. There also occur lipids of another kind, phosphoglycerides, which contain only two acyl groups and, in place of the third, a *phosphate* group. The parent structure is *diacylglycerol phosphate*, or *phosphatidic acid.*



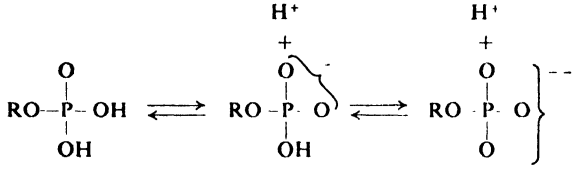
Phosphatidic acid
(A phosphoglyceride)

Phosphoglycerides are, then, not only carboxylate esters but phosphate esters as well. Just what are phosphate esters like? It will be well for us to learn something about them since we shall be encountering them again and again: phospholipids make up the membranes of cells (Sec. 33.8); adenosine triphosphate lies at the heart of the energy system of organisms, and it does its job by converting hosts of other compounds into phosphate esters (Sec. 37.3); nucleic acids, which control heredity, are polyesters of phosphoric acid.

To begin with, phosphates come in various kinds. Phosphoric acid contains three hydroxy groups and can form esters in which one, two, or three of these have been replaced by alkoxy groups. Phosphoric acid is highly acidic, and so are the



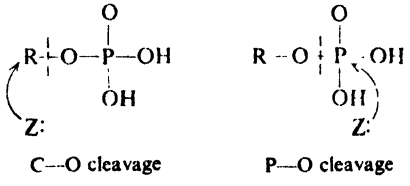
monoalkyl and dialkyl esters; in aqueous solution they tend to exist as anions, the exact extent of ionization depending, of course, upon the acidity of the medium. For example:



Like other esters, phosphates undergo hydrolysis to the parent acid and alcohol. Here, the acidity of —OH attached to phosphorus has several effects. In the first place, since acidic phosphate esters can undergo ionization, there may be many species present in the hydrolysis solution. A monoalkyl ester, for example, could exist as dianion, monoanion, neutral ester, and protonated ester; any or all of these could conceivably be undergoing hydrolysis. Actually, the situation is not quite that complicated. From the dissociation constants of these acidic esters, one can calculate the fraction of ester in each form in a given solution. The dependence of rate on acidity of the solution often shows which species is the principal reactant.

In carboxylates, we remember, attack generally occurs at acyl carbon, and in sulfonates, at alkyl carbon, with a resulting difference in point of cleavage. In

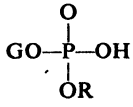
hydrolytic behavior, phosphates are intermediate between carboxylates and sulfonates. Cleavage can occur at either position, depending on the nature of the alcohol group.



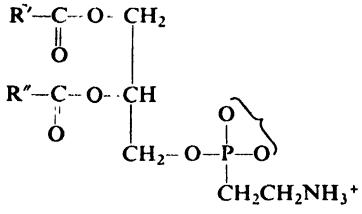
Here again the acidity of phosphoric acids comes in. Cleavage of the alkyl–oxygen bond in carboxylates is difficult because the carboxylate anion is strongly basic and a poor leaving group; in sulfonates such cleavage is favored because the weakly basic sulfonate anion is a very good leaving group. Phosphoric acid is intermediate in acidity between carboxylic and sulfonic acid; as a result, the phosphate anion is a better leaving group than carboxylate but a poorer one than sulfonate. In these esters, phosphorus is bonded to four groups; but it can accept more—witness stable pentacovalent compounds like PCl_5—and nucleophilic attack at phosphorus competes with attack at alkyl carbon.

In acidic solution, phosphate esters are readily cleaved to phosphoric acid. In alkaline solution, however, only trialkyl phosphates, $(RO)_3PO$, are hydrolyzed, and only one alkoxy group is removed. Monoalkyl and dialkyl esters, $ROPO(OH)_2$ and $(RO)_2PO(OH)$, are inert to alkali, even on long treatment. This may seem unusual behavior, but it has a perfectly rational explanation. The monoalkyl and dialkyl esters contain acidic —OH groups on phosphorus, and in alkaline solution exist as anions; repulsion between like charges prevents attack on these anions by hydroxide ion.

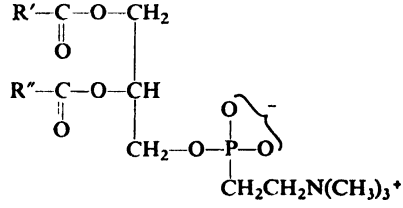
In most phospholipids, phosphate is of the kind



in which G is the glyceryl group—with its two carboxylates—and R is derived from some other alcohol, ROH, most often *ethanolamine*, $HOCH_2CH_2NH_2$, or *choline*, $HOCH_2CH_2N(CH_3)_3^+$. Since the remaining —OH on phosphorus is



Phosphatidyl ethanolamine
(Ethanolamine phosphoglyceride)



Phosphatidyl choline
(Choline phosphoglyceride)

highly acidic, the ester exists mostly in the ionic form. Furthermore, since the alcohol ROH usually contains an amino group, the phosphate unit carries both positive and negative charges, and the phospholipid is—at this end—a *dipolar ion*. On hydrolysis, these phosphates generally undergo cleavage between phosphorus and oxygen, P–O–R.

Problem 33.4 Consider hydrolysis of $(RO)_2PO(OH)$ by aqueous hydroxide, and grant that for electrostatic reasons attack by OH^- cannot occur. Even so, why does not attack by the nucleophile water lead to hydrolysis? After all, water *is* the successful nucleophile in acidic hydrolysis. (*Hint:* See Sec. 20.18.)

33.8 Phospholipids and cell membranes

The fats are found, we said, in storage fat cells of plants and animals. Their function rests on their chemical properties: through oxidation, they are consumed to help provide energy for the life processes.

The phospholipids, on the other hand, are found in the membranes of cells—all cells and are a basic structural element of living organisms. This vital function depends, in a fascinating way, on their physical properties.

Phosphoglyceride molecules are amphipathic, and in this respect differ from fats--but resemble soaps and detergents. The hydrophobic part is, again, the long fatty acid chains. The hydrophilic part is the dipolar ionic end: the substituted phosphate group with its positive and negative charges. In aqueous solution, as we would expect, phosphoglycerides form micelles. In certain situations, however—at an aperture between two aqueous solutions, for example--they tend to form bilayers: two rows of molecules are lined up, back to back, with their polar ends projecting into water on the two surfaces of the bilayer (Fig. 33.3). Although the

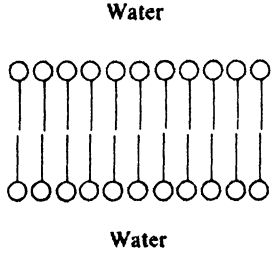


Figure 33.3. A phospholipid bilayer. Hydrophobic fatty chains held together by van der Waals forces. Hydrophilic ends dissolve in water.

polar groups are needed to hold molecules in position, the bulk of the bilayer is made up of the fatty acid chains. Non-polar molecules can therefore dissolve in this mostly hydrocarbon wall and pass through it, but it is an effective barrier to polar molecules and ions.

It is in the form of bilayers that phosphoglycerides are believed to exist in cell membranes. They constitute walls that not only enclose the cell but also very selectively control the passage, in and out, of the various substances—nutrients, waste products, hormones, etc.—even from a solution of low concentration to a solution of high concentration. Now, many of these substances that enter and leave the cells are highly polar molecules like carbohydrates and amino acids, or ions like sodium and potassium. How can these molecules pass through cell membranes when they cannot pass through simple bilayers? And how can permeability be so highly selective?

The answer to both these questions seems to involve the proteins that are also found in cell membrane: embedded in the bilayer, and even extending clear through

it. Proteins, as we shall see in Chap. 36, are very long chain amides, polymers of twenty-odd different amino acids. Protein chains can be looped and coiled in a variety of ways; the conformation that is favored for a particular protein molecule depends on the exact sequence of amino acids along its chain.

It has been suggested that transport through membranes happens in the following way. A protein molecule, coiled up to turn its hydrophobic parts outward, is dissolved in the bilayer, forming a part of the cell wall. A molecule approaches: a potassium ion, say. If the particular protein is the one designed to handle potassium ion, it receives the ion into its polar interior. Hidden in this hydrophobic wrapping, the ion is smuggled through the bilayer and released on the other side.

Now, if the transport protein is to do its job, it must be free to move within the membrane. The molecules of the bilayer, while necessarily aligned, must not be locked into a rigid crystalline lattice—as they would be if all the fatty acid chains were saturated. Actually, some of the chains in the membrane phospholipids are unsaturated and these, with their *cis* stereochemistry and the accompanying bend (Fig. 33.1), disrupt the alignment enough to make the membrane semiliquid at physiological temperatures.

Here, we have had a glimpse of just one complex biological process. Yet we can begin to see how the understanding of biology rests on basic chemical concepts: van der Waals forces and ion–dipole bonds; polarity and solubility; melting point and molecular shape; configuration and conformation; and, ultimately, the sequence of atoms in molecular chains.

Problem 33.5 The degree of unsaturation of the membrane lipids in the legs of reindeer is higher in cells near the hooves than in cells near the body. What survival value does this unsaturation gradient have?

PROBLEMS

1. From saponification of cerebrosides, lipids found in the membranes of brain and nerve cells, there is obtained *nervonic acid.* This acid rapidly decolorizes dilute $KMnO_4$ and Br_2/CCl_4 solutions. Hydrogenation in the presence of nickel yields tetracosanoic acid, $n\text{-}C_{23}H_{47}COOH$. Vigorous oxidation of nervonic acid yields one acid of neutralization equivalent 156 ± 3 and another acid of neutralization equivalent 137 ± 2. What structure or structures are possible for nervonic acid?

2. When peanut oil is heated very briefly with a little sodium methoxide, its properties are changed dramatically—it becomes so viscous it can hardly be poured—yet saponification yields the same mixture of fatty acids as did the untreated oil. What has probably happened?

3. On oxidation with O_2, methyl oleate (methyl 9-*cis*-octadecenoate) was found to yield a mixture of hydroperoxides of formula $C_{18}H_{34}O_4$. In these, the —OOH group was found attached not only to C–8 and C–11 but also to C–9 and C–10. What is the probable structure of these last two hydroperoxides? How did they arise? Show all steps in a likely mechanism for the reaction.

4. Although alkaline hydrolysis of monoalkyl or monoaryl phosphates is ordinarily very difficult, 2,4-dinitrophenyl phosphate, $2,4\text{-}(NO_2)_2C_6H_3OPO_3H_2$, does react with aqueous base, and with cleavage at the phosphorus–oxygen bond. Suggest an explanation for this.

5. *Spermaceti* (a wax from the head of the sperm whale) resembles high-molecular weight hydrocarbons in physical properties and inertness toward Br_2/CCl_4 and $KMnO_4$; on qualitative analysis it gives positive tests only for carbon and hydrogen. However, its infrared spectrum shows the presence of an ester group, and quantitative analysis gives the empirical formula $C_{16}H_{32}O$.

A solution of the wax and KOH in ethanol is refluxed for a long time. Titration of an aliquot shows that one equivalent of base has been consumed for every 475 ± 10 grams of wax. Water and ether are added to the cooled reaction mixture, and the aqueous and ethereal layers are separated. Acidification of the aqueous layer yields a solid A, m.p. 62–3°, neutralization equivalent 260 ± 5. Evaporation of the ether layer yields a solid B, m.p. 48–9°. (a) What is a likely structure of spermaceti? (b) Reduction by $LiAlH_4$ of either spermaceti or A gives B as the only product. Does this confirm the structure you gave in (a)?

6. As the acidity of the solution is increased, the rate of hydrolysis of monoalkyl phosphates, $ROPO(OH)_2$, rises from essentially zero in alkaline solution, and *passes through a maximum* at the point (moderate acidity, pH about 4) where the concentration of monoanion, $ROPO(OH)(O^-)$, is greatest. Cleavage is at the phosphorus–oxygen bond.

(a) Can you suggest a mechanism or mechanisms that might account for the fact that this species is more reactive than either the dianion, $ROPO(O^-)_2$, or the neutral ester?

(b) At still higher acidity, the rate rises again and continues to rise. To what is the high reactivity now due?

7. On the basis of the following synthesis, give the structure of *vaccenic acid.*

n-hexyl chloride + sodium acetylide ⟶ C (C_8H_{14})
C + Na, NH_3; then $I(CH_2)_9Cl$ ⟶ D ($C_{17}H_{31}Cl$)
D + KCN ⟶ E ($C_{18}H_{31}N$)
E + OH^-, heat; then H^+ ⟶ F ($C_{18}H_{32}O_2$)
F + H_2, Pd ⟶ *vaccenic acid* ($C_{18}H_{34}O_2$)

8. From the lipids of *Corynebacterium diphtherium* there is obtained *corynomycolenic acid.* Its structure was confirmed by the following synthesis.

n-$C_{13}H_{27}CH_2Br$ + sodiomalonic ester ⟶ G ($C_{21}H_{40}O_4$)
G + exactly one mole alc. KOH ⟶ H ($C_{19}H_{36}O_4$)
H + dihydropyran (Problem 16, p. 692) ⟶ I ($C_{24}H_{44}O_5$)
cis-9-hexadecenoic acid + $SOCl_2$ ⟶ J ($C_{16}H_{29}OCl$)
I + Na, then J ⟶ K ($C_{39}H_{72}O_4$)
K + dilute acid ⟶ L ($C_{34}H_{64}O_3$)
L + $NaBH_4$ ⟶ M ($C_{34}H_{66}O_3$)
M + OH^-, heat; then H^+ ⟶ (±)-*corynomycolenic acid* ($C_{32}H_{62}O_3$)

What is the structure of corynomycolenic acid?

9. From saponification of the fatty capsule of the tubercle bacillus, there is obtained *tuberculostearic acid.* Its structure was established by the following synthesis.

2-decanol + PBr_3 ⟶ N ($C_{10}H_{21}Br$)
N + sodiomalonic ester; then OH^-, heat; then H^+; then heat ⟶ O ($C_{12}H_{24}O_2$)
O + $SOCl_2$ ⟶ P $\xrightarrow{C_2H_5OH}$ Q ($C_{14}H_{28}O_2$)
Q + $LiAlH_4$ ⟶ R ($C_{12}H_{26}O$)
R + PBr_3 ⟶ S ($C_{12}H_{25}Br$)
S + Mg; $CdCl_2$; then $C_2H_5OOC(CH_2)_5COCl$ ⟶ T ($C_{21}H_{40}O_3$)
T + Zn, HCl ⟶ U ($C_{21}H_{42}O_2$)
U + OH^-, heat; then H^+ ⟶ tuberculostearic acid ($C_{19}H_{38}O_2$)

What is the structure of tuberculostearic acid?

10. Besides tuberculostearic acid (preceding problem), the capsule of the tubercle bacillus yields *C_{27}-phthienoic acid,* which on injection into animals causes the lesions typical of tuberculosis. On the basis of the following data, assign a structure to this acid.

C_{27}-phthienoic acid ($C_{27}H_{52}O_2$) + O_3; then Zn, H_2O $\longrightarrow$ $CH_3COCOOH$
C_{27}-phthienoic acid + $KMnO_4$ $\longrightarrow$ acid V ($C_{24}H_{48}O_2$)
methyl ester of V + $2C_6H_5MgBr$; then H_2O $\longrightarrow$ W ($C_{36}H_{58}O$)
W + H^+, heat $\longrightarrow$ X ($C_{36}H_{56}$)
X + CrO_3 $\longrightarrow$ $(C_6H_5)_2CO$ + ketone Y ($C_{23}H_{46}O$)
Y + I_2, NaOH $\longrightarrow$ CHI_3
V + Br_2, P $\longrightarrow$ Z $\xrightarrow{\text{alc. KOH}}$ acid AA ($C_{24}H_{46}O_2$)
AA + $KMnO_4$ $\longrightarrow$ among other things, BB ($C_{20}H_{40}O$)
Compound BB was shown to be identical with a sample of $CH_3(CH_2)_{17}COCH_3$.

Caution: $KMnO_4$ is a vigorous reagent, and not all the cleavage occurs at the double bond. Compare the number of carbons in AA and BB.

11. On the basis of the following nmr spectra, assign likely structures to the isomeric fatty acids, CC and DD, of formula $C_{17}H_{35}COOH$.

Isomer CC
a triplet, δ 0.8, 3H
b broad band, δ 1.35, 30H
c triplet, δ 2.3, 2H
d singlet, δ 12.0, 1H

Isomer DD
a triplet, δ 0.8, 3H
b doublet, δ 1.15, 3H
c broad band, δ 1.35, 28H
d multiplet, δ 2.2, 1H
e singlet, δ 12.05, 1H

12. *Juvenile hormones* take part in the delicate balance of hormonal activity that controls development of insects. Applied artificially, they prevent maturing, and thus offer a highly specific way to control insect population.

The structure of the juvenile hormone of the moth *Hyalophora cecropia* was confirmed by the following synthesis. (At each stage where geometric isomers were obtained, these were separated and the desired one—(*Z*) or (*E*)—was selected on the basis of its nmr spectrum.)

2-butanone + $[(CH_3O)_2P(O)CHCOOCH_3]^-Na^+$ (See Sec. 26.2)
$\longrightarrow$ EE ($C_9H_{19}O_6P$)
EE $\xrightarrow{\text{distillation}}$ H_3PO_4 + FF ((*Z*)-$C_7H_{12}O_2$)
FF + $LiAlH_4$ $\longrightarrow$ GG ($C_6H_{12}O$) $\xrightarrow{PBr_3}$ HH ($C_6H_{11}Br$)
HH + $[CH_3CH_2COCHCOOC_2H_5]^-Na^+$ $\longrightarrow$ II ($C_{13}H_{22}O_3$)
II + OH^-, heat; then H^+; then heat $\longrightarrow$ JJ ($C_{10}H_{18}O$)
JJ + $[(CH_3O)_2P(O)CHCOOCH_3]^-Na^+$ $\longrightarrow$ KK $\xrightarrow{\text{distillation}}$ LL ((*E*)-$C_{13}H_{22}O_2$)
LL + $LiAlH_4$ $\longrightarrow$ MM $\xrightarrow{PBr_3}$ NN ($C_{12}H_{21}Br$)
NN + sodioacetoacetic ester $\longrightarrow$ OO ($C_{18}H_{30}O_3$)
OO + OH^-, heat; then H^+; then heat $\longrightarrow$ PP ($C_{15}H_{26}O$)
PP + $[(CH_3O)_2P(O)CHCOOCH_3]^-Na^+$ $\longrightarrow$ QQ $\xrightarrow{\text{distillation}}$ RR ((*E*)-$C_{18}H_{30}O_2$)
RR + *m*-$ClC_6H_4CO_2OH$ $\longrightarrow$ SS (*racemic*-$C_{18}H_{30}O_3$)

SS was a mixture of positional isomers, corresponding to attack by perbenzoic acid at various double bonds in RR. Of these, one isomer (a *racemic* modification) was found to be identical, in physical and biological properties, to the natural juvenile hormone. This isomer was the one resulting from reaction at the double bond first introduced into the molecule.

What is the structure of the juvenile hormone of *Hyalophora cecropia*? Account for the fact that the synthesis yields a racemic modification.

Chapter 34 Carbohydrates, I. Monosaccharides

34.1 Introduction

In the leaf of a plant, the simple compounds carbon dioxide and water are combined to form the sugar (+)-**glucose**. This process, known as *photosynthesis*, requires catalysis by the green coloring matter *chlorophyll*, and requires energy in the form of light. Thousands of (+)-glucose molecules can then be combined to form the much larger molecules of **cellulose**, which constitutes the supporting framework of the plant. (+)-Glucose molecules can also be combined, in a somewhat different way, to form the large molecules of **starch**, which is then stored in the seeds to serve as food for a new, growing plant.

When eaten by an animal, the starch—and in the case of certain animals also the cellulose—is broken down into the original (+)-glucose units. These can be carried by the bloodstream to the liver to be recombined into **glycogen**, or animal starch; when the need arises, the glycogen can be broken down once more into (+)-glucose. (+)-Glucose is carried by the bloodstream to the tissues, where it is oxidized, ultimately to carbon dioxide and water, with the release of the energy originally supplied as sunlight. Some of the (+)-glucose is converted into fats; some reacts with nitrogen-containing compounds to form amino acids, which in turn are combined to form the proteins that make up a large part of the animal body.

(+)-Glucose, cellulose, starch, and glycogen all belong to the class of organic compounds known as **carbohydrates**. Carbohydrates are the ultimate source of most of our food: we eat starch-containing grain, or feed it to animals to be converted into meat and fat which we then eat. We clothe ourselves with cellulose in the form of cotton and linen, rayon and cellulose acetate. We build houses and furniture from cellulose in the form of wood. Thus carbohydrates quite literally provide us with the necessities of life: food, clothing, and shelter.

Basic necessities aside, our present civilization depends to a surprising degree

upon cellulose, particularly as *paper*: the books and newspapers we read, the letters we write, the bills we pay and the money and checks with which we pay them; marriage licenses, drivers' licenses, birth certificates, mortgages; paper in the form of bags and boxes, sheets and rolls.

The study of carbohydrates is one of the most exciting fields of organic chemistry. It extends from the tremendously complicated problem of understanding the process of photosynthesis to the equally difficult problem of unraveling the tangled steps in the enzyme-catalyzed reconversion of (+)-glucose into carbon dioxide and water. Between these two biochemical problems there lie the more traditional problems of the organic chemist: determination of the structure and properties of the carbohydrates, and the study of their conversion into other organic compounds.

In this book we shall learn something of the fundamental chemical properties of the carbohydrates, knowledge that is basic to any further study of these compounds.

34.2 Definition and classification

Carbohydrates are polyhydroxy aldehydes, polyhydroxy ketones, or compounds that can be hydrolyzed to them. A carbohydrate that cannot be hydrolyzed to simpler compounds is called a **monosaccharide**. A carbohydrate that can be hydrolyzed to two monosaccharide molecules is called a **disaccharide**. A carbohydrate that can be hydrolyzed to many monosaccharide molecules is called a **polysaccharide**.

A monosaccharide may be further classified. If it contains an aldehyde group, it is known as an **aldose**; if it contains a keto group, it is known as a **ketose**. Depending upon the number of carbon atoms it contains, a monosaccharide is known as a **triose**, **tetrose**, **pentose**, **hexose**, and so on. An **aldohexose**, for example, is a six-carbon monosaccharide containing an aldehyde group; a **ketopentose** is a five-carbon monosaccharide containing a keto group. Most naturally occurring monosaccharides are pentoses or hexoses.

Carbohydrates that reduce Fehling's (or Benedict's) or Tollens' reagent (p. 1075) are known as **reducing sugars**. All monosaccharides, whether aldose or ketose, are reducing sugars. Most disaccharides are reducing sugars; sucrose (common table sugar) is a notable exception, for it is a non-reducing sugar.

34.3 (+)-Glucose: an aldohexose

Because it is the unit of which starch, cellulose, and glycogen are made up, and because of its special role in biological processes, **(+)-glucose** is by far the most abundant monosaccharide—there are probably more (+)-glucose units in nature than any other organic group—and by far the most important monosaccharide.

Most of what we need to know about monosaccharides we can learn from the study of just this one compound, and indeed from the study of just one aspect: its structure, and how that structure was arrived at. In learning about the structure of (+)-glucose, we shall at the same time learn about its properties, since it is from these properties that the structure has been deduced. (+)-Glucose is typical mono-

saccharide, so that in learning about its structure and properties, we shall be learning about the structure and properties of the other members of this family.

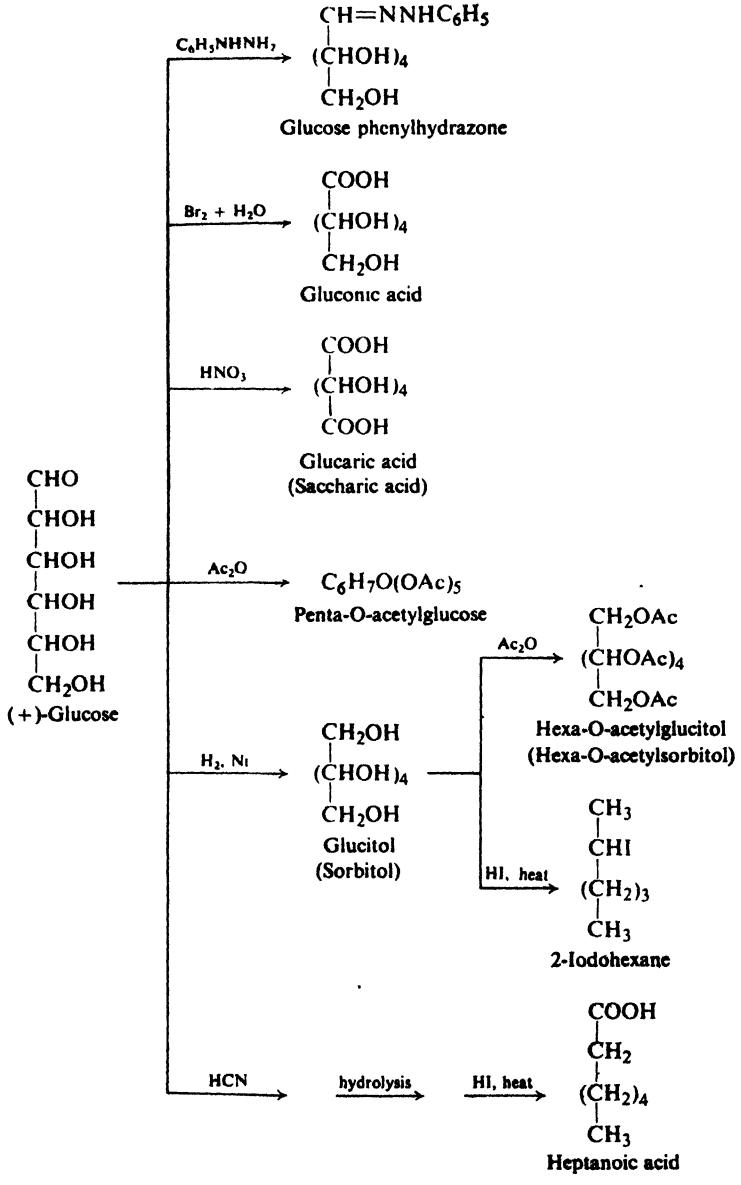


Figure 34.1. (+)-Glucose as an aldohexose.

(+)-Glucose has the molecular formula $C_6H_{12}O_6$, as shown by elemental analysis and molecular weight determination. In Fig. 34.1 is summarized other evidence about its structure: evidence consistent with the idea that (+)-glucose is

a six-carbon, straight-chain, pentahydroxy aldehyde, that is, that (+)-glucose is an aldohexose. But this is only the beginning. There are, as we shall see, 16 possible aldohexoses, all stereoisomers of each other, and we want to know which one (+)-glucose is. Beyond this, there is the fact that (+)-glucose exists in *alpha* and *beta* forms, indicating still further stereochemical possibilities that are not accommodated by the simple picture of a pentahydroxy aldehyde. Finally, we must pinpoint the predominant conformation in which the compound exists. All this is the structure of (+)-glucose and, when we have arrived at it, we shall see the features that make it the very special molecule that it is.

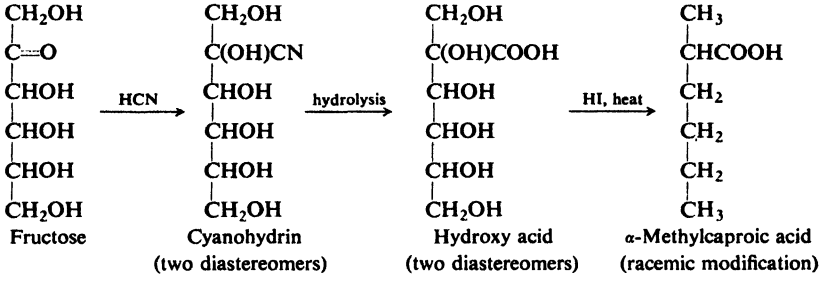
Problem 34.1 Assume that you start knowing only the molecular formula of (+)-glucose. You carry out each of the reactions of Fig. 34.1, and study each of the products obtained: characterize the product as to family; determine its molecular weight and, if any, its neutralization equivalent. You identify 2-iodohexane and heptanoic acid by comparison with authentic samples.

(a) For each product, tell what you would actually observe. (b) Take each piece of evidence in turn, and tell what it shows about the structure of (+)-glucose.

34.4 (−)-Fructose: a 2-ketohexose

The most important ketose is **(−)-fructose**, which occurs widely in fruits and, combined with glucose, in the disaccharide *sucrose* (common table sugar).

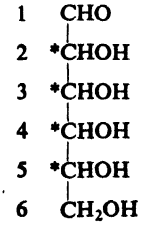
The following sequence shows that (−)-fructose is a ketone rather than an aldehyde, and gives the position of the keto group in the chain:



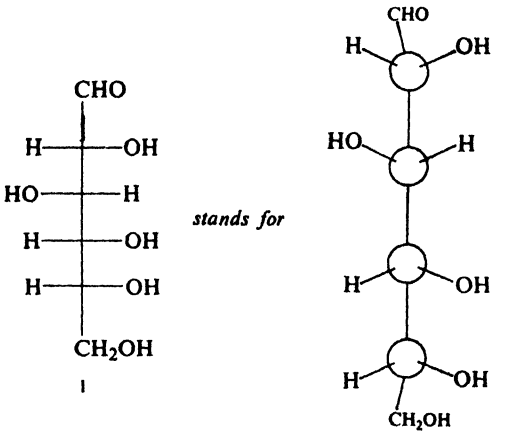
Fructose is thus a 2-ketohexose.

34.5 Stereoisomers of (+)-glucose. Nomenclature of aldose derivatives

If we examine the structural formula we have drawn for glucose, we see that it contains four chiral centers (marked by asterisks):



Each of the possible stereoisomers is commonly represented by a "cross" formula, as, for example, in I. As always in formulas of this kind, it is understood that



horizontal lines represent bonds coming *toward us* out of the plane of the paper, and *vertical* lines represent bonds going *away from us* behind the plane of the paper.

Only molecular models can show us what is really meant by formulas like I. A correct model of one of these stereoisomers is difficult to build unless we follow certain rules first clearly stated by the great carbohydrate chemist Emil Fischer:

(1) Construct a chain of carbon atoms with a —CHO group at one end and a —CH_2OH group at the other. (2) Hold the —CHO group in one hand and let the rest of the chain hang down. (3) Take the —CH_2OH group at the bottom end in the other hand and bring it up *behind* the chain until it touches the —CHO group. (4) Now one hand can hold both groups firmly and the rest of the chain will form a rather rigid ring projecting *toward you.* (This is the object of the whole operation up to this point: to impart rigidity to an otherwise flexible chain.) By this procedure you have —CHO above —CH_2OH as in formula I, and both these groups directed *away from you.* (5) Finally, still holding the ring as described above, look in turn at each carbon atom, and attach the —OH or —H to the right or to the left just as it appears in the "cross" formula. In each case, these groups will be directed *toward you.*

The dissimilarity of the two ends of an aldohexose molecule prevents the existence of *meso* compounds (Sec. 4.18), and hence we expect that there should be 2^4 or 16 stereoisomers—eight pairs of enantiomers. All 16 of these possible stereoisomers are now known, through either synthesis in the laboratory or isolation from natural sources; only three—(+)-glucose, (+)-mannose, (+)-galactose—are found in abundance.

Problem 34.2 Draw a "cross" formula of one enantiomer of each of these eight pairs, placing —CHO at the top, —CH_2OH at the bottom, and —OH on the right on the lowest chiral center (C-5).

Of these 16 isomers, only one is the (+)-glucose that we have described as the most abundant monosaccharide. A second isomer is (−)-glucose, the enantiomer of the naturally occurring compound. The other 14 isomers are all diastereomers

of (+)-glucose, and are given names of their own, for example, *mannose*, *galactose*, *gulose*, etc. As we might expect, these other aldohexoses undergo the same set of reactions that we have described for glucose. Although as diastereomers they undergo these reactions at different rates and yield different individual compounds, the chemistry is essentially the same.

The products obtained from these other aldohexoses are generally given names that correspond to the names of the products obtained from glucose. This principle is illustrated in Table 34.1 for the aldohexose (+)-mannose, which occurs naturally in many plants (the name is derived from the Biblical word *manna*).

Table 34.1 NAMES OF ALDOSE DERIVATIVES

Type of Compound	Type Name	Examples of Specific Names	
Monosaccharide $HOCH_2(CHOH)_nCHO$	Aldose	*Glucose*	*Mannose*
Monocarboxylic acid $HOCH_2(CHOH)_nCOOH$	Aldonic acid	*Gluconic acid*	*Mannonic acid*
Dicarboxylic acid $HOOC(CHOH)_nCOOH$	Aldaric acid	*Glucaric acid* (*Saccharic acid*)	*Mannaric acid* (*Mannosaccharic acid*)
Polyhydroxy alcohol $HOCH_2(CHOH)_nCH_2OH$	Alditol	*Glucitol* (*Sorbitol*)	*Mannitol*
Aldehydo acid $HOOC(CHOH)_nCHO$	Uronic acid	*Glucuronic acid*	*Mannuronic acid*

The structural formula we have drawn to represent (+)-glucose so far could actually represent any of the 16 aldohexoses. Only when we have specified the configuration about each of the chiral centers will we have the structural formula that applies only to (+)-glucose itself. Before we can discuss the brilliant way in which the configuration of (+)-glucose was worked out, we must first learn a little more about the chemistry of monosaccharides.

Problem 34.3 (a) How many chiral centers are there in (−)-fructose? (b) How many stereoisomeric 2-ketohexoses should there be? (c) Draw a "cross" formula of one enantiomer of each pair, placing C=O near the top, and —OH on the right on the lowest chiral center (C-5).

34.6 Oxidation. Effect of alkali

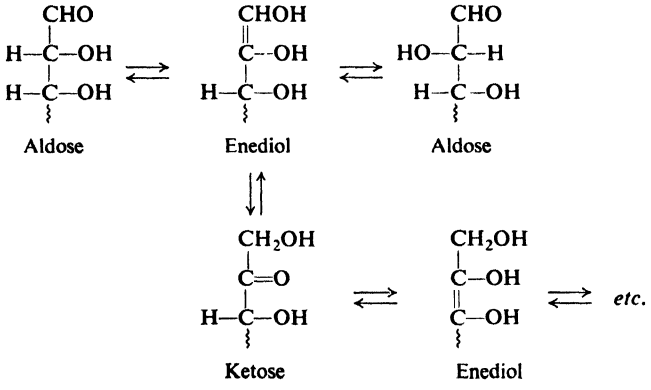
Aldoses can be oxidized in four important ways: (a) by Fehling's or Tollens' reagent; (b) by bromine water; (c) by nitric acid; and (d) by periodic acid, HIO_4.

Aldoses reduce **Tollens' reagent**, as we would expect aldehydes to do. They also reduce **Fehling's solution**, an alkaline solution of cupric ion complexed with tartrate ion (or **Benedict's solution**, in which complexing is with citrate ion); the deep-blue color of the solution is discharged, and red cuprous oxide precipitates. These reactions are less useful, however, than we might at first have expected.

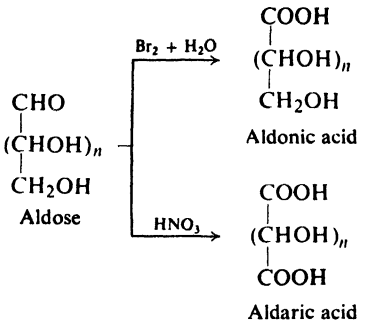
In the first place, they cannot be used to differentiate aldoses from ketoses.

Ketoses, too, reduce Fehling's and Tollens' reagents; this behavior is characteristic of α-hydroxy ketones.

In the second place, oxidation by Fehling's or Tollens' reagent cannot be used for the preparation of aldonic acids (monocarboxylic acids) from aldoses. Both Fehling's and Tollens' reagents are alkaline reagents, and the treatment of sugars with alkali can cause extensive isomerization and even decomposition of the chain. Alkali exerts this effect, in part at least, by establishing an equilibrium between the monosaccharide and an enediol structure.



Bromine water oxidizes aldoses, but not ketoses; as an acidic reagent it does not cause isomerization of the molecule. It can therefore be used to differentiate an aldose from a ketose, and is the reagent chosen to synthesize the *aldonic acid* (monocarboxylic acid) from an aldose.



Treatment of an aldose with the more vigorous oxidizing agent **nitric acid** brings about oxidation not only of the —CHO group but also of the —CH_2OH group, and leads to the formation of the *aldaric acid* (dicarboxylic acid).

Like other compounds that contain two or more —OH or =O groups on *adjacent* carbon atoms, carbohydrates undergo oxidative cleavage by **periodic acid**, HIO_4 (Sec. 16.12). This reaction, introduced in 1928 by L. Malaprade (at the University of Nancy, France), is one of the most useful tools in modern research on carbohydrate structure.

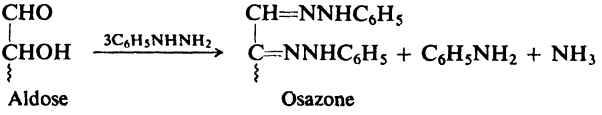
Problem 34.4 Treatment of (+)-glucose with HIO_4 gives results that confirm its aldohexose structure. What products should be formed, and how much HIO_4 should be consumed?

Problem 34.5 Identify each of the following glucose derivatives:

$$A + 4HIO_4 \longrightarrow 3HCOOH + HCHO + OHC{-}COOH$$
$$B + 5HIO_4 \longrightarrow 4HCOOH + 2HCHO$$
$$C + 3HIO_4 \longrightarrow 2HCOOH + 2OHC{-}COOH$$
$$D + 4HIO_4 \longrightarrow 4HCOOH + OHC{-}COOH$$

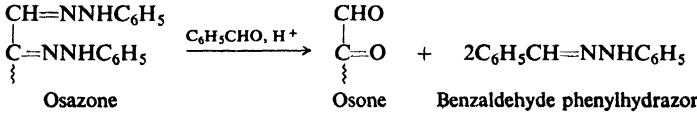
34.7 Osazone formation. Epimers

As aldehydes, aldoses react with phenylhydrazine to form phenylhydrazones. If an excess of phenylhydrazine is used, the reaction proceeds further to yield products known as **osazones**, which contain two phenylhydrazine residues per molecule; a third molecule of the reagent is turned into aniline and ammonia. (Just how the —OH group is oxidized is not quite clear.)



Osazone formation is not limited to carbohydrates, but is typical of α-hydroxy aldehydes and α-hydroxy ketones in general (e.g., *benzoin*, $C_6H_5CHOHCOC_6H_5$).

Removal of the phenylhydrazine groups yields dicarbonyl compounds known as **osones**. For example:



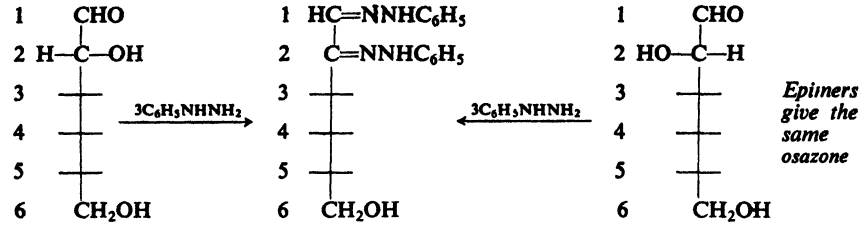
Problem 34.6 Aldehydes are more easily reduced than ketones. On this basis what product would you expect from the reduction of glucosone by zinc and acetic acid? Outline a sequence of reactions by which an aldose can be turned into a 2-ketose.

In 1858 Peter Griess (in time taken from his duties in an English brewery) discovered diazonium salts. In 1875 Emil Fischer (at the University of Munich) found that reduction of benzenediazonium chloride by sulfur dioxide yields phenylhydrazine. Nine years later, in 1884, Fischer reported that the phenylhydrazine he had discovered could be used as a powerful tool in the study of carbohydrates.

One of the difficulties of working with carbohydrates is their tendency to form sirups; these are fine for pouring on pancakes at breakfast, but hard to work with in the laboratory. Treatment with phenylhydrazine converts carbohydrates into solid osazones, which are readily isolated and purified, and can be identified by their characteristic crystalline forms.

Fischer found osazone formation to be useful not only in identifying carbohydrates, but also—and this was much more important—in determining their configurations. For example, the two diastereomeric aldohexoses (+)-glucose and

(+)-mannose yield the same osazone. Osazone formation destroys the configuration about C–2 of an aldose, but does not affect the configuration of the rest of the molecule.



It therefore follows that (+)-glucose and (+)-mannose differ only in configuration about C–2, and have the same configuration about C–3, C–4, and C–5. We can see that whenever the configuration of either of these compounds is established, the configuration of the other is immediately known through this osazone relationship. *A pair of diastereomeric aldoses that differ only in configuration about C–2 are called* **epimers**. One way in which a pair of aldoses can be identified as epimers is through the formation of the same osazone.

Problem 34.7 When the ketohexose (−)-fructose is treated with phenylhydrazine, it yields an osazone that is identical with the one prepared from either (+)-glucose or (+)-mannose. How is the configuration of (−)-fructose related to those of (+)-glucose and (+)-mannose?

34.8 Lengthening the carbon chain of aldoses. The Kiliani-Fischer synthesis

In the next few sections we shall examine some of the ways in which an aldose can be converted into a different aldose. These conversions can be used not only to synthesize new carbohydrates, but also, as we shall see, to help determine their configurations.

First, let us look at a method for converting an aldose into another aldose containing one more carbon atom, that is, at a method for lengthening the carbon chain. In 1886, Heinrich Kiliani (at the Technische Hochschule in Munich) showed that an aldose can be converted into two aldonic acids of the next higher carbon number by addition of HCN and hydrolysis of the resulting cyanohydrins. In 1890, Fischer reported that reduction of an aldonic acid (in the form of its lactone, Sec. 20.15) can be controlled to yield the corresponding aldose. In Fig. 34.2, the entire **Kiliani-Fischer synthesis** is illustrated for the conversion of an aldopentose into two aldohexoses.

Addition of cyanide to the aldopentose generates a new chiral center, about which there are two possible configurations. As a result, two diastereomeric cyanohydrins are obtained, which yield diastereomeric carboxylic acids (aldonic acids) and finally diastereomeric aldoses.

The situation is strictly analogous to that in Sec. 7.7. Using models, we can see that the particular configuration obtained here depends upon which face of the carbonyl group is attacked by cyanide ion. Since the aldehyde is already chiral, attack at the two faces is not equally likely. Both possible diastereomeric products are formed, and in unequal amounts.

Since a six-carbon aldonic acid contains —OH groups in the γ- and δ-positions, we would expect it to form a lactone under acidic conditions (Sec. 20.15). This occurs, the γ-lactone generally being the more stable product. It is the lactone that is actually reduced to an aldose in the last step of a Kiliani-Fischer synthesis.

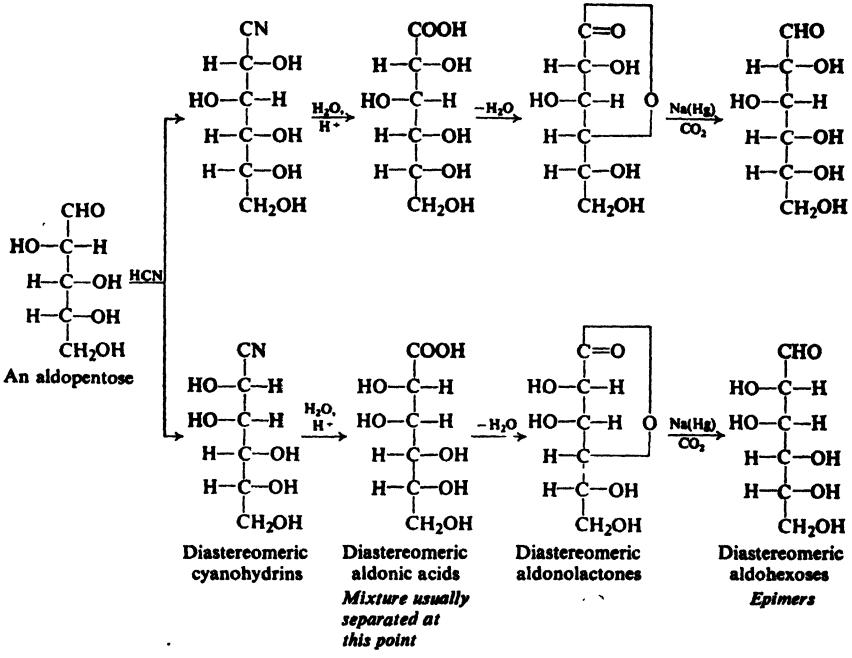


Figure 34.2. An example of the Kiliani-Fischer synthesis.

The pair of aldoses obtained from the sequence differ only in configuration about C-2, and hence are epimers. A pair of aldoses can be recognized as epimers not only by their conversion into the same osazone (Sec. 34.7), but also by their formation in the same Kiliani-Fischer synthesis.

Like other diastereomers, these epimers differ in physical properties and therefore are separable. However, since carbohydrates are difficult to purify, it is usually more convenient to separate the diastereomeric products at the acid stage, where crystalline salts are easily formed, so that a single pure lactone can be reduced to a single pure aldose.

Problem 34.8 As reducing agent, Fischer used sodium amalgam and acid. Today, lactones are reduced to aldoses by the addition of $NaBH_4$ to an aqueous solution of lactone. If, however, lactone is added to the $NaBH_4$, another product, not the aldose, is obtained. What do you think this other product is? Why is the order of mixing of reagents crucial?

Problem 34.9 (a) Using cross formulas to show configuration, outline all steps in a Kiliani-Fischer synthesis, starting with the aldotriose R-(+)-glyceraldehyde, $CH_2OHCHOHCHO$. How many aldotetroses would be expected? (b) Give configura-

tions of the aldopentoses expected from each of these aldotetroses by a Kiliani-Fischer synthesis; of the aldohexoses expected from each of these aldopentoses.

(c) Make a "family tree" showing configurations of these aldoses hypothetically descended from R-(+)-glyceraldehyde. If the —CHO is placed at the top in each case, what configurational feature is the same in all these formulas? Why?

Problem 34.10 (a) Give the configuration of the dicarboxylic acid (aldaric acid) that would be obtained from each of the tetroses in Problem 34.9 by nitric acid oxidation. (b) Assume that you have actually carried out the chemistry in part (a). In what simple way could you assign configuration to each of your tetroses?

34.9 Shortening the carbon chain of aldoses. The Ruff degradation

There are a number of ways in which an aldose can be converted into another aldose of one less carbon atom. One of these methods for shortening the carbon chain is the **Ruff degradation**. An aldose is oxidized by bromine water to the aldonic acid; oxidation of the calcium salt of this acid by hydrogen peroxide in the presence of ferric salts yields carbonate ion and an aldose of one less carbon atom (see Fig. 34.3).

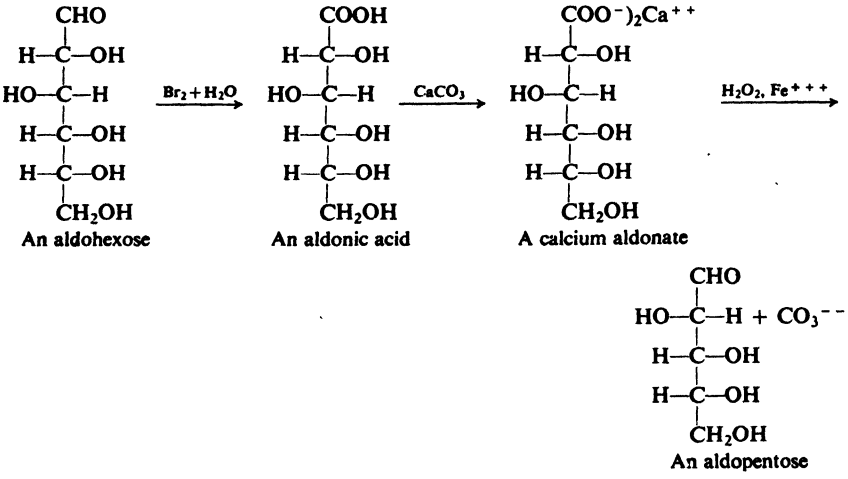


Figure 34.3. An example of the Ruff degradation.

34.10 Conversion of an aldose into its epimer

In the presence of a tertiary amine, in particular pyridine (Sec. 31.6), an equilibrium is established between an aldonic acid and its epimer. This reaction is the basis of the best method for converting an aldose into its epimer, since the only configuration affected is that at C-2. The aldose is oxidized by bromine water to the aldonic acid, which is then treated with pyridine. From the equilibrium mixture thus formed, the epimeric aldonic acid is separated, and reduced (in the form of its lactone) to the epimeric aldose. See, for example, Fig. 34.4.

34.11 Configuration of (+)-glucose. The Fischer proof

Let us turn back to the year 1888. Only a few monosaccharides were known, among them (+)-glucose, (−)-fructose, (+)-arabinose. (+)-Mannose had just

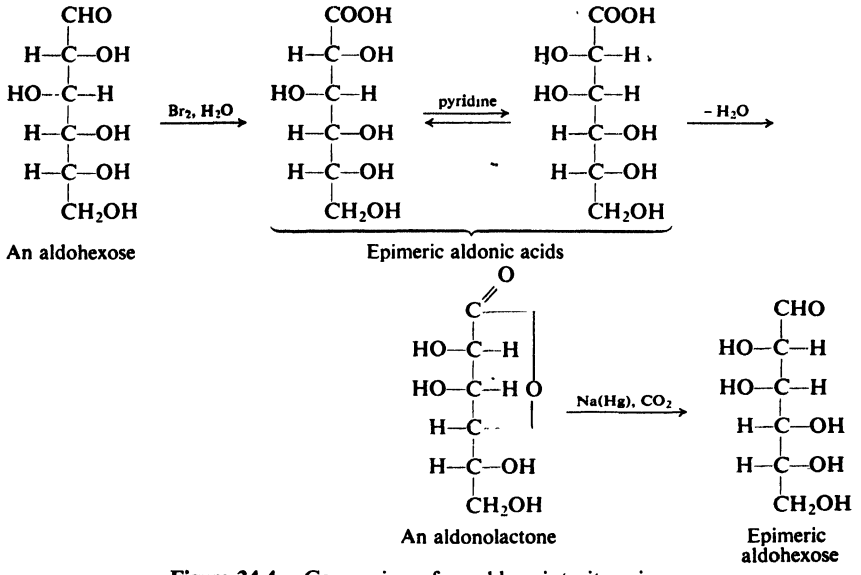


Figure 34.4. Conversion of an aldose into its epimer.

been synthesized. It was known that (+)-glucose was an aldohexose and that (+)-arabinose was an aldopentose. Emil Fischer had discovered (1884) that phenylhydrazine could convert carbohydrates into osazones. The Kiliani cyanohydrin method for lengthening the chain was just two years old.

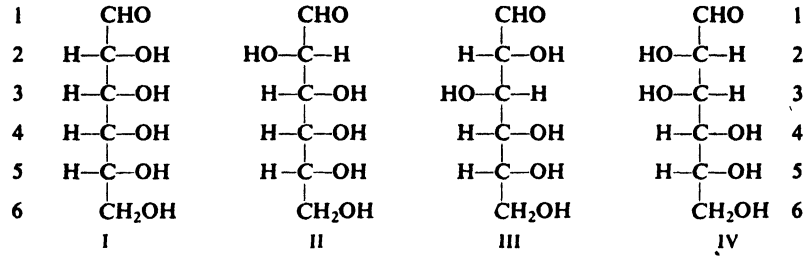
It was known that aldoses could be reduced to alditols, and could be oxidized to the monocarboxylic aldonic acids and to the dicarboxylic aldaric acids. A theory of stereoisomerism and optical activity had been proposed (1874) by van't Hoff and Le Bel. Methods for separating stereoisomers were known and optical activity could be measured. The concepts of racemic modifications, *meso* compounds, and epimers were well established.

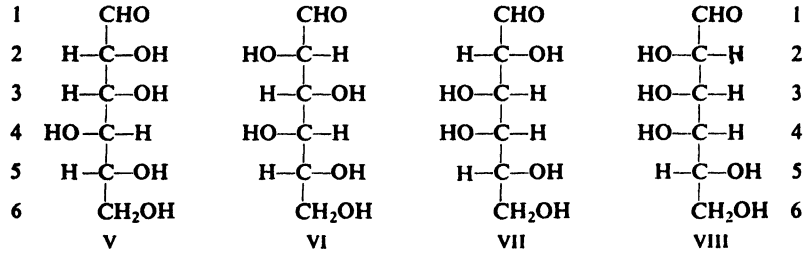
(+)-Glucose was known to be an aldohexose; but as an aldohexose it could have any one of 16 possible configurations. The question was: *which* configuration did it have? In 1888, Emil Fischer (at the University of Würzburg) set out to find the answer to that question, and in 1891 announced the completion of a most remarkable piece of chemical research, for which he received the Nobel Prize in 1902. Let us follow Fischer's steps to the configuration of (+)-glucose. Although somewhat modified, the following arguments are essentially those of Fischer.

The 16 possible configurations consist of eight pairs of enantiomers. Since methods of determining absolute configuration were not then available, Fischer realized that he could at best limit the configuration of (+)-glucose to a pair of enantiomeric configurations; he would not be able to tell which one of the pair was the correct absolute configuration.

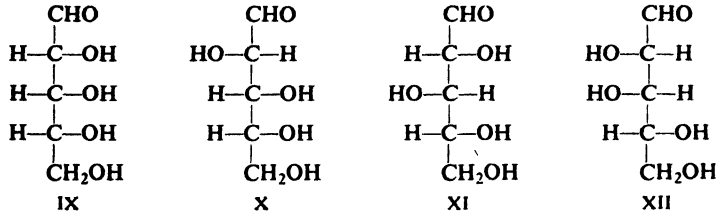
To simplify the problem, Fischer therefore rejected eight of the possible configurations, arbitrarily retaining only those (I–VIII) in which C-5 carried the —OH on the right (with the understanding that —H and —OH project toward the observer). He realized that any argument that led to the selection of one of these formulas applied with equal force to the mirror image of that formula. (As it

turned out, his arbitrary choice of an —OH on the right of C-5 in (+)-glucose was the correct one.)



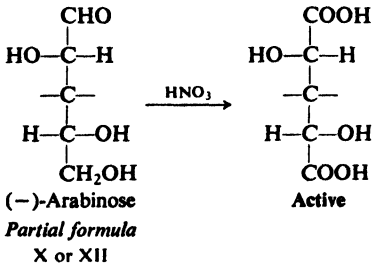


Since his proof depended in part on the relationship between (+)-glucose and the aldopentose (−)-arabinose, Fischer also had to consider the configurations of the five-carbon aldoses. Of the eight possible configurations, he retained only four, IX–XII, again those in which the bottom chiral center carried the —OH on the right.

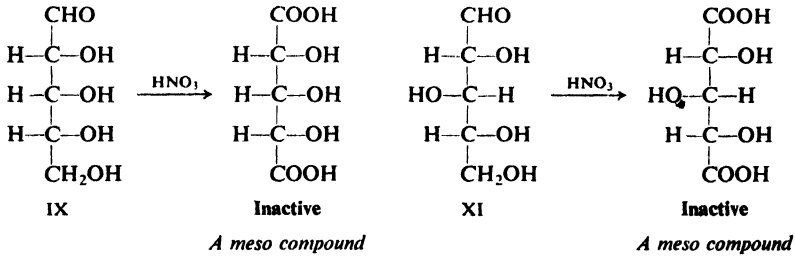


The line of argument is as follows:

(1) Upon oxidation by nitric acid, (—)-arabinose yields an optically active dicarboxylic acid. Since the —OH on the lowest chiral center is arbitrarily placed on the right, this fact means that the —OH on the uppermost chiral center is on the left (as in X or XII),



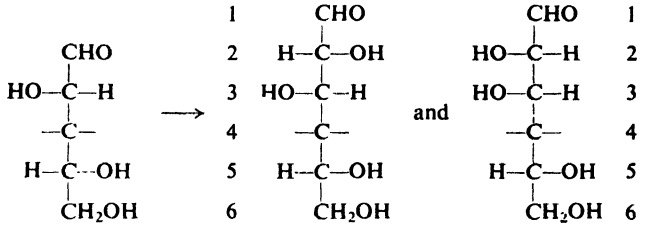
for if it were on the right (as in IX or XI), the diacid would necessarily be an inactive *meso* acid.



IX — Inactive — *A meso compound*

XI — Inactive — *A meso compound*

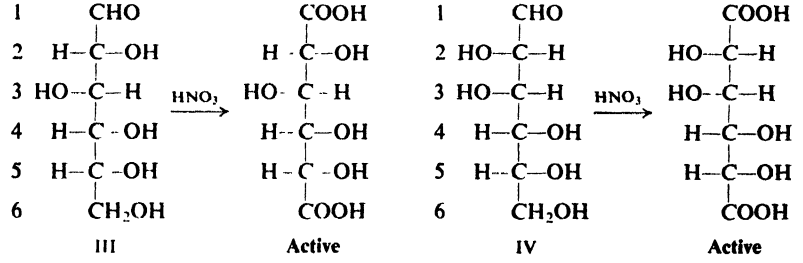
(2) (−)-Arabinose is converted by the Kiliani-Fischer synthesis into (+)-glucose and (+)-mannose. (+)-Glucose and (+)-mannose therefore are epimers, differing only in configuration about C–2, and have the same configuration about C–3, C–4, and C–5 as does (−)-arabinose. (+)-Glucose and (+)-mannose must be III and IV, or VII and VIII.



(−)-Arabinose — *Partial formula* X or XII

(+)-Glucose and (+)-Mannose: epimers — *Partial formulas* III and IV, or VII and VIII

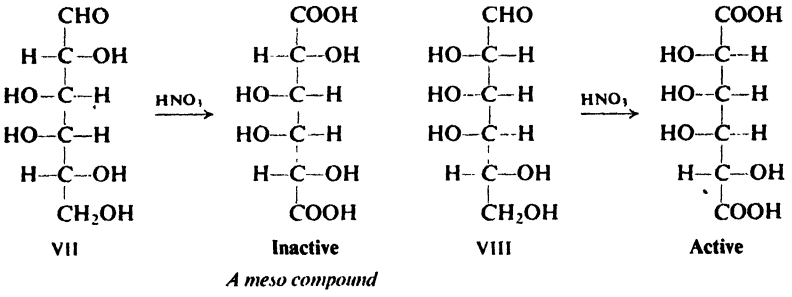
(3) Upon oxidation by nitric acid, both (+)-glucose and (+)-mannose yield dicarboxylic acids that are optically active. This means that the —OH on C–4 is on the right, as in III and IV,



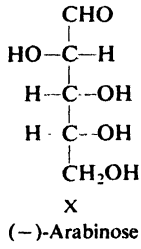
III — Active

IV — Active

for if it were on the left, as in VII and VIII, *one* of the aldaric acids would necessarily be an inactive *meso* acid.



(−)-Arabinose must also have that same —OH on the right, and hence has configuration X.

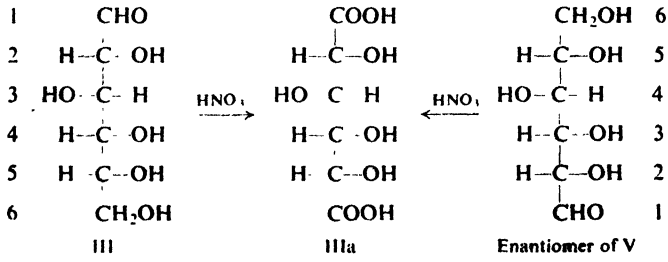


X

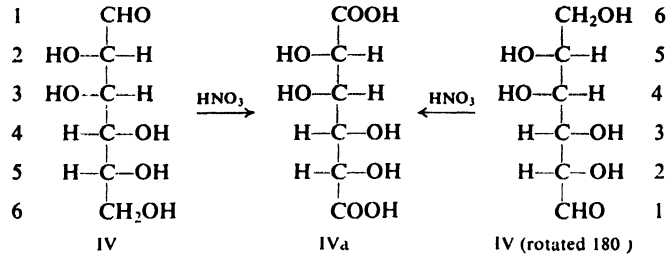
(−)-Arabinose

(+)-Glucose and (+)-mannose have configurations III and IV, but one question remains: which compound has which configuration? One more step is needed.

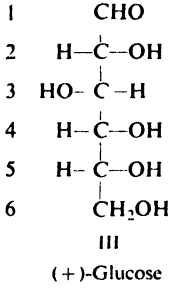
(4) Oxidation of another hexose, (+)-gulose, yields the same dicarboxylic acid, (+)-glucaric acid, as does oxidation of (+)-glucose. (The gulose was synthesized for this purpose by Fischer.) If we examine the two possible configurations for (+)-glucaric acid, IIIa and IVa, we see that only IIIa can be derived from two different hexoses: from III and the enantiomer of V.



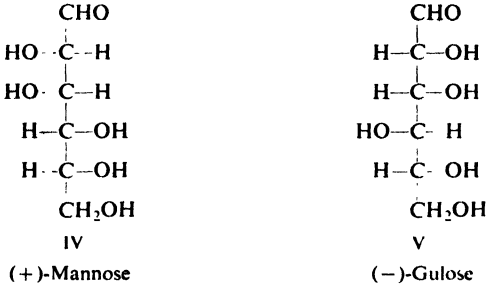
The acid IVa can be derived from just one hexose: from IV.



It follows that (+)-glucaric acid has configuration IIIa, and therefore that (+)-glucose has configuration III.



(+)-Mannose, of course, has configuration IV, and (−)-gulose (the enantiomer of the one used by Fischer) has configuration V.



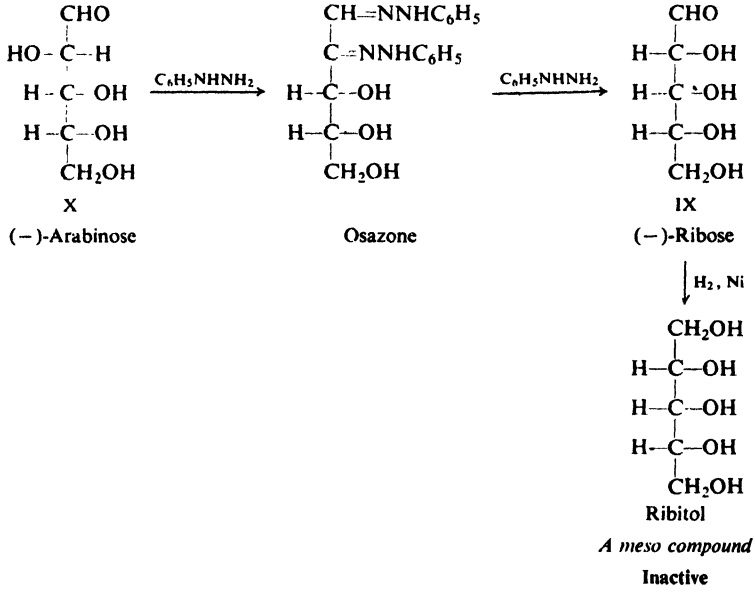
34.12 Configurations of aldoses

Today all possible aldoses (and ketoses) of six carbons or less, and many of more than six carbons, are known; most of these do not occur naturally and have been synthesized. The configurations of all these have been determined by application of the same principles that Fischer used to establish the configuration of (+)-glucose; indeed, twelve of the sixteen aldohexoses were worked out by Fischer and his students.

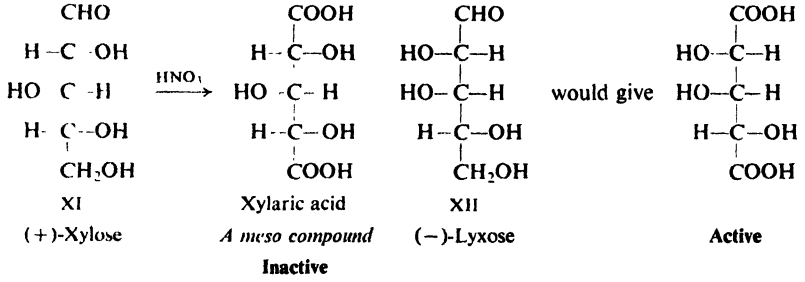
So far in our discussion, we have seen how configurations III, IV, V, and X of the previous section were assigned to (+)-glucose, (+)-mannose, (−)-gulose,

and (−)-arabinose, respectively. Let us see how configurations have been assigned to some other monosaccharides.

The aldopentose **(—)-ribose** forms the same osazone as (−)-arabinose. Since (−)-arabinose was shown to have configuration X, (−)-ribose must have configuration IX. This configuration is confirmed by the reduction of (−)-ribose to the optically inactive (*meso*) pentahydroxy compound *ribitol*.

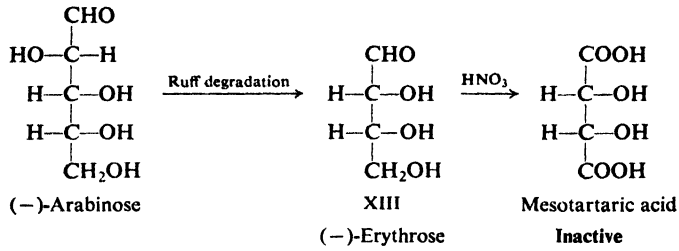


The two remaining aldopentoses, **(+)-xylose** and **(—)-lyxose**, must have the configurations XI and XII. Oxidation by nitric acid converts (+)-xylose into an

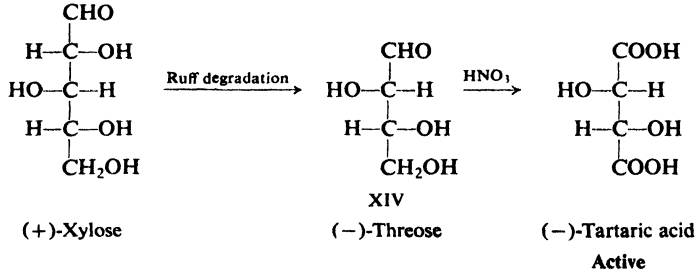


optically inactive (*meso*) aldaric acid. (+)-Xylose must therefore be XI and (−)-lyxose must be XII.

Degradation of (−)-arabinose yields the tetrose **(—)-erythrose**, which therefore has configuration XIII. In agreement with this configuration, (−)-erythrose is found to yield *meso*tartaric acid upon oxidation by nitric acid.



Degradation of (+)-xylose by the Ruff method yields the tetrose (**−**)-**threose**, which must therefore have configuration XIV. This is confirmed by oxidation of (−)-threose to optically active (−)-tartaric acid.



Problem 34.11 Assign a name to I, II, VI, VII, and VIII (p. 1082) on the basis of the following evidence and the configurations already assigned:

(a) The aldohexoses **(+)-galactose** and **(+)-talose** yield the same osazone. Degradation of (+)-galactose yields (−)-lyxose. Oxidation of (+)-galactose by nitric acid yields an inactive *meso* acid, *galactaric acid* (also called *mucic acid*).

(b) (−)-Ribose is converted by the Kiliani-Fischer synthesis into the two aldohexoses **(+)-allose** and **(+)-altrose**. Oxidation of (+)-altrose yields optically active *(+)-altraric acid.* Reduction of (+)-allose to a hexahydroxy alcohol yields optically **inactive** *allitol.*

(c) The aldohexose **(−)-idose** yields the same osazone as (−)-gulose.

Problem 34.12 Go back to the "family tree" you constructed in Problem 34.9, p. 1079, and assign names to all structures.

Problem 34.13 What is the configuration of the 2-ketohexose (−)-fructose? (See Problem 34.7, p. 1078.)

Problem 34.14 Give the configurations of (−)-glucose, (−)-mannose, and (+)-fructose.

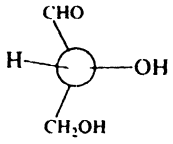
34.13 Optical families. D and L

Before we can explore further the structure of (+)-glucose and its relatives, we must examine a topic of stereochemistry we have not yet touched on: use of the prefixes D and L.

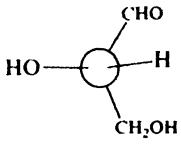
Most applications of stereochemistry, as we have already seen, are based upon the *relative* configurations of different compounds, not upon their absolute configurations. We are chiefly interested in whether the configurations of a reactant and its product are the same or different, not in what either configuration actually is. In the days before any absolute configurations had been determined,

there was the problem not only of determining the relative configurations of various optically active compounds, but also of indicating these relationships once they had been established. This was a particularly pressing problem with the carbohydrates.

The compound **glyceraldehyde**, $CH_2OHCHOHCHO$, was selected as a standard of reference, because it is the simplest carbohydrate—an aldotriose—capable of optical isomerism. Its configuration could be related to those of the carbohydrates, and because of its highly reactive functional groups, it could be converted into, and thus related to, many other kinds of organic compounds. (+)-Glyceraldehyde was arbitrarily assigned configuration I, and was designated D-glyceraldehyde; (−)-glyceraldehyde was assigned configuration II and was designated L-glyceraldehyde. Configurations were assigned to the glyceraldehydes purely for



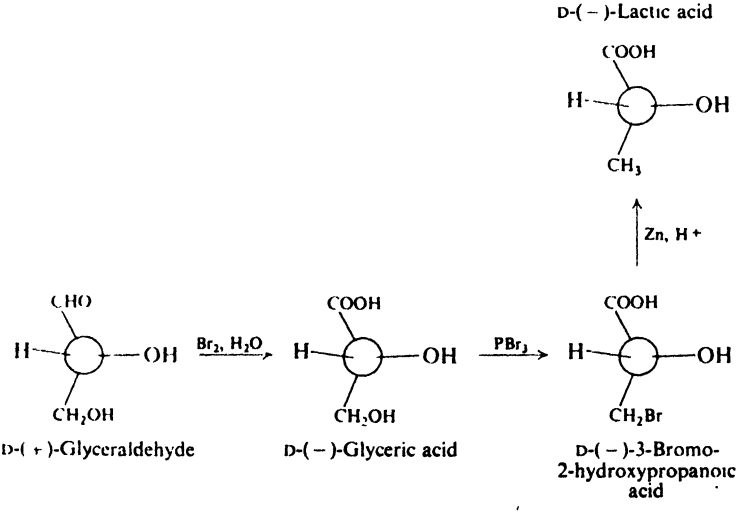
I
D-Glyceraldehyde



II
L-Glyceraldehyde

convenience; the particular assignment had a 50:50 chance of being correct, and, as it has turned out, the configuration chosen actually is the correct absolute configuration.

Other compounds could be related configurationally to one or the other of the glyceraldehydes by means of reactions that did not involve breaking bonds to a chiral center (Sec. 7.5). On the basis of the *assumed* configuration of the glyceraldehyde, these related compounds could be assigned configurations, too. As it has turned out, these configurations are the correct absolute ones; in any case, for



D-(+)-Glyceraldehyde

D-(−)-Glyceric acid

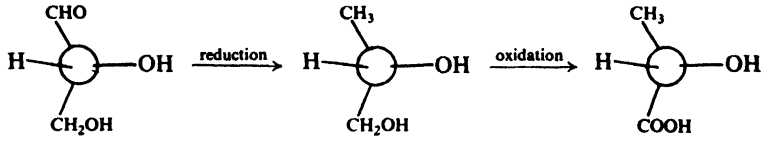
D-(−)-3-Bromo-2-hydroxypropanoic acid

Figure 34.5. Relating configurations to glyceraldehyde.

many years they served as a convenient way of indicating structural relationships. See, for example, Fig. 34.5.

To indicate the relationship thus established, compounds related to D-glyceraldehyde are given the designation D, and compounds related to L-glyceraldehyde are given the designation L. The symbols D and L (pronounced "dee" and "ell") thus refer to configuration, not to sign of rotation, so that we have, for example, D-(−)-glyceric acid and L-(+)-lactic acid. (One frequently encounters the prefixes *d* and *l*, pronounced "dextro" and "levo," but their meaning is not always clear. Today they usually refer to direction of rotation; in some of the older literature they refer to optical family. It was because of this confusion that D and L were introduced.)

Unfortunately, the use of the designations D and L is not unambiguous. In relating glyceraldehyde to lactic acid, for example, we might envision carrying out a sequence of steps in which the —CH_2OH rather than the —CHO group is converted into the —COOH group:



(+)-Glyceraldehyde (+)-1,2-Propanediol (+)-Lactic acid

By this series of reactions, (+)-glyceraldehyde would yield (+)-lactic acid; by the previous sequence, (+)-glyceraldehyde yields (−)-lactic acid. It would appear that, depending upon the particular sequence used, we could designate either of the lactic acids as D-lactic acid; the first sequence is the more direct, and by convention is the accepted one. We should notice that, whatever the ambiguity associated with the use of D and L, there is no ambiguity about the configurational relationship; we arrive at the proper configurations for (+)- and (−)-lactic acids whichever route we use.

The prefixes R and S enable us to specify unambiguously the absolute configuration of a compound, because their use does not depend on a relationship to any other compound. But, by the same token, the letters R and S do not immediately reveal configurational relationships between two compounds; we have to work out and compare the configurations in each case.

The designations D and L, on the other hand, tell us nothing of the configuration of the compound unless we know the route by which the configurational relationship has been established. However, in the case of the carbohydrates (and the amino acids, Chap. 36), there are certain conventions about this which make these designations extremely useful.

Problem 34.15 Which specification, R or S, would you give to the following? (a) D-(+)-glyceraldehyde; (b) D-(−)-glyceric acid; (c) D-(−)-3-bromo-2-hydroxypropionic acid; (d) D-(−)-lactic acid.

Problem 34.16 The transformation of L-(+)-lactic acid into (+)-2-butanol was accomplished by the following sequence of reactions:

L-(+)-lactic acid $\xrightarrow{C_2H_5OH,\ H_2SO_4}$ A $\xrightarrow{Na,\ C_2H_5OH}$ B $\xrightarrow{HBr}$ C

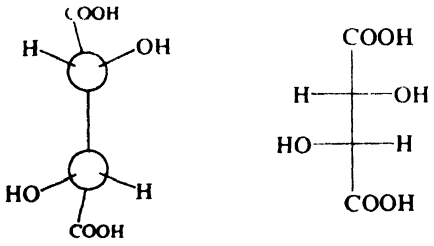
C $\xrightarrow{KCN}$ D $\xrightarrow{H_2O,\ HCl,\ heat}$ E $\xrightarrow{CH_3OH,\ HCl}$ F $\xrightarrow{Na,\ CH_3COOH}$ G

G $\xrightarrow{HI}$ H $\xrightarrow{H_2,\ Pd}$ (+)-2-butanol

What is the absolute configuration of (+)-2-butanol?

34.14 Tartaric acid

Tartaric acid, HOOCCHOHCHOHCOOH, has played a key role in the development of stereochemistry, and particularly the stereochemistry of the carbohydrates. In 1848 Louis Pasteur, using a hand lens and a pair of tweezers, laboriously separated a quantity of the sodium ammonium salt of racemic tartaric acid into two piles of mirror-image crystals and, in thus carrying out the first resolution of a racemic modification, was led to the discovery of enantiomerism. Almost exactly 100 years later, in 1949, Bijvoet, using x-ray diffraction—and also laboriously—determined the actual arrangement in space of the atoms of the sodium rubidium salt of (+)-tartaric acid, and thus made the first determination of the absolute configuration of an optically active substance.



(+)-Tartaric acid

As we shall see in the next section, tartaric acid is the stereochemical link between the carbohydrates and our standard of reference, glyceraldehyde. In 1917, the configurational relationship between glyceraldehyde and tartaric acid was worked out. When the reaction sequence outlined in Fig. 34.6 was carried out starting with D-glyceraldehyde, two products were obtained, one inactive and

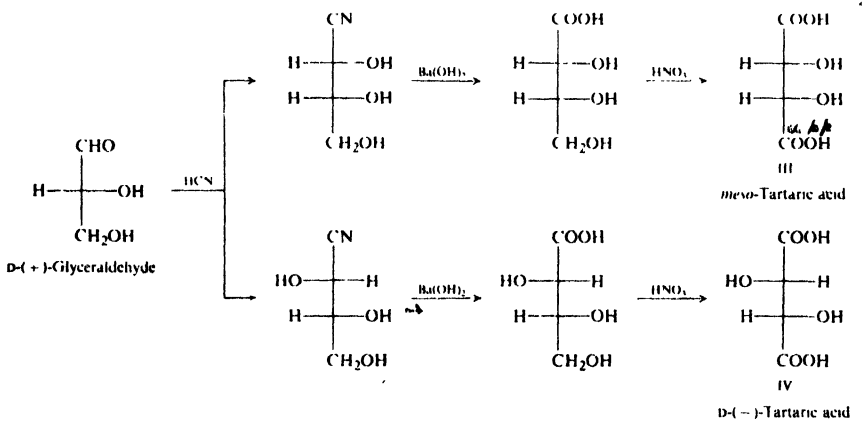
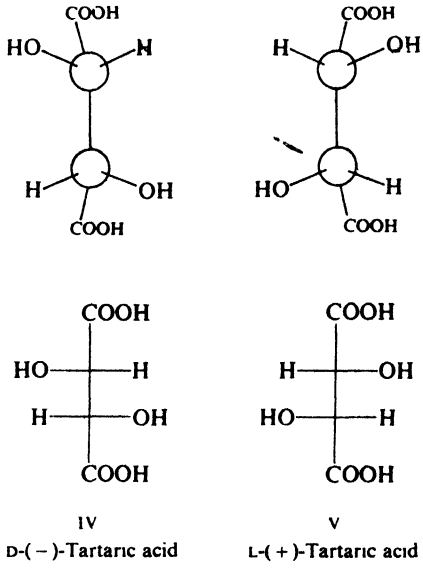


Figure 34.6. Configurational relationship between glyceraldehyde and tartaric acid.

one which rotated the plane of polarized light to the left. The inactive product was, of course, mesotartaric acid, III. The active (−)-tartaric acid thus obtained was assigned configuration IV; since it is related to D-glyceraldehyde, we designate it D-(−)-tartaric acid.

On the basis of the assumed configuration of D-(+)-glyceraldehyde, then, L-(+)-tartaric acid, the enantiomer of D-(−)-tartaric acid, would have configuration V, the mirror image of IV. When Bijvoet determined the absolute configuration



IV
D-(−)-Tartaric acid

V
L-(+)-Tartaric acid

of (+)-tartaric acid, he found that it actually has the configuration that had been previously assumed. The assumed configurations of the glyceraldehydes, and hence the assumed configurations of all compounds related to them, were indeed the correct ones.

The designation of even the tartaric acids is subject to ambiguity. In this book, we have treated the tartaric acids as one does carbohydrates: by considering —CHO of glyceraldehyde as the position from which the chain is lengthened, via the cyanohydrin reaction. Some chemists, on the other hand, view the tartaric acids as one does the amino acids (Sec. 36.5) and, considering —COOH to be derived from —CHO of glyceraldehyde, designate (−)-tartaric acid as L, and (+)-tartaric acid as D.

Regardless of which convention one follows, this fact remains: (−)- and (+)-tartaric acid—and (+)- and (−)-glyceraldehyde—have the absolute configurations shown on p. 1088 and above.

Problem 34.17 Give the specification by the R/S system of: (a) (−)-tartaric acid; (b) (+)-tartaric acid; (c) mesotartaric acid.

Problem 34.18 (a) From the sequence of Fig. 34.6 the ratio of products III:IV is about 1:3. Why would you have expected to obtain III and IV in unequal amounts?

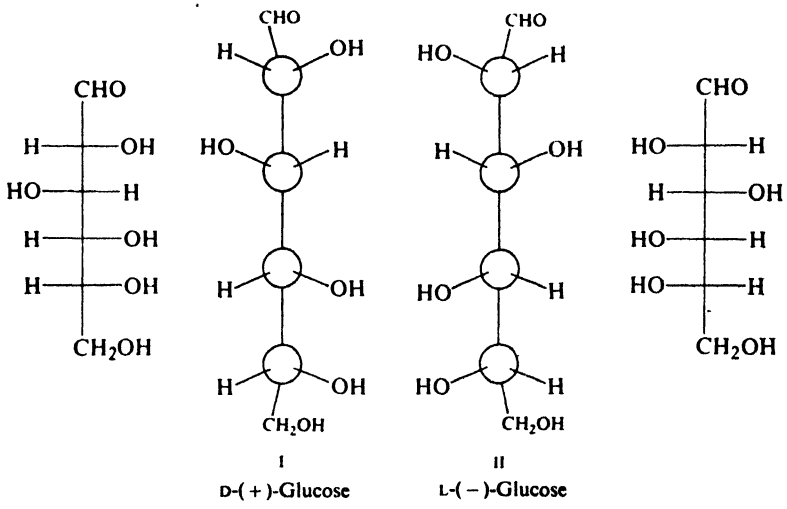
(b) Outline the same sequence starting from L-(−)-glyceraldehyde. Label each

product with its name, showing its rotation and D/L designation. In what ratio will these products be obtained?

(c) Outline the same sequence starting from racemic (±)-glyceraldehyde. How do you account for the fact that only inactive material is obtained in spite of the unequal amounts of diastereomeric products formed from each of the enantiomeric glyceraldehydes?

34.15 Families of aldoses. Absolute configuration

The evidence on which Fischer assigned a configuration to (+)-glucose leads to either of the enantiomeric structures I and II. Fischer, we have seen, arbitrarily selected I, in which the lowest chiral center carries —OH on the right.



We recognize I as the ènantiomer that would hypothetically be derived from D-(+)-glyceraldehyde by a series of Kiliani-Fischer syntheses, the chiral center of (+)-glyceraldehyde being retained as the *lowest* chiral center of the aldoses derived from it. (See Problem 34.9, p. 1079.) That (+)-glucose is related to D-(+)-glyceraldehyde has been established by a number of reaction sequences, one of which is shown in Fig. 34.7. On this basis, then, structure I becomes D-(+)-glucose, and structure II becomes L-(−)-glucose.

In 1906 the American chemist Rosanoff (then an instructor at New York University) proposed glyceraldehyde as the standard to which the configurations of carbohydrates should be related. Eleven years later experiment showed that it is the *dextrorotatory* (+)-glyceraldehyde that is related to (+)-glucose. On that basis, (+)-glyceraldehyde was then given the designation D and was assigned a configuration to conform with the one arbitrarily assigned to (+)-glucose by Fischer. Although rejected by Fischer, the Rosanoff convention became universally accepted.

Regardless of the direction in which they rotate polarized light, all monosaccharides are designated as D or L on the basis of the configuration about the lowest chiral center, the carbonyl group being at the top: D if the —OH is on the

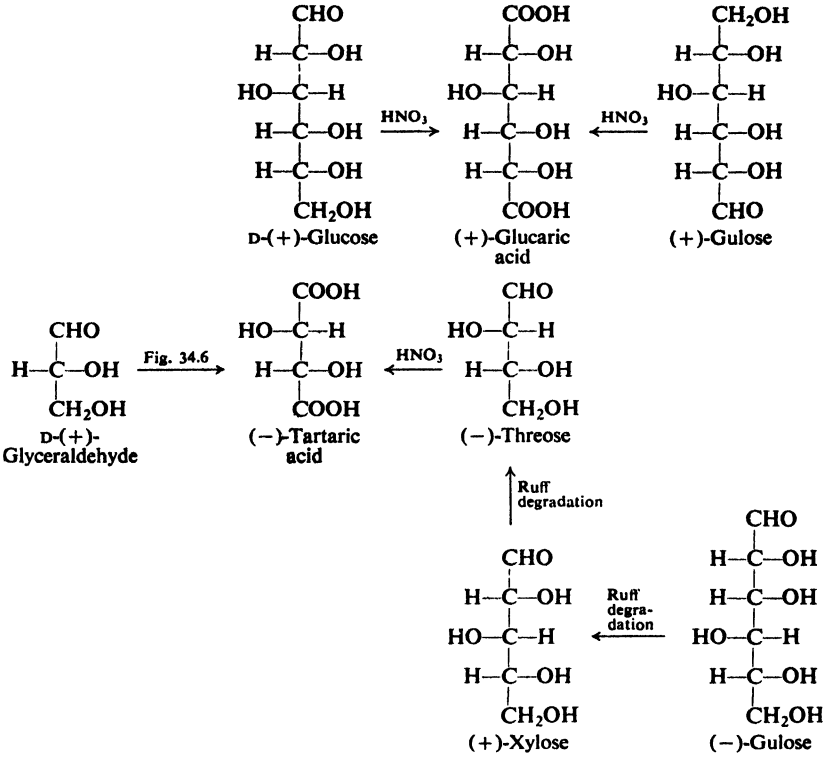


Figure 34.7. Relating (+)-glucose to D-(+)-glyceraldehyde.

right, L if the —OH is on the left. (As always, it is understood that —H and —OH project toward us from the plane of the paper.) (+)-Mannose and (−)-arabinose, for example, are both assigned to the D-family on the basis of their relationship to D-(+)-glucose, and, through it, to D-(+)-glyceraldehyde.

Until 1949, these configurations were accepted on a purely empirical basis; they were a convenient way to show configurational relationships among the various carbohydrates, and between them and other organic compounds. But so far as anyone knew, the configurations of these compounds might actually have been the mirror images of those assigned; the lowest chiral center in the D-series of monosaccharides might have carried —OH on the left. As we have seen, however, when Bijvoet determined the absolute configuration of (+)-tartaric acid by x-ray analysis in 1949, he found that it actually has the configuration that had been up to then merely assumed. The arbitrary choice that Emil Fischer made in 1891 was the correct one; the configuration he assigned to (+)-glucose—and, through it, to every carbohydrate—is the correct absolute configuration.

Problem 34.19 The (+)-gulose that played such an important part in the proof of configuration of D-(+)-glucose was synthesized by Fischer via the following sequence:

D-(+)-glucose $\xrightarrow{HNO_3}$ (+)-glucaric acid $\xrightarrow{-H_2O}$ A and B (lactones, separated)

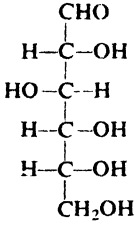
A $\xrightarrow{Na(Hg)}$ C (aldonic acid) $\xrightarrow{-H_2O}$ D (lactone) $\xrightarrow{Na(Hg),\ acid}$ D-(+)-glucose

B $\xrightarrow{Na(Hg)}$ E (aldonic acid) $\xrightarrow{-H_2O}$ F (lactone) $\xrightarrow{Na(Hg),\ acid}$ (+)-gulose

Give the structures of A through F. What is the configuration of (+)-gulose? Is it a member of the D-family or of the L-family? Why?

34.16 Cyclic structure of D-(+)-glucose. Formation of glucosides

We have seen evidence indicating that D-(+)-glucose is a pentahydroxy aldehyde. We have seen how its configuration has been established. It might seem, therefore, that D-(+)-glucose had been definitely proved to have structure I.



I

D-(+)-Glucose

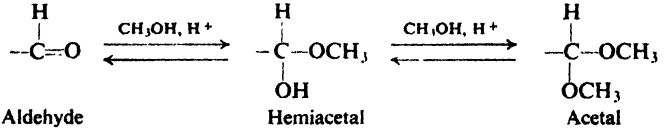
But during the time that much of the work we have just described was going on, certain facts were accumulating that were inconsistent with this structure of D-(+)-glucose. By 1895 it had become clear that the picture of D-(+)-glucose as a pentahydroxy aldehyde had to be modified.

Among the facts that had still to be accounted for were the following:

(a) **D-(+)-Glucose fails to undergo certain reactions typical of aldehydes.** Although it is readily oxidized, it gives a negative Schiff test and does not form a bisulfite addition product.

(b) **D-(+)-Glucose exists in two isomeric forms which undergo mutarotation.** When crystals of ordinary D-(+)-glucose of m.p. 146° are dissolved in water, the specific rotation gradually drops from an initial +112° to +52.7°. On the other hand, when crystals of D-(+)-glucose of m.p. 150° (obtained by crystallization at temperatures above 98°) are dissolved in water, the specific rotation gradually rises from an initial +19° to +52.7°. The form with the higher positive rotation is called **α-D-(+)-glucose** and that with lower rotation **β-D-(+)-glucose**. The change in rotation of each of these to the equilibrium value is called **mutarotation.**

(c) **D-(+)-Glucose forms two isomeric methyl D-glucosides.** Aldehydes, we remember, react with alcohols in the presence of anhydrous HCl to form acetals (Sec. 19.15). If the alcohol is, say, methanol, the acetal contains two methyl groups:



When D-(+)-glucose is treated with methanol and HCl, the product, **methyl D-glucoside**, contains only one $-CH_3$ group; yet it has properties resembling those of a full acetal. It does not spontaneously revert to aldehyde and alcohol on contact with water, but requires hydrolysis by aqueous acids.

Furthermore, not just one but two of these monomethyl derivatives of D-(+)-glucose are known, one with m.p. 165° and specific rotation +158°, and the other with m.p. 107° and specific rotation −33°. The isomer of higher positive rotation is called **methyl α-D-glucoside**, and the other is called **methyl-β-D-glucoside.** These glucosides do not undergo mutarotation, and do not reduce Tollens' or Fehling's reagent.

To fit facts like these, ideas about the structure of D-(+)-glucose had to be changed. In 1895, as a result of work by many chemists, including Tollens, Fischer, and Tanret, there emerged a picture of D-(+)-glucose as a *cyclic* structure. In 1926 the ring size was corrected, and in recent years the preferred conformation has been elucidated.

D-(+)-Glucose has the cyclic structure represented crudely by IIa and IIIa, more accurately by IIb and IIIb, and best of all by IIc and IIIc (Fig. 34.8).

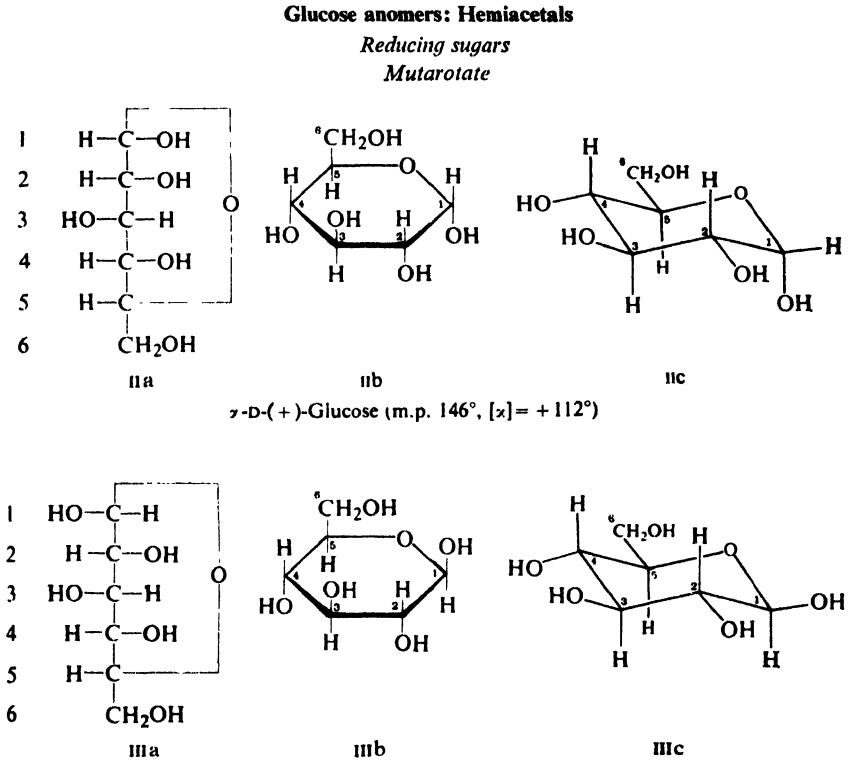


Figure 34.8. Cyclic structures of D-(+)-glucose.

D-(+)-Glucose is the hemiacetal corresponding to reaction between the aldehyde group and the C-5 hydroxyl group of the open-chain structure (I). It has a cyclic structure simply because aldehyde and alcohol are part of the same molecule.

There are two isomeric forms of D-(+)-glucose because this cyclic structure has one more chiral center than Fischer's original open-chain structure (I). α-D-(+)-Glucose and β-D-(+)-glucose are diastereomers, differing in configuration about C-1. Such a pair of diastereomers are called **anomers.**

As hemiacetals, α- and β-D-(+)-glucose are readily hydrolyzed by water. In aqueous solution either anomer is converted—via the open-chain form—into an equilibrium mixture containing both cyclic isomers. Thus mutarotation results from the ready opening and closing of the hemiacetal ring (Fig. 34.9).

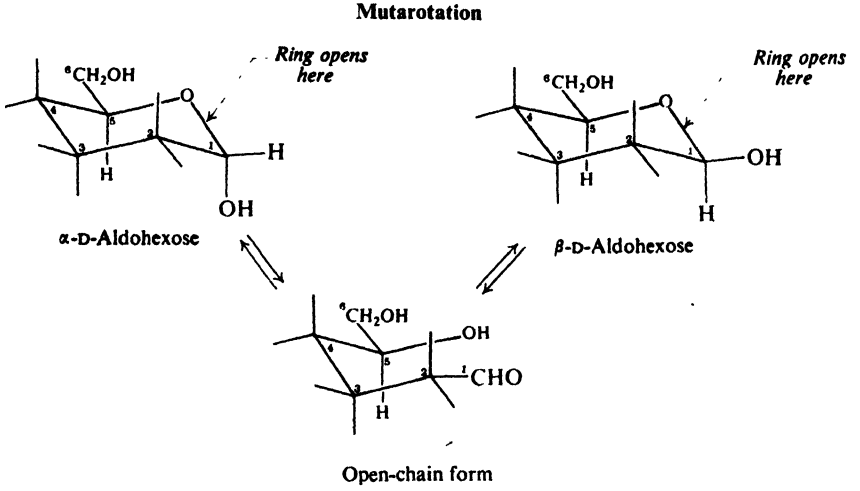


Figure 34.9. Mutarotation.

The typical aldehyde reactions of D-(+)-glucose—osazone formation, and perhaps reduction of Tollens' and Fehling's reagents—are presumably due to a small amount of open-chain compound, which is replenished as fast as it is consumed. The concentration of this open-chain structure is, however, too low (less than 0.5%) for certain easily reversible aldehyde reactions like bisulfite addition and the Schiff test.

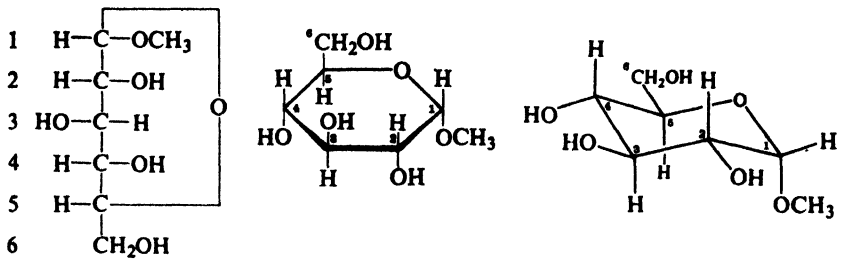
The isomeric forms of methyl D-glucoside are anomers and have the cyclic structures IV and V (Fig. 34.10).

Although formed from only one mole of methanol, they are nevertheless full acetals, the other mole of alcohol being D-(+)-glucose itself through the C-5 hydroxyl group. The glucosides do not undergo mutarotation since, being acetals, they are fairly stable in aqueous solution. On being heated with aqueous acids, they undergo hydrolysis to yield the original hemiacetals (II and III). Toward bases glycosides, like acetals generally, are stable. Since they are not readily hydrolyzed to the open-chain aldehyde by the alkali in Tollens' or Fehling's reagent, glucosides are non-reducing sugars.

Glucoside anomers: Acetals

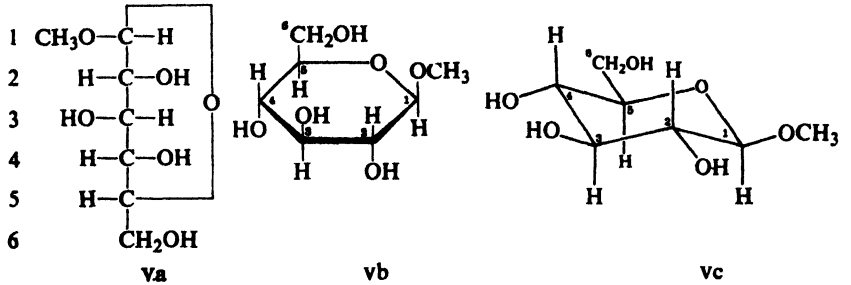
Non-reducing sugars

Do not mutarotate



IVa IVb IVc

Methyl α-D-glucoside (m.p. 165°, [α] = +158°)



Va vb vc

Methyl β-D-glucoside (m.p. 107°, [α] = −33°)

Figure 34.10. Cyclic structures of methyl D-glucosides.

Like D-(+)-glucose, other monosaccharides exist in anomeric forms capable of mutarotation, and react with alcohols to yield anomeric **glycosides.**

We have represented the cyclic structures of D-glucose and methyl D-glucoside in several different ways: β-D-glucose, for example, by IIIa, IIIb, and IIIc. At this point we should convince ourselves that all three representations correspond to the same structure, and that the configurations about C-2, C-3, C-4, and C-5 are the same as in the open-chain structure worked out by Fischer. These relationships are best seen by use of models.

We can convert the open-chain model of D-glucose into a cyclic model by joining oxygen of the C-5 —OH to the aldehyde carbon C-1. Whether we end up with the α- or β-structure depends upon which face of the flat carbonyl group we join the C-5 oxygen to. IIb and IIIb represent this ring lying on its side, so that groups that were on the right in the vertical model are directed downward, and groups that were on the left in the vertical model are directed upward. (Note particularly that the —CH_2OH group points *upward.*) In the more accurate representations IIc and IIIc, the disposition of these groups is modified by puckering of the six-membered ring, which will be discussed furtner in Sec. 34.20.

Problem 34.20 (a) From the values for the specific rotations of aqueous solutions of pure α- and β-D-(+)-glucose, and for the solution after mutarotation, calcu-

late the relative amounts of α- and of β-forms at equilibrium (assuming a negligible amount of open-chain form).

(b) From examination of structures IIc and IIIc, suggest a reason for the greater proportion of one isomer. (*Hint:* See Sec. 9.14.)

Problem 34.21 From what you learned in Secs. 19.8 and 19.15, suggest a mechanism for the acid-catalyzed mutarotation of D-(+)-glucose.

Problem 34.22 (+)-Glucose reacts with acetic anhydride to give two isomeric pentaacetyl derivatives neither of which reduces Fehling's or Tollens' reagent. Account for these facts.

34.17 Configuration about C-1

Knowledge that aldoses and their glycosides have cyclic structures immediately raises the question: what is the configuration about C-1 in each of these anomeric structures?

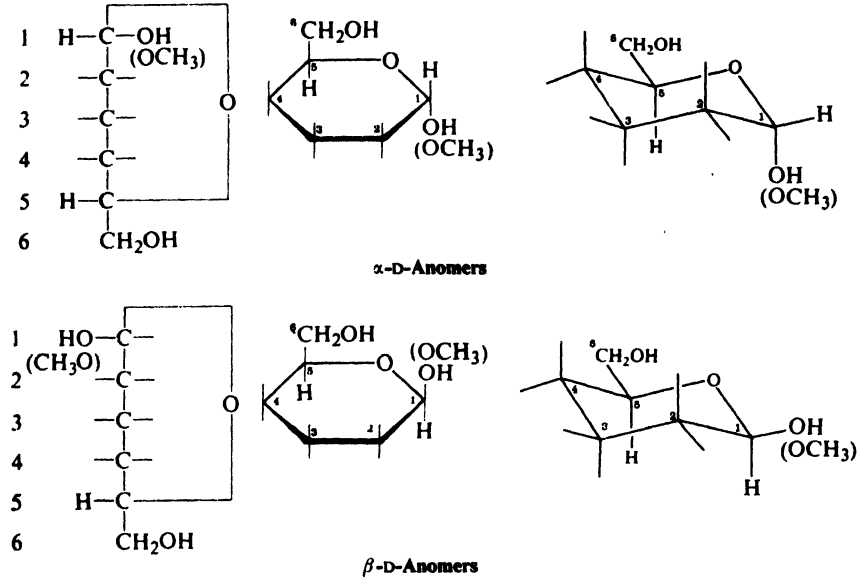


Figure 34.11. Configuration of anomers of aldohexoses.

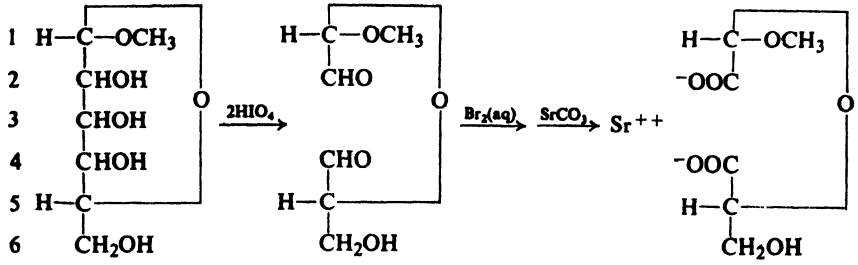
In 1909 C. S. Hudson (of the U.S. Public Health Service) made the following proposal. *In the* D-*series the more dextrorotatory member of an* α,β-*pair of anomers is to be named* α-D-, *the other being named* β-D. *In the* L *series the more levorotatory member of such a pair is given the name* α-L *and the other* β-L. Thus the enantiomer of α-D-(+)-glucose is α-L-(−)-glucose.

Furthermore, *the* —OH *or* —OCH_3 *group on* C-1 *is on the right in an* α-D-*anomer and on the left in a* β-D-*anomer*, as shown for Fig. 34.11 for aldohexoses. (Notice that "on the right" means "down" in the cyclic structure.)

Hudson's proposals have been adopted generally. Although they were originally based upon certain apparent but unproved relationships between configuration and optical rotation, all the evidence indicates that the assigned configurations are the correct ones. For example:

α-D-Glucose and methyl α-D-glucoside have the same configuration, as do β-D-glucose and methyl β-D-glucoside. *Evidence:* enzymatic hydrolysis of methyl α-D-glucoside liberates initially the more highly rotating α-D-glucose, and hydrolysis of methyl β-D-glucoside liberates initially β-D-glucose.

The configuration about C–1 is the same in the methyl α-glycosides of all the D-aldohexoses. *Evidence:* they all yield the same compound upon oxidation by HIO_4.



Oxidation destroys the chiral centers at C–2, C–3, and C–4, but configuration is preserved about C–1 and C–5. Configuration about C–5 is the same for all members of the D-family. The same products can be obtained from all these glycosides *only* if they also have the same configuration about C–1.

The C–1 —OH is on the right in the α-D-series and on the left in the β-D-series. *Evidence:* results of x-ray analysis.

Problem 34.23 (a) What products would be formed from the strontium salts shown above by treatment with dilute HCl?

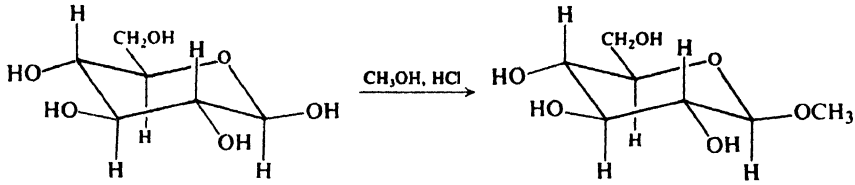
(b) An oxidation of this sort was used to confirm the configurational relationship between (+)-glucose and (+)-glyceraldehyde. How was this done?

34.18 Methylation

Before we can go on to the next aspect of the structure of D-(+)-glucose, determination of ring size, we must first learn a little more about the methylation of carbohydrates.

As we know, treatment of D-(+)-glucose with methanol and dry hydrogen chloride yields the methyl D-glucosides:

Acetal formation



and α-anomer

β-D-(+)-Glucose
Reducing sugar

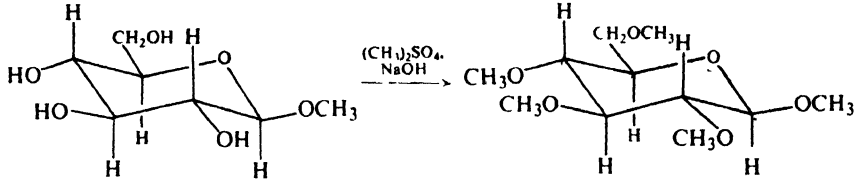
and α-anomer

Methyl β-D-glucoside
Non-reducing sugar

In this reaction, an aldehyde (or more exactly, its hemiacetal) is converted into an acetal in the usual manner.

Treatment of a methyl D-glucoside with methyl sulfate and sodium hydroxide brings about methylation of the four remaining —OH groups, and yields a methyl tetra-O-methyl-D-glucoside:

Ether formation



Methyl β-D-glucoside
Non-reducing sugar

Methyl β-2,3,4,6-tetra-O-methyl-D-glucoside
Non-reducing sugar

In this reaction, ether linkages are formed by a modification of the Williamson synthesis that is possible here because of the comparatively high acidity of these —OH groups. (Why are these —OH groups more acidic than those of an ordinary alcohol?)

There is now an $-OCH_3$ group attached to every carbon in the carbohydrate except the one joined to C-1 through the acetal linkage; if the six-membered ring structure is correct, there is an $-OCH_3$ group on every carbon except C-5.

Treatment of the methyl tetra-O-methyl-D-glucoside with dilute hydrochloric acid removes only one of these $-OCH_3$ groups, and yields a tetra-O-methyl-D-glucose (Fig. 34.12). Only the reactive acetal linkage is hydrolyzed under these mild conditions; the other four $-OCH_3$ groups, held by ordinary ether linkages, remain intact.

What we have just described for D-(+)-glucose is typical of the methylation of any monosaccharide. A fully methylated carbohydrate contains both acetal linkages and ordinary ether linkages; these are formed in different ways and are hydrolyzed under different conditions.

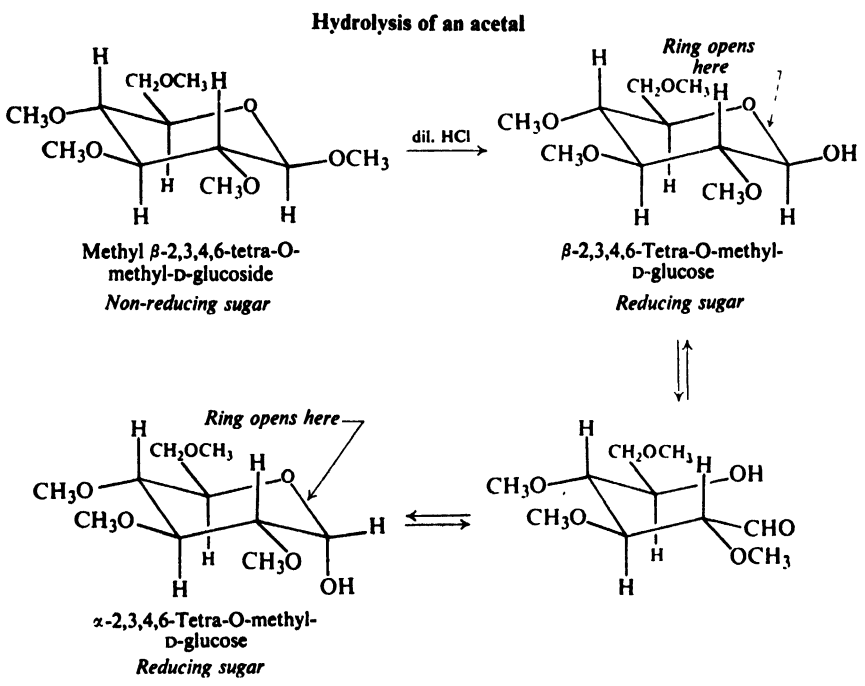


Figure 34.12. Hydrolysis of a methyl glucoside.

34.19 Determination of ring size

In the cyclic structures that we have used so far for α- and β-D-(+)-glucose and the glucosides, oxygen has been shown as joining together C–1 and C–5; that is, these compounds are represented as containing a six-membered ring. But other ring sizes are possible, in particular, a five-membered ring, one in which C–1 is joined to C–4. What is the evidence that these compounds actually contain a six-membered ring?

When methyl β-D-glucoside is treated with methyl sulfate and sodium hydroxide, and the product is hydrolyzed by dilute hydrochloric acid, there is obtained a tetra-O-methyl-D-glucose. This compound is a cyclic hemiacetal which, in solution, exists in equilibrium with a little of the open-chain form (Fig. 34.13, p. 1102).

This open-chain tetra-O-methyl-D-glucose contains an aldehyde group and four —OCH_3 groups. It also contains a free, unmethylated —OH group at whichever carbon was originally involved in the acetal ring—on C–5, if the six-membered ring is correct. *Determination of ring size becomes a matter of finding out which carbon carries the free* —OH *group.*

What would we expect to happen if the tetra-O-methyl-D-glucose were vigorously oxidized by nitric acid? The —CHO and the free —OH group should be oxidized to yield a keto acid. But, from what we know about ketones (Sec. 19.9), we would not expect oxidation to stop here: the keto acid should be cleaved on one side or the other of the carbonyl group.

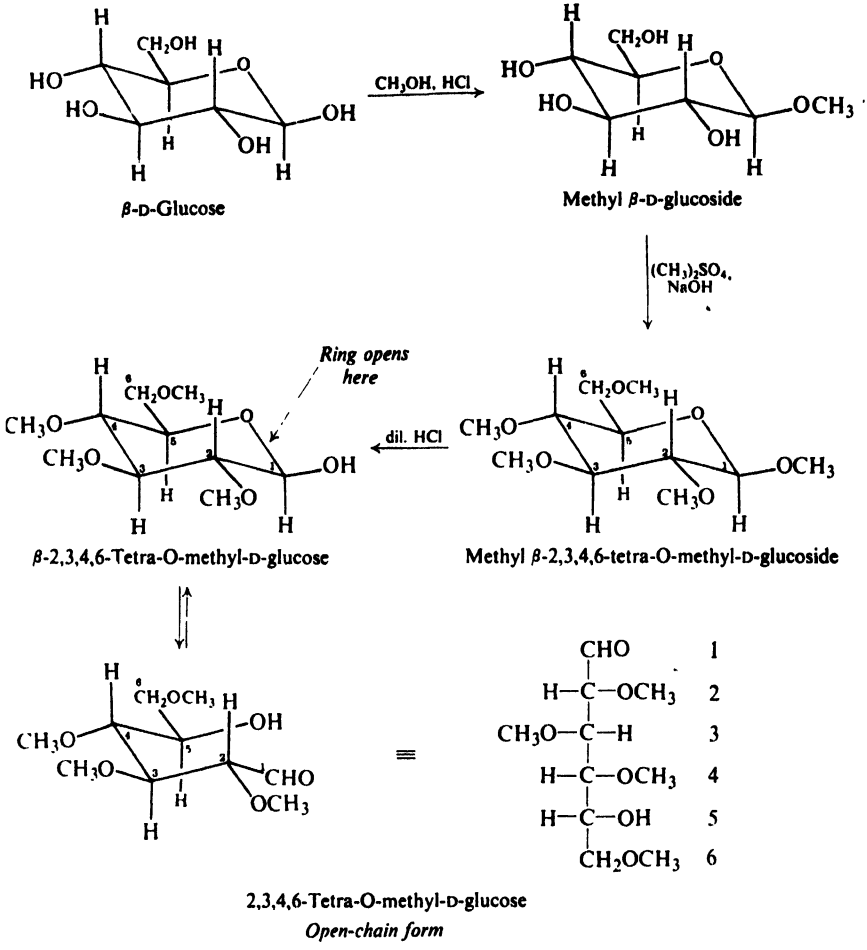


Figure 34.13. Determination of ring size. Methylation of D-glucose, followed by hydrolysis.

Oxidation actually yields a trimethoxyglutaric acid and a dimethoxysuccinic acid (Fig. 34.14, p. 1103). A mixture of five-carbon and four-carbon acids could be formed only by cleavage on either side of C–5. It must be C–5, therefore, that carries the carbonyl oxygen of the intermediate keto acid, C–5 that carries the free —OH group in the tetra-O-methyl-D-glucose, C–5 that is involved in the acetal ring of the original glucoside. Methyl β-D-glucoside must contain a six-membered ring.

By the method just described, and largely through the work of Nobel Prize winner Sir W. N. Haworth (of the University of Birmingham, England), it has been established that the six-membered ring is the common one in the glycosides of aldohexoses. Evidence of other kinds (enzymatic hydrolysis, x-ray analysis) indicates that the *free* aldohexoses, too, contain six-membered rings.

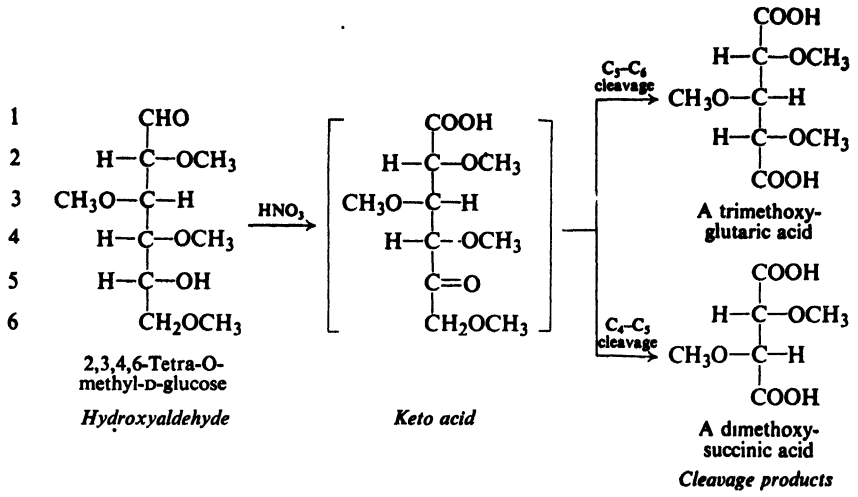


Figure 34.14. Oxidation of 2,3,4,6-tetra-O-methyl-D-glucose.

Problem 34.24 The products of HIO_4 oxidation of the methyl α-glycosides of the D-aldohexoses are shown in Sec. 34.17. What products would have been obtained if these glycosides had contained five-membered rings?

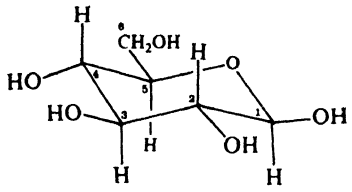
Problem 34.25 When either methyl α-L-arabinoside or methyl β-D-xyloside is methylated, hydrolyzed, and then oxidized by nitric acid, there is obtained a trimethoxyglutaric acid. (a) What ring size is indicated for these aldopentosides? (b) Predict the products of HIO_4 oxidation of each of these aldopentosides.

Problem 34.26 When crystalline methyl α-D-fructoside is methylated, hydrolyzed, oxidized by $KMnO_4$ and then nitric acid, there is obtained a trimethoxyglutaric acid. (a) What ring size is indicated for this 2-ketohexoside? (b) How does this acid compare with the one obtained from methyl α-L-arabinoside?

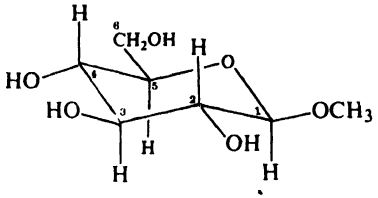
Problem 34.27 The crystalline methyl α- and β-D-glycosides we have discussed are usually prepared using methanolic HCl at 120°. When D-(+)-glucose is methylated *at room temperature*, there is obtained a liquid methyl D-glucoside. When this so-called "γ"-glucoside is methylated, hydrolyzed, and oxidized by nitric acid, there is obtained a dimethoxysuccinic acid. (a) What ring size is indicated for this "γ"-glucoside? (b) Should the dimethoxysuccinic acid be optically active or inactive? What is its absolute configuration? (c) When the liquid "γ"-glycoside obtained from D-(−)-fructose is methylated, hydrolyzed, and oxidized by nitric acid, there is also obtained a dimethoxysuccinic acid. How does this acid compare with the one in (b)?

If the name of a carbohydrate is exactly to define a particular structure, it must indicate ring size. Following a suggestion made by Haworth, carbohydrates are named to show their relationship to one of the heterocycles *pyran* or *furan*.

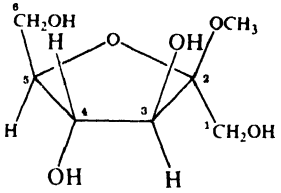
A glycose containing a six-membered ring is thus a **pyranose** and its glycosides are **pyranosides**. A glycose containing a five-membered ring is a **furanose** and its glycosides are **furanosides**. For example:



β-D-Glucopyranose



Methyl β-D-glucopyranoside



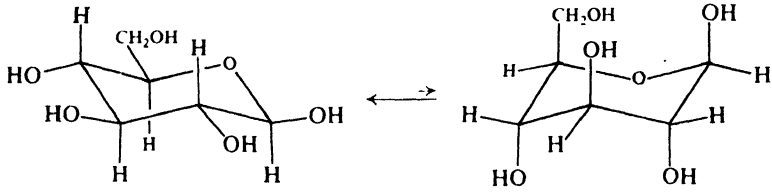
Methyl β-D-fructofuranoside

34.20 Conformation

We have followed the unraveling of the structure of D-(+)-glucose, and with it structures of the other monosaccharides, to the final working out of the ring size in 1926. Left to be discussed is one aspect whose importance has been realized only since about 1950: **conformation.**

D-(+)-Glucose contains the six-membered, pyranose ring. Since the C—O—C bond angle (111°) is very nearly equal to the tetrahedral angle (109.5°), the pyranose ring should be quite similar to the cyclohexane ring (Sec. 9.14). It should be puckered and, to minimize torsional and van der Waals strain, should exist in chair conformations in preference to twist-boat conformations. X-ray analysis shows this reasoning to be correct.

But there are *two* chair conformations possible for a D-(+)-glucopyranose anomer: I and II for β-D-(+)-glucopyranose, for example.



I

More stable:
all bulky groups equatorial

II

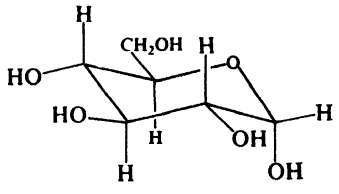
Less stable:
all bulky groups axial

β-D-(+)-Glucopyranose

Which of these is the more stable one, the one in which the molecules spend most of the time? For β-D-(+)-glucopyranose, the answer seems clear: I, in which all bulky substituents (—CH_2OH and —OH) occupy roomy equatorial positions, should certainly be much more stable than II, in which all bulky groups are crowded into axial positions. Again, x-ray analysis shows this reasoning to be correct.

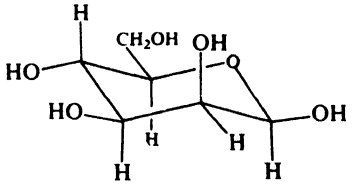
What can we say about α-D-(+)-glucose and the other aldohexoses? This problem has been largely worked out by R. E. Reeves (then at the U.S. Southern Regional Research Laboratory) through study of copper complexes.

In general, the more stable conformation is the one in which the bulkiest group, —CH_2OH, occupies an equatorial position. For example:



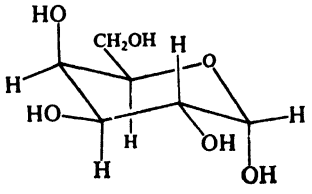
III

α-D-Glucopyranose
Stable conformation



IV

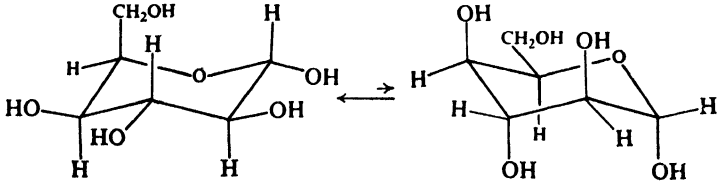
β-D-Mannopyranose
Stable conformation



V

α-D-Galactopyranose
Stable conformation

In an extreme case, to permit many —OH groups to take up equatorial positions, the —CH_2OH group may be forced into an axial position. For example:



More stable:
4 equatorial OH's,
1 axial —CH_2OH

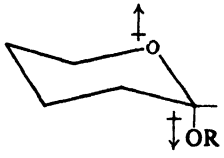
Less stable:
4 axial OH's,
1 equatorial —CH_2OH

α-D-Idopyranose

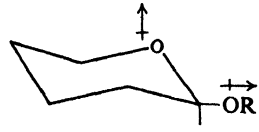
We notice that *of all* D-*aldohexoses it is* β-D-(+)-*glucose that can assume a conformation in which every bulky group occupies an equatorial position.* It is hardly accidental that β-D-(+)-glucose is the most widely occurring organic group in nature.

In drawing structural formulas or making models for the aldohexoses, a convenient point of reference is β-D-(+)-glucose. We draw the ring as shown in I—with C-1 down, C-4 up, and oxygen at the right-hand back corner—and place all —OH groups and the $-CH_2OH$ group in equatorial positions. We draw the structures of other D-family aldohexoses merely by taking into account their differences from I. Thus α-D-(+)-glucose (III) differs in configuration at C-1; β-D-mannose (IV) differs in configuration at C-2; α-D-galactose (V) differs at C-1 and C-4. L-Family compounds are, of course, mirror images of these.

In methylated and acetylated pyranoses, too, bulky groups tend to occupy equatorial positions, with one general exception: a methoxy or acetoxy group on C-1 tends to be axial. This *anomeric effect* is attributed to repulsion between the dipoles associated with the C-1 oxygen and the oxygen of the ring.



More stable



Anomeric effect

As we would expect for dipole–dipole interactions, the anomeric effect weakens as the polarity of the solvent increases (Sec. 9.10). For free sugars dissolved in water, the anomeric effect is usually outweighed by other factors; D-glucose, for example, exists predominantly as the β-anomer, with the —OH on C-1 equatorial.

Problem 34.28 Draw the conformation you predict to be the most stable for:

(a) β-D-allopyranose
(b) β-D-gulopyranose
(c) β-D-xylopyranose
(d) α-D-arabinopyranose
(e) β-L-(−)-glucopyranose
(f) β-D-(−)-fructopyranose

PROBLEMS

1. Give structures and, where possible, names of the principal products of the reaction (if any) of D-(+)-galactose with:

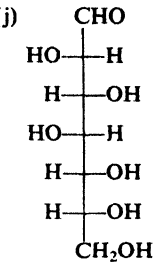
(a) hydroxylamine
(b) phenylhydrazine
(c) bromine water
(d) HNO_3
(e) HIO_4
(f) acetic anhydride
(g) benzoyl chloride, pyridine
(h) CH_3OH, HCl
(i) CH_3OH, HCl; then $(CH_3)_2SO_4$, NaOH
(j) reagents of (i), then dilute HCl
(k) reagents of (i) and (j), then vigorous oxidation
(l) H_2, Ni
(m) $NaBH_4$
(n) CN^-, H^+; then hydrolysis; then one mole $NaBH_4$
(o) H_2, Ni; then oxidation to monocarboxylic acid
(p) Br_2(aq); then pyridine; then H^+; then Na(Hg), CO_2
(q) phenylhydrazine; then benzaldehyde, H^+
(r) reagents of (q), then reduction to monocarbonyl compound

(s) Br_2(aq); then $CaCO_3$; then H_2O_2, Fe^{+++}
(t) reagents of (i), then NaOH
(u) CH_3OH, HCl; then HIO_4
(v) reagents of (u); then Br_2(aq); then dilute HCl

2. Write equations to show how D-(+)-glucose could be converted into:

(a) methyl β-D-glucoside
(b) methyl β-2,3,4,6-tetra-O-methyl-D-glucoside
(c) 2,3,4,6-tetra-O-methyl-D-glucose
(d) D-mannose
(e) L-gulose
(f) D-arabinose
(g) mesotartaric acid
(h) hexa-O-acetyl-D-glucitol
(i) D-fructose

(j)



3. Besides D-fructose, there are three D-2-ketohexoses: D-*psicose*, D-*sorbose*, and D-*tagatose*. (a) Draw the possible configurations for these three ketoses. (b) Given the configurations of all aldohexoses, tell how you could assign definite configurations to the ketoses.

4. Draw stereochemical formulas for products A through O, and tell what aldoses E, E′, F, H, I, I′, N, and O are related to.

(a) $ClCH_2CHO + BrMgC{\equiv}CMgBr + OHCCH_2Cl \longrightarrow$ A ($C_6H_8O_2Cl_2$), mainly *meso*
meso-A + KOH $\longrightarrow$ B ($C_6H_6O_2$), a diepoxide
B + H_2O, OH^- $\longrightarrow$ C ($C_6H_{10}O_4$)
C + H_2, $Pd/CaCO_3$ $\longrightarrow$ D ($C_6H_{12}O_4$)
D + cold dilute $KMnO_4$ $\longrightarrow$ E and E′ (both $C_6H_{14}O_6$)
D + peroxyformic acid $\longrightarrow$ F ($C_6H_{14}O_6$)
C + Na, NH_3 $\longrightarrow$ G ($C_6H_{12}O_4$)
G + cold dilute $KMnO_4$ $\longrightarrow$ H ($C_6H_{14}O_6$)
G + peroxyformic acid $\longrightarrow$ I and I′ (both $C_6H_{14}O_6$)

(b) *trans*-2-penten-4-yn-1-ol + $HCO_2OH \longrightarrow$ J ($C_5H_8O_3$), 4-pentyn-1,2,3-triol
J + acetic anhydride, then $Pd/CaCO_3$ + H_2 $\longrightarrow$ K ($C_{11}H_{16}O_6$)
K + HOBr $\longrightarrow$ L and M (both $C_{11}H_{17}O_7Br$)
L + hydrolysis $\longrightarrow$ N ($C_5H_{12}O_5$)
M + hydrolysis $\longrightarrow$ O ($C_5H_{12}O_5$), a racemic modification

(c) Starting with 2-butyn-1,4-diol, outline a synthesis of erythritol; of DL-threitol.

(d) 2-Butyn-1,4-diol (above) is made by reaction under pressure of acetylene with formaldehyde. What kind of reaction is this?

5. When borneol (ROH) is fed to a dog, this toxic substance is excreted as compound P, $C_6H_9O_6$—OR, where R stands for the bornyl group. Compound P does not reduce Benedict's solution. It reacts with aqueous $NaHCO_3$ with the liberation of a gas. Treatment of P with aqueous acid yields borneol (ROH) and D-glucuronic acid (Table 34.1), which is oxidized by bromine water to D-glucaric acid.

(a) What is the structure of P?

(b) Hydrolysis of the polysaccharide *pectin* (from fruits and berries) gives chiefly D-galacturonic acid; hydrolysis of the polysaccharide *algin* (from seaweed) yields D-mannuronic acid. Give the structures of these uronic acids.

(c) There are two uronic acids related to D-fructose. Draw their structures. Give the name and family of the aldonic acids formed from each "fructuronic acid" by reduction of the carbonyl group.

(d) What compound would you expect from the treatment of D-glucosone with bromine water?

6. The rate of oxidation of reducing sugars by cupric ion is found to be proportional to sugar and $[OH^-]$, and to be independent of $[Cu^{++}]$. What does the kinetics suggest about the mechanism of oxidation?

7. Upon oxidation by HIO_4 the methyl glycoside Q yields the same product (shown on p. 1099) as that obtained from methyl α-glycosides of the D-aldohexoses; however, it consumes only one mole of HIO_4 and yields *no* formic acid.

(a) How many carbon atoms are there in Q, and what is the ring size? (b) For which carbon atoms do you know the configuration? (c) When Q is methylated, hydrolyzed, and then vigorously oxidized, the dicarboxylic acid obtained is the di-O-methyl ether of (−)-tartaric acid. What is the complete structure and configuration of Q?

8. *Salicin*, $C_{13}H_{18}O_7$, found in willow (*Salix*, whence the name *salicylic*), is hydrolyzed by emulsin to D-glucose and saligenin, $C_7H_8O_2$. Salicin does not reduce Tollens' reagent. Oxidation of salicin by nitric acid yields a compound that can be hydrolyzed to D-glucose and salicylaldehyde.

Methylation of salicin gives pentamethylsalicin, which on hydrolysis gives 2,3,4,6-tetra-O-methyl-D-glucose.

What is the structure of salicin?

9. The optically inactive carbohydrate *bio-inonose*, $C_6H_{10}O_6$, reduces Benedict's solution, but does not react with bromine water. It is reduced to R and S, of formula $C_6H_{12}O_6$. Compounds R and S are oxidized by HIO_4 to six moles of HCOOH, and react with acetic anhydride to yield products of formula $C_{18}H_{24}O_{12}$. Vigorous oxidation of bio-inonose yields DL-idaric acid (the dicarboxylic acid from idose) as the only six-carbon fragment.

What is the structure of bio-inonose? Of R and S?

10. Much of what is known about photosynthesis has been learned by determining the fate of radioactive carbon dioxide, $^{14}CO_2$. The ^{14}C was found in many products, including glucose, fructose, and sucrose. To measure the radioactivity of each carbon atom in a particular molecule, degradations to one-carbon fragments were carried out.

Tell which position or positions in the molecule each of the following one-carbon products came from.

Show how the activity of the carbon atom in every position could be figured out.

(a) glucose $\xrightarrow{\text{Ruff degradation}}$ CO_2 + arabinose $\xrightarrow{\text{Ruff degradation}}$ CO_2

glucose + HIO_4 ⟶ HCHO

glucose + CH_3OH, HCl; then HIO_4 ⟶ HCOOH

glucose $\xrightarrow{\text{Lactobacillus casei}}$ 2 lactic acid (carboxyls are C-3 and C-4)

$\downarrow KMnO_4$

CO_2 + CH_3CHO $\xrightarrow{NaOI}$ CHI_3 + HCOOH

(b) ribulose (a 2-ketopentose) + HIO_4 ⟶ $HOCH_2COOH$ + 2HCOOH + HCHO

ribulose + H_2, Pt; then HIO_4 ⟶ 2HCHO + 3HCOOH

ribulose + $C_6H_5NHNH_2$ ⟶ ribosazone

ribosazone + HIO_4 ⟶ HCHO + HCOOH + $\begin{matrix} HC{=}NNHC_6H_5 \\ | \\ C{=}NNHC_6H_5 \\ | \\ CHO \end{matrix}$

11. *Nucleic acids*, the compounds that control heredity on the molecular level, are polymers composed of nucleotide units. The structures of nucleotides have been determined in the following way, as illustrated for *adenylic acid*, a nucleotide isolated from yeast cells.

Hydrolysis of adenylic acid yields one molecule each of a heterocyclic base, a sugar

T, and phosphoric acid. The base is called *adenine*, and will be represented as R_2NH. Adenylic acid has the formula $R_2N-C_5H_8O_3-OPO_3H_2$.

The sugar T is levorotatory and has the formula $C_5H_{10}O_5$; it reduces Tollens' reagent and Benedict's solution. T is oxidized by bromine water to optically active $C_5H_{10}O_6$, and by nitric acid to optically inactive $C_5H_8O_7$. T forms an osazone that is identical with the osazone obtained from another pentose, (−)-U. Degradation of (−)-U, followed by oxidation by nitric acid, yields optically inactive $C_4H_6O_6$.

(a) What is T?

Careful acidic hydrolysis of adenylic acid yields adenine and a phosphate of T, $C_5H_9O_4-OPO_3H_2$. Reduction of the phosphate with H_2/Pt yields optically inactive V, $C_5H_{11}O_4-OPO_3H_2$. Hydrolysis of V yields optically inactive W, $C_5H_{12}O_5$, which reacts with acetic anhydride to yield optically inactive X, $C_{15}H_{22}O_{10}$.

(b) What is the structure of the phosphate of T?

Adenylic acid does not reduce Tollens' reagent or Benedict's solution. When hydrolyzed by aqueous ammonia, adenylic acid yields phosphoric acid and the nucleoside *adenosine*. Treatment of adenosine with methyl sulfate and NaOH, followed by acidic hydrolysis, yields Y, a methylation product of T. Compound Y has the formula $C_8H_{16}O_5$. Vigorous oxidation of Y yields 2,3-di-O-methylmesotartaric acid and no larger fragments.

Synthesis of adenosine shows that a nitrogen atom of adenine is joined to a carbon atom in T; synthesis also shows that T has the β-configuration.

(c) Give the structure of adenylic acid, using R_2NH for the adenine unit.

(Check your answers in Fig. 37.5, p. 1179.)

12. Give structural formulas for compounds Z through II. Tell what each piece of information—(a), (b), (c), etc.—shows about the structures of Z and AA.

(a) D-glucose + CH_3COCH_3, H_2SO_4 ⟶ Z ($C_{12}H_{20}O_6$) + AA ($C_9H_{16}O_6$)
Z or AA + H_2O, OH ⟶ no reaction
Z $\xrightarrow{H_2O,\ H^+}$ AA $\xrightarrow{H_2O,\ H^+}$ D-glucose + CH_3COCH_3
To what class of compounds do Z and AA belong?

(b) Z or AA + Benedict's solution ⟶ no reaction

(c) Z + $(CH_3)_2SO_4$, NaOH ⟶ BB ($C_{13}H_{22}O_6$)
BB + H_2O, H^+ ⟶ CC ($C_7H_{14}O_6$)
CC + $C_6H_5NHNH_2$ ⟶ DD (an osazone)

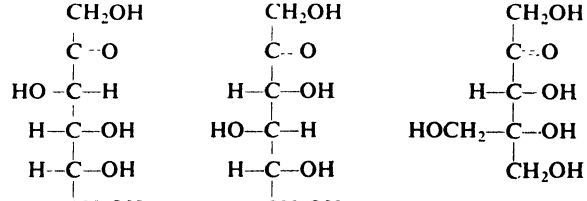
(d) AA + $(CH_3)_2SO_4$, NaOH ⟶ EE ($C_{12}H_{22}O_6$)
EE + H_2O, H^+ ⟶ FF ($C_9H_{18}O_6$)
FF + $C_6H_5NHNH_2$ ⟶ GG (an osazone)

(e) FF + $(CH_3)_2SO_4$, NaOH ⟶ 2,3,5,6-tetra-O-methyl-D-glucofuranose

(f) CC + HNO_3 ⟶ HH ($C_6H_{12}O_8$)

(g) CC + HCN, then H_2O, H^+ ⟶ II (a δ-lactone)

13. When either D-glyceraldehyde or dihydroxyacetone, $HOCH_2COCH_2OH$, is treated with base, there is obtained a mixture of the following compounds:



D-Fructose D-Sorbose and enantiomer D,L-Dendroketose

Suggest a possible mechanism for this reaction. (*Hints:* See Sec. 34.6. Count the carbons in reactants and products, and consider the reagent used.)

14. In dilute acid, hydrolysis of D-glucose-1-phosphate differs from ordinary alkyl esters of its type ($ROPO_3H_2$) in two ways: it is abnormally fast, and it takes place with cleavage of the carbon–oxygen bond. Can you suggest an explanation for its unusual behavior?

15. In Chap. 13, we learned about certain relationships between nmr spectra and the conformations of six-membered rings: in Problems 8 and 9 (p. 447), that a given proton absorbs farther downfield when in an equatorial position than when in an axial position; in Sec. 13.11, that the coupling constant, J, between *anti* protons (axial,axial) is bigger than between *gauche* protons (axial,equatorial or equatorial,equatorial). It was in the study of carbohydrates that those relationships were first recognized, chiefly by R. U. Lemieux (p. 1119).

(a) In the nmr spectra of aldopyranoses and their derivatives, the signal from one proton is found at lower fields than any of the others. Which proton is this, and why?

(b) In the nmr spectra of the two anomers of p-tetra-O-acetylxylopyranose the downfield peak appears as follows:

Anomer JJ: doublet, δ 5.39, J = 6 Hz
Anomer KK: doublet, δ 6.03, J = 3 Hz

Identify JJ and KK; that is, tell which is the α-anomer, and which is the β-anomer. Explain your answer.

(c) Answer (b) for the anomers of D-tetra-O-acetylribopyranose:

Anomer LL: doublet, δ 5.72, J = 5 Hz
Anomer MM: doublet, δ 5.82, J = 2 Hz

(d) Consider two pairs of anomers: NN and OO, and PP and QQ. One pair are the D-penta-O-acetylglucopyranoses, and the other pair are the D-penta-O-acetylmannopyranoses.

Anomer NN: doublet, δ 5.97, J = 3 Hz
Anomer OO: doublet, δ 5.68, J = 3 Hz
Anomer PP: doublet, δ 5.54, J = 8 Hz
Anomer QQ: doublet, δ 5.99, J = 3 Hz

Identify NN, OO, PP, and QQ. Explain your answer.

16. The rare sugar (−)-*mycarose* occurs as part of the molecules of several antibiotics. Using the following evidence, work out the structure and configuration of mycarose.

(i) lactone of $CH_3CH(OH)CH{=}C(CH_3)CH_2COOH \xrightarrow{syn\text{-hydroxylation}} RR\ (C_7H_{12}O_4)$
$RR + KBH_4 \longrightarrow (\pm)$-mycarose

(ii) In the nmr spectrum of (−)-mycarose and several derivatives, the coupling constant between C_4–H and C_5–H is 9.5–9.7 Hz.

(iii) methyl mycaroside + $HIO_4 \longrightarrow SS\ (C_8H_{14}O_4)$
SS + cold $KMnO_4 \longrightarrow TT\ (C_8H_{14}O_5)$
$TT \xrightarrow{hydrolysis}$ L-lactic acid

(a) Disregarding stereochemistry, what is the structure of mycarose?

(b) What are the relative configurations about C-3 and C-4? About C-4 and C-5?

(c) What is the absolute configuration at C-5?

(d) What is the absolute configuration of (−)-mycarose? To which family, D or L, does it belong? In what conformation does it preferentially exist?

(e) (−)-Mycarose can be converted into two methyl mycarosides. In the nmr spectrum of one of these, the downfield peak appears as a triplet with J = 2.4 Hz. Which anomer, α or β, is this one likely to be? What would you expect to see in the nmr spectrum of the other anomer?

(f) In the nmr spectrum of free (−)-mycarose, the downfield peak (1H) appears as two doublets with J = 9.5 and 2.5 Hz. Which anomer of mycarose, α or β, does this appear to be?

17. How do you account for the following facts? (a) In an equilibrium mixture of methyl α-D-glucoside and methyl β-D-glucoside, the α-anomer predominates. (b) In the more stable conformation of *trans*-2,5-dichloro-1,4-dioxane, both chlorines occupy axial positions.

18. From study of the nmr spectra of many compounds, Lemieux (p. 1119) found that the protons of axial acetoxy groups (—$OOCCH_3$) generally absorb at lower field than those of equatorial acetoxy groups.

(a) Draw the two chair conformations of tetra-O-acetyl-β-L-arabinopyranose. On steric grounds, which would you expect to be the more stable? Taking into account the anomeric effect, which would you expect to be the more stable?

(b) In the nmr spectrum of this compound, absorption by the acetoxy protons appears upfield as two equal peaks, at δ 1.92 and δ 2.04. How do you account for the equal sizes of these peaks? What, if anything, does this tell about the relative abundances of the two anomers?

(c) When the acetoxy group on C-1 is replaced by the deuteriated group —$OOCCD_3$, the total area of the upfield peaks is decreased, of course, from 12H to 9H. The ratio of peak areas δ 1.92:δ 2.04 is now 1.46.1.00. Which anomer predominates, and by how much? Is the predominant anomer the one you predicted to be the more stable?

Chapter 35 Carbohydrates II. Disaccharides and Polysaccharides

35.1 Disaccharides

Disaccharides are carbohydrates that are made up of two monosaccharide units. On hydrolysis a molecule of disaccharide yields two molecules of monosaccharide.

We shall study four disaccharides: (+)-**maltose** (malt sugar), (+)-**cellobiose**, (+)-**lactose** (milk sugar), and (+)-**sucrose** (cane or beet sugar). As with the monosaccharides, we shall focus our attention on the structure of these molecules: on which monosaccharides make up the disaccharide, and how they are attached to each other. In doing this, we shall also learn something about the properties of these disaccharides.

35.2 (+)-Maltose

(+)-Maltose can be obtained, among other products, by partial hydrolysis of starch in aqueous acid. (+)-Maltose is also formed in one stage of the fermentation of starch to ethyl alcohol; here hydrolysis is catalyzed by the enzyme *diastase*, which is present in malt (sprouted barley).

Let us look at some of the facts from which the structure of (+)-maltose has been deduced.

(+)-Maltose has the molecular formula $C_{12}H_{22}O_{11}$. It reduces Tollens' and Fehling's reagents and hence is a reducing sugar. It reacts with phenylhydrazine to yield an osazone, $C_{12}H_{20}O_9(=NNHC_6H_5)_2$. It is oxidized by bromine water to a monocarboxylic acid, $(C_{11}H_{21}O_{10})COOH$, *maltobionic acid*. (+)-Maltose exists in *alpha* ($[\alpha] = +168°$) and *beta* ($[\alpha] = +112°$) forms which undergo mutarotation in solution (equilibrium $[\alpha] = +136°$).

All these facts indicate the same thing: (+)-maltose contains a carbonyl group that exists in the reactive hemiacetal form as in the monosaccharides we have studied. It contains only one such "free" carbonyl group, however, since

(a) the osazone contains only two phenylhydrazine residues, and (b) oxidation by bromine water yields only a *mono*carboxylic acid.

When hydrolyzed in aqueous acid, or when treated with the enzyme *maltase* (from yeast), (+)-maltose is completely converted into D-(+)-glucose. This indicates that (+)-maltose ($C_{12}H_{22}O_{11}$) is made up of two D-(+)-glucose units joined together in some manner with the loss of one molecule of water:

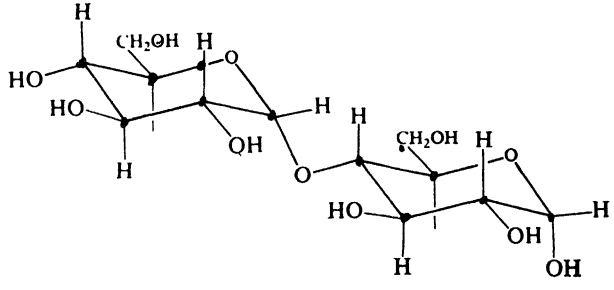
$$2C_6H_{12}O_6 - H_2O = C_{12}H_{22}O_{11}$$

Hydrolysis by acid to give a new reducing group (two reducing D-(+)-glucose molecules in place of one (+)-maltose molecule) is characteristic of glycosides; hydrolysis by the enzyme maltase is characteristic of *alpha*-glucosides. A glycoside is an acetal formed by interaction of an alcohol with a carbonyl group of a carbohydrate (Sec. 34.16); in this case the alcohol concerned can only be a second molecule of D-(+)-glucose. We conclude that (+)-maltose contains two D-(+)-glucose units, joined by an *alpha*-glucoside linkage between the carbonyl group of one D-(+)-glucose unit and an —OH group of the other.

Two questions remain: which —OH group is involved, and what are the sizes of the rings in the two D-(+)-glucose units? Answers to both these questions are given by the sequence of oxidation, methylation, and hydrolysis shown in Fig. 35.1.

Oxidation by bromine water converts (+)-maltose into the monocarboxylic acid D-maltobionic acid. Treatment of this acid with methyl sulfate and sodium hydroxide yields octa-O-methyl-D-maltobionic acid. Upon hydrolysis in acidic solution, the methylated acid yields two products, 2,3,5,6-tetra-O-methyl-D-gluconic acid and 2,3,4,6-tetra-O-methyl-D-glucose.

These facts indicate that (+)-maltose has structure I, which is given the name 4-O-(α-D-glucopyranosyl)-D-glucopyranose. It is the —OH group on C-4 that serves as the alcohol in the glucoside formation; both halves of the molecule contain the six-membered, pyranose ring.



I

(+)-Maltose (β-anomer)

4-O-(α-D-Glucopyranosyl)-D-glucopyranose

Let us see how we arrive at structure I from the experimental facts.

First of all, the initial oxidation labels (with a —COOH group) the D-glucose unit that contains the "free" aldehyde group. Next, methylation labels (as —OCH_3) every free —OH group. Finally, upon hydrolysis, the absence of a methoxyl group shows which —OH groups were *not* free.

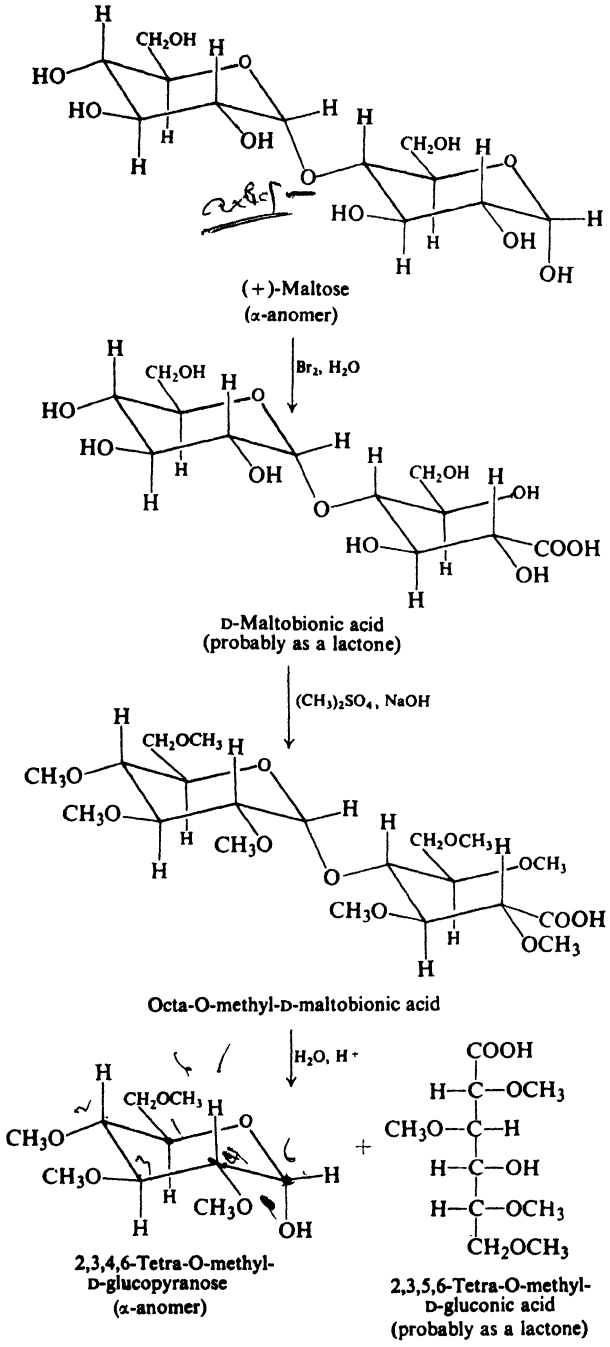


Figure 35.1. Sequence of oxidation, methylation, and hydrolysis shows that (+)-maltose is 4-O-(α-D-glucopyranosyl)-D-glucopyranose.

The oxidized product, 2,3,5,6-tetra-O-methyl-D-gluconic acid, must have arisen from the reducing (oxidizable) D-glucose unit. The presence of a free —OH group at C–4 shows that this position was not available for methylation at the maltobionic acid stage; hence it is the —OH on C–4 that is tied up in the glucoside linkage of maltobionic acid and of (+)-maltose itself. This leaves only the —OH group on C–5 to be involved in the ring of the reducing (oxidizable) unit in the original disaccharide. On the basis of these facts, therefore, we designate one D-(+)-glucose unit as a 4-O-substituted-D-glucopyranose.

The unoxidized product, 2,3,4,6-tetra-O-methyl-D-glucose, must have arisen from the non-reducing (non-oxidizable) D-glucose unit. The presence of the free —OH group at C–5 indicates that this position escaped methylation at the maltobionic acid stage; hence it is the —OH on C–5 that is tied up as a ring in maltobionic acid and in (+)-maltose itself. On the basis of these facts, therefore, we designate the second D-(+)-glucose unit as an α-D-glucopyranosyl group.

Problem 35.1 Formula I shows the structure of only the α-form of (+)-maltose. What is the structure of the β-(+)-maltose that in solution is in equilibrium with I?

Problem 35.2 The position of the free —OH group in 2,3,4,6-tetra-O-methyl-D-glucose was shown by the products of oxidative cleavage, as described in Sec. 34.19. What products would be expected from oxidative cleavage of 2,3,5,6-tetra-O-methyl-D-gluconic acid?

Problem 35.3 What products would be obtained if (+)-maltose itself were subjected to methylation and hydrolysis? What would this tell us about the structure of (+)-maltose? What uncertainty would remain in the (+)-maltose structure? Why was it necessary to oxidize (+)-maltose first before methylation?

Problem 35.4 When (+)-maltose is subjected to two successive one-carbon degradations, there is obtained a disaccharide that reduces Tollens' and Fehling's reagents but does not form an osazone. What products would be expected from the acidic hydrolysis of this disaccharide? What would these facts indicate about the structure of (+)-maltose?

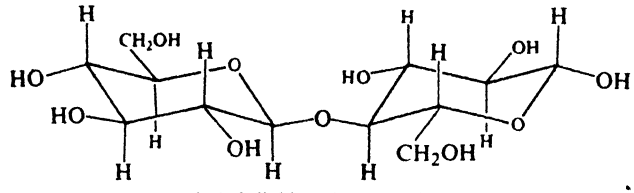
35.3 (+)-Cellobiose

When cellulose (cotton fibers) is treated for several days with sulfuric acid and acetic anhydride, a combination of acetylation and hydrolysis takes place; there is obtained the octaacetate of (+)-cellobiose. Alkaline hydrolysis of the octaacetate yields (+)-cellobiose itself.

Like (+)-maltose, (+)-cellobiose has the molecular formula $C_{12}H_{22}O_{11}$, is a reducing sugar, forms an osazone, exists in *alpha* and *beta* forms that undergo mutarotation, and can be hydrolyzed to two molecules of D-(+)-glucose. The sequence of oxidation, methylation, and hydrolysis (as described for (+)-maltose) shows that (+)-cellobiose contains two pyranose rings and a glucoside linkage to an —OH group on C–4.

(+)-Cellobiose differs from (+)-maltose in one respect: it is hydrolyzed by the enzyme *emulsin* (from bitter almonds), not by maltase. Since emulsin is known to hydrolyze only β-glucoside linkages, we can conclude that the structure of (+)-cellobiose differs from that of (+)-maltose in only one respect: the D-glucose units

are joined by a *beta* linkage rather than by an *alpha* linkage. (+)-Cellobiose is therefore 4-O-(β-D-glucopyranosyl)-D-glucopyranose.



(+)-Cellobiose (β-anomer)
4-O-(β-D-Glucopyranosyl)-D-glucopyranose

Although the D-glucose unit on the right in the formula of (+)-cellobiose may look different from the D-glucose unit on the left, this is only because it has been turned over to permit a reasonable bond angle at the glycosidic oxygen atom.

Problem 35.5 Why is *alkaline* hydrolysis of cellobiose octaacetate (better named octa-O-acetylcellobiose) to (+)-cellobiose preferred over acidic hydrolysis?

Problem 35.6 Write equations for the sequence of oxidation, methylation, and hydrolysis as applied to (+)-cellobiose.

35.4 (+)-Lactose

(+)-Lactose makes up about 5% of human milk and of cow's milk. It is obtained commercially as a by-product of cheese manufacture, being found in the *whey*, the aqueous solution that remains after the milk proteins have been coagulated. Milk *sours* when lactose is converted into lactic acid (sour, like all acids) by bacterial action (e.g., by *Lactobacillus bulgaricus*).

(+)-Lactose has the molecular formula $C_{12}H_{22}O_{11}$, is a reducing sugar, forms an osazone, and exists in *alpha* and *beta* forms which undergo mutarotation. Acidic hydrolysis or treatment with emulsin (which splits β-linkages only) converts (+)-lactose into equal amounts of D-(+)-glucose and D-(+)-galactose. (+)-Lactose is evidently a β-glycoside formed by the union of a molecule of D-(+)-glucose and a molecule of D-(+)-galactose.

The question next arises: which is the reducing monosaccharide unit and which the non-reducing unit? Is (+)-lactose a glucoside or a galactoside? Hydrolysis of lactosazone yields D-(+)-galactose and D-glucosazone; hydrolysis of *lactobionic acid* (monocarboxylic acid) yields D-gluconic acid and D-(+)-galactose (see Fig. 35.2). Clearly, it is the D-(+)-glucose unit that contains the "free" aldehyde group and undergoes osazone formation or oxidation to the acid. (+)-Lactose is thus a substituted D-glucose in which a D-galactosyl unit is attached to one of the oxygens; it is a galactoside, not a glucoside.

The sequence of oxidation, methylation, and hydrolysis gives results analogous to those obtained with (+)-maltose and (+)-cellobiose: the glycoside linkage involves an —OH group on C–4, and both units exist in the six-membered, pyranose form. (+)-Lactose is therefore 4-O-(β-D-galactopyranosyl)-D-glucopyranose.

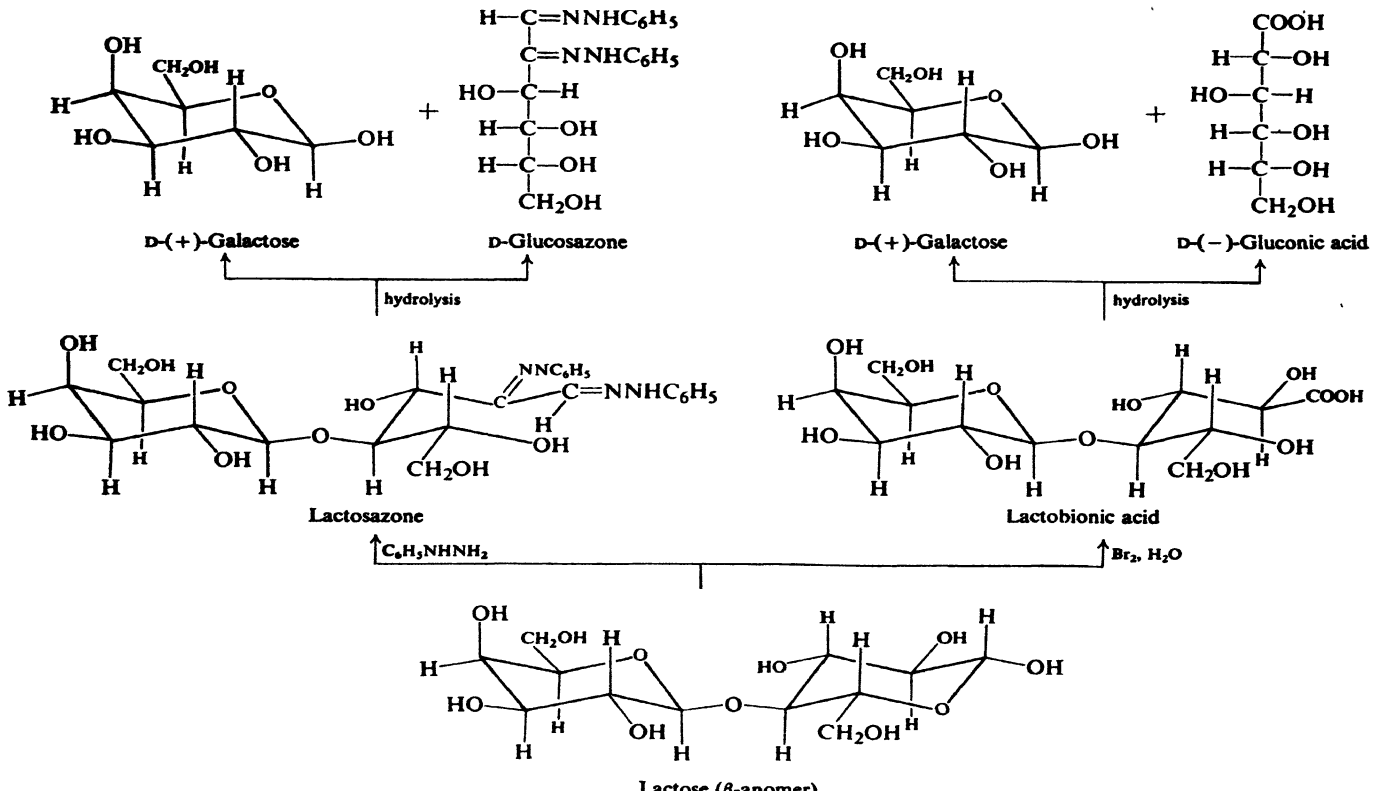


Figure 35.2. Hydrolysis of (+)-lactose derivatives. Shows that glucose is the reducing unit. (+)-Lactose is 4-O-(β-D-galactopyranosyl)-D-glucopyranose.

Problem 35.7 (a) Write equations for the sequence of oxidation, methylation, and hydrolysis as applied to (+)-lactose.

(b) What compounds would be expected from oxidative cleavage of the final products of (a)?

Problem 35.8 What products would be expected if (+)-lactose were subjected to two successive one-carbon degradations followed by acidic hydrolysis?

35.5 (+)-Sucrose

(+)-Sucrose is our common table sugar, obtained from sugar cane and sugar beets. Of organic chemicals, it is the one produced in the largest amount in pure form.

(+)-Sucrose has the molecular formula $C_{12}H_{22}O_{11}$. It does not reduce Tollens' or Fehling's reagent. It is a non-reducing sugar, and in this respect it differs from the other disaccharides we have studied. Moreover, (+)-sucrose does not form an osazone, does not exist in anomeric forms, and does not show mutarotation in solution. All these facts indicate that (+)-sucrose does not contain a "free" aldehyde or ketone group.

When (+)-sucrose is hydrolyzed by dilute aqueous acid, or by the action of the enzyme *invertase* (from yeast), it yields equal amounts of D-(+)-glucose and D-(−)-fructose. This hydrolysis is accompanied by a change in the sign of rotation from positive to negative; it is therefore often called the *inversion* of (+)-sucrose, and the levorotatory mixture of D-(+)-glucose and D-(−)-fructose obtained has been called *invert sugar*. (Honey is mostly invert sugar; the bees supply the invertase.) While (+)-sucrose has a specific rotation of +66.5° and D-(+)-glucose has a specific rotation of +52.7°, D-(−)-fructose has a large negative specific rotation of −92.4°, giving a net negative value for the specific rotation of the mixture. (Because of their opposite rotations and their importance as components of (+)-sucrose, D-(+)-glucose and D-(−)-fructose are commonly called **dextrose** and **levulose**.)

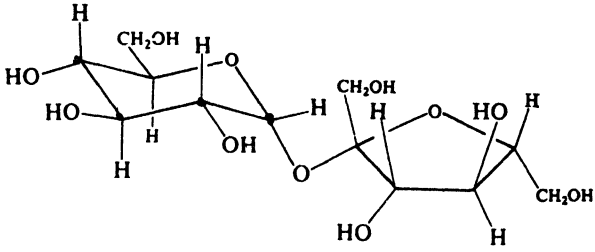
Problem 35.9 How do you account for the experimentally observed $[\alpha] = -19.9°$ for invert sugar?

(+)-Sucrose is made up of a D-glucose unit and a D-fructose unit; since there is no "free" carbonyl group, it must be both a D-glucoside and a D-fructoside. The two hexose units are evidently joined by a glycoside linkage between C–1 of glucose and C–2 of fructose, for only in this way can the single link between the two units effectively block *both* carbonyl functions.

Problem 35.10 What would be the molecular formula of (+)-sucrose if C–1 of glucose were attached to, say, C–4 of fructose, and C–2 of fructose were joined to C–4 of glucose? Would this be a reducing or non-reducing sugar?

Determination of the stereochemistry of the D-glucoside and D-fructoside linkages is complicated by the fact that both linkages are hydrolyzed at the same

time. The weight of evidence, including the results of x-ray studies and finally the synthesis of (+)-sucrose (1953), leads to the conclusion that (+)-sucrose is a *beta* D-fructoside and an *alpha* D-glucoside. (The synthesis of sucrose, by R. U. Lemieux of the Prairie Regional Laboratory, Saskatoon, Saskatchewan, has been described as "the Mount Everest of organic chemistry.")



(+)-Sucrose

α-D-Glucopyranosyl β-D-fructofuranoside

β-D-Fructofuranosyl α-D-glucopyranoside

(no anomers; *non-mutarotating*)

Problem 35.11 When (+)-sucrose is hydrolyzed enzymatically, the D-glucose initially obtained mutarotates *downward* to +52.7°. What does this fact indicate about the structure of (+)-sucrose?

Methylation and hydrolysis show that (+)-sucrose contains a D-glucopyranose unit and a D-fructofuranose unit. (The unexpected occurrence of the relatively rare five-membered, furanose ring caused no end of difficulties in both structure proof and synthesis of (+)-sucrose.) (+)-Sucrose is named equally well as either α-D-glucopyranosyl β-D-fructofuranoside or β-D-fructofuranosyl α-D-glucopyranoside.

Problem 35.12 (a) Write equations for the sequence of methylation and hydrolysis as applied to (+)-sucrose.

(b) What compounds would be expected from oxidative cleavage of the final products of (a)?

35.6 Polysaccharides

Polysaccharides are compounds made up of many—hundreds or even thousands—monosaccharide units per molecule. As in disaccharides, these units are held together by glycoside linkages, which can be broken by hydrolysis.

Polysaccharides are naturally occurring polymers, which can be considered as derived from aldoses or ketoses by condensation polymerization. A polysaccharide derived from hexoses, for example, has the general formula $(C_6H_{10}O_5)_n$. This formula, of course, tells us very little about the structure of the polysaccharide. We need to know what the monosaccharide units are and how many there are in each molecule; how they are joined to each other; and whether the huge molecules thus formed are straight-chained or branched, looped or coiled.

By far the most important polysaccharides are **cellulose** and **starch**. Both are produced in plants from carbon dioxide and water by the process of photosynthesis, and both, as it happens, are made up of D-(+)-glucose units. Cellulose is the chief structural material of plants, giving the plants rigidity and form. It is probably the most widespread organic material known. Starch makes up the reserve food supply of plants and occurs chiefly in seeds. It is more water-soluble than cellulose, more easily hydrolyzed, and hence more readily digested.

Both cellulose and starch are, of course, enormously important to us. Generally speaking, we use them in very much the same way as the plant does. We use cellulose for its structural properties: as wood for houses, as cotton or rayon for clothing, as paper for communication and packaging. We use starch as a food: potatoes, corn, wheat, rice, cassava, etc.

35.7 Starch

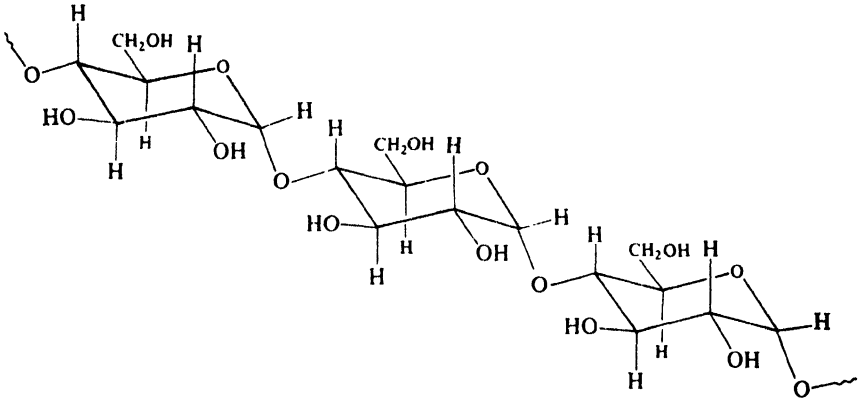
Starch occurs as granules whose size and shape are characteristic of the plant from which the starch is obtained. When intact, starch granules are insoluble in cold water; if the outer membrane has been broken by grinding, the granules swell in cold water and form a gel. When the intact granule is treated with warm water, a soluble portion of the starch diffuses through the granule wall; in hot water the granules swell to such an extent that they burst.

In general, starch contains about 20% of a water-soluble fraction called **amylose**, and 80% of a water-insoluble fraction called **amylopectin**. These two fractions appear to correspond to different carbohydrates of high molecular weight and formula $(C_6H_{10}O_5)_n$. Upon treatment with acid or under the influence of enzymes, the components of starch are hydrolyzed progressively to dext in (a mixture of low molecular weight polysaccharides), (+)-maltose, and finally D-(+)-glucose. (A mixture of all these is found in corn sirup, for example.) Both amylose and amylopectin are made up of D-(+)-glucose units, but differ in molecular size and shape.

35.8 Structure of amylose. End group analysis

(+)-Maltose is the only disaccharide that is obtained by hydrolysis of amylose, and D-(+)-glucose is the only monosaccharide. To account for this, it has been proposed that amylose is made up of chains of many D-(+)-glucose units, each unit joined by an *alpha* glycoside linkage to C-4 of the next one.

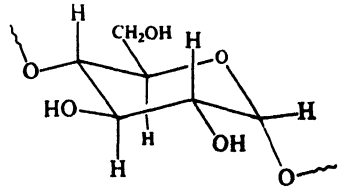
We could conceive of a structure for amylose in which α- and β-linkages regularly alternate. However, a compound of such a structure would be expected to yield (+)-cellobiose as well as (+)-maltose unless hydrolysis of the β-linkages occurred much faster than hydrolysis of the α-linkages. Since hydrolysis of the β-linkage in (+)-cellobiose is actually slower than hydrolysis of the α-linkage in (+)-maltose, such a structure seems unlikely.



Amylose
(chair conformations assumed)

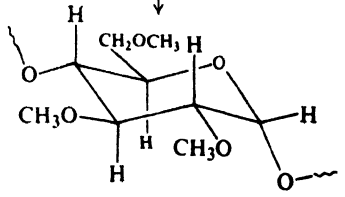
How many of these α-D-(+)-glucose units are there per molecule of amylose, and what are the shapes of these large molecules? These are difficult questions, and attempts to find the answers have made use of chemical and enzymatic methods, and of physical methods like x-ray analysis, electron microscopy, osmotic pressure and viscosity measurements, and behavior in an ultracentrifuge.

Valuable information about molecular size and shape has been obtained by the combination of methylation and hydrolysis that was so effective in studying the structures of disaccharides. D-(+)-Glucose, a monosaccharide, contains five free —OH groups and forms a pentamethyl derivative, methyl tetra-O-methyl-D-glucopyranoside. When two D-(+)-glucose units are joined together, as in (+)-maltose, each unit contains four free —OH groups; an octamethyl derivative is formed. If each D-(+)-glucose unit in amylose is joined to two others, it contains only three free —OH groups; methylation of amylose should therefore yield a compound containing only three —OCH_3 groups per glucose unit. What are the facts?



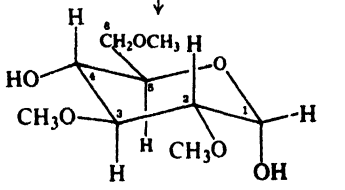
Amylose

$(CH_3)_2SO_4$, NaOH



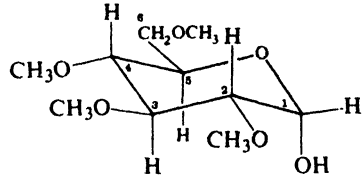
Methylated amylose

HCl



2,3,6-Tri-O-methyl-D-glucose
(α-anomer)

When amylose is methylated and hydrolyzed there is obtained, as expected, 2,3,6-tri-O-methyl-D-glucose. But there is also obtained a little bit of 2,3,4,6-tetra-O-methyl-D-glucose, amounting to about 0.2–0.4% of the total product.



2,3,4,6-Tetra-O-methyl-D-glucose
(α-anomer)

Consideration of the structure of amylose shows that this, too, is to be expected, and an important principle emerges: that of **end group analysis** (Fig. 35.3).

Each D-glucose unit in amylose is attached to two other D-glucose units, one through C–1 and the other through C–4, with C–5 in every unit tied up in the pyranose ring. As a result, free —OH groups at C–2, C–3, and C–6 are available for methylation. But this is not the case for *every* D-glucose unit. Unless the amylose chain is cyclic, it must have two ends. At one end there should be a D-glucose unit that contains a "free" aldehyde group. At the other end there should be a D-glucose unit that has a free —OH on C–4. This last D-glucose unit should undergo methylation at *four* —OH groups, and on hydrolysis should give a molecule of 2,3,4,6-tetra-O-methyl-D-glucose.

Thus each molecule of completely methylated amylose that is hydrolyzed should yield one molecule of 2,3,4,6-tetra-O-methyl-D-glucose; from the number of molecules of tri-O-methyl-D-glucose formed *along with* each molecule of the tetramethyl compound, we can calculate the length of the amylose chain.

Here we see an example of the use of end group analysis to determine chain length. A methylation that yields 0.25% of tetra-O-methyl-D-glucose shows that for every end group (with a free —OH on C–4) there are about 400 chain units.

But physical methods suggest that the chains are even longer than this. Molecular weights range from 150,000 to 600,000, indicating 1000 to 4000 glucose units per molecule. Evidently some degradation of the chain occurs during the methylation step; hydrolysis of only a few glycoside linkages in the alkaline medium would break the chain into much shorter fragments.

Problem 35.13 Consider an amylose chain of 4000 glucose units. At how many places must cleavage occur to lower the average length to 2000 units? To 1000? To 400? What percentage of the total number of glycoside links are hydrolyzed in each case?

Amylose, then, is believed to be made up of long chains, each containing 1000 or more D-glucose units joined together by α-linkages as in (+)-maltose; there is little or no branching of the chain.

Amylose is the fraction of starch that gives the intense blue color with iodine. X-ray analysis shows that the chains are coiled in the form of a helix (like a spiral

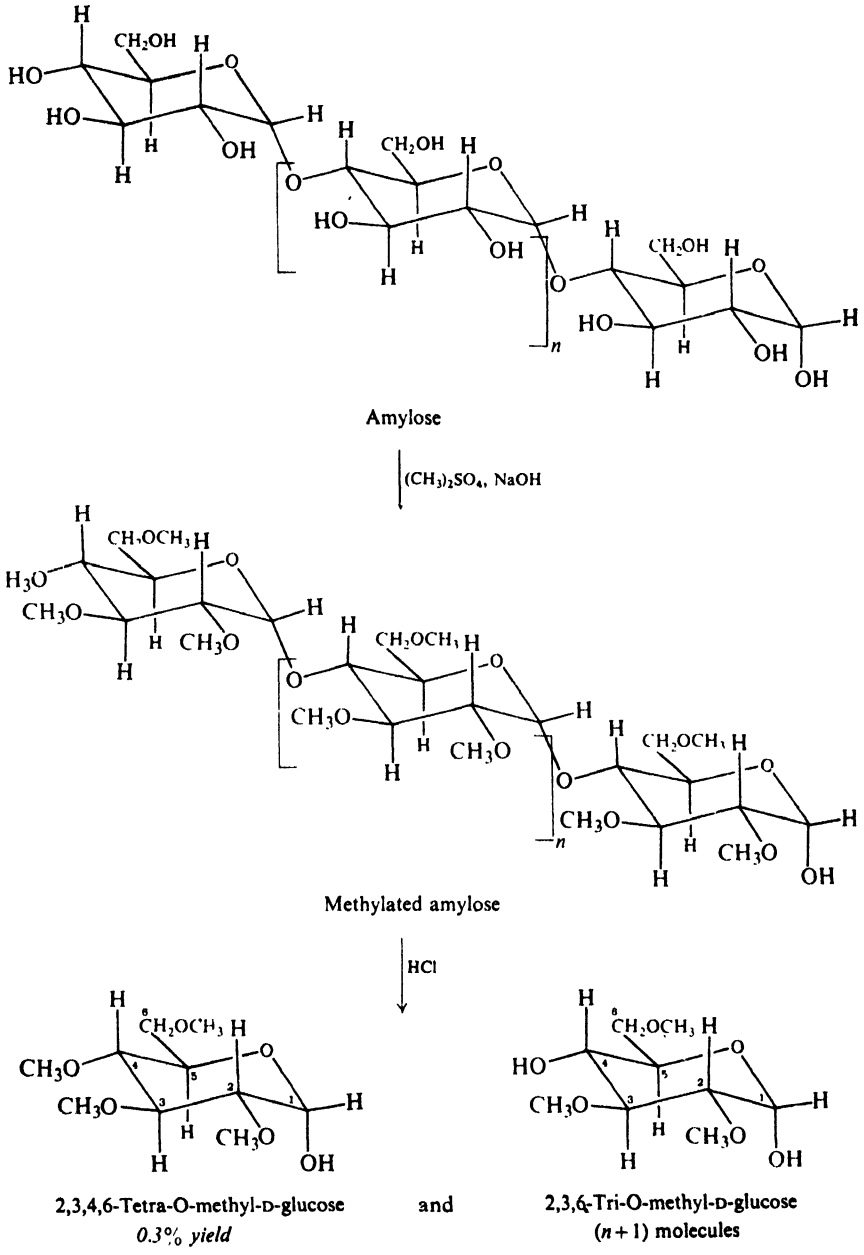
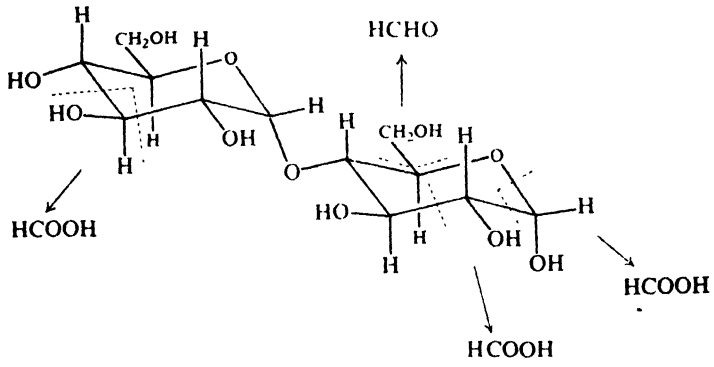


Figure 35.3. End group analysis. Hydrolysis of methylated amylose. End unit of long molecule gives 2,3,4,6-tetra-O-methyl-D-glucose; other units give 2,3,6-tri-O-methyl-D-glucose.

staircase), inside which is just enough space to accommodate an iodine molecule; the blue color is due to entrapped iodine molecules.

Problem 35.14 On the basis of certain evidence, it has been suggested that the rings of amylose have a twist-boat conformation, rather than the usual chair conformation. (a) What feature would tend to make any chair conformation unstable? (b) Suggest a twist-boat conformation that would avoid this difficulty. (*Hint:* What are the largest groups attached to a ring in amylose?)

Problem 35.15 When one mole of a disaccharide like (+)-maltose is treated with periodic acid (under conditions that minimize hydrolysis of the glycoside link), three moles of formic acid (and one of formaldehyde) are obtained.



(a) Show what would happen to amylose (see formula on p. 1123) when treated with HIO_4. (b) How could this reaction be used to determine chain length? (c) Oxidation by HIO_4 of 540 mg of amylose (from the sago plant) yielded 0.0102 millimoles of HCOOH. What is the chain length of this amylose?

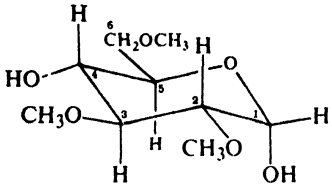
35.9 Structure of amylopectin

Amylopectin is hydrolyzed to the single disaccharide (+)-maltose; the sequence of methylation and hydrolysis yields chiefly 2,3,6-tri-O-methyl-D-glucose. Like amylose, amylopectin is made up of chains of D-glucose units, each unit joined by an *alpha* glycoside linkage to C–4 of the next one. However, its structure is more complex than that of amylose.

Molecular weights determined by physical methods show that there are up to a million D-glucose units per molecule. Yet hydrolysis of methylated amylopectin gives as high as 5% of 2,3,4,6-tetra-O-methyl-D-glucose, indicating only 20 units per chain. How can these facts be reconciled by the same structure?

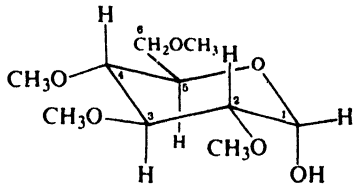
The answer is found in the following fact: along with the trimethyl and tetramethyl compounds, hydrolysis yields 2,3-di-O-methyl-D-glucose and in an amount nearly equal to that of the tetramethyl derivative.

Methylated amylopectin



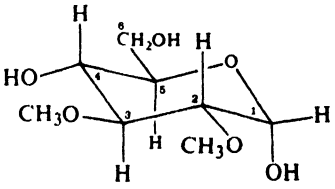
2,3,6-Tri-O-methyl-D-glucose
~90%

and



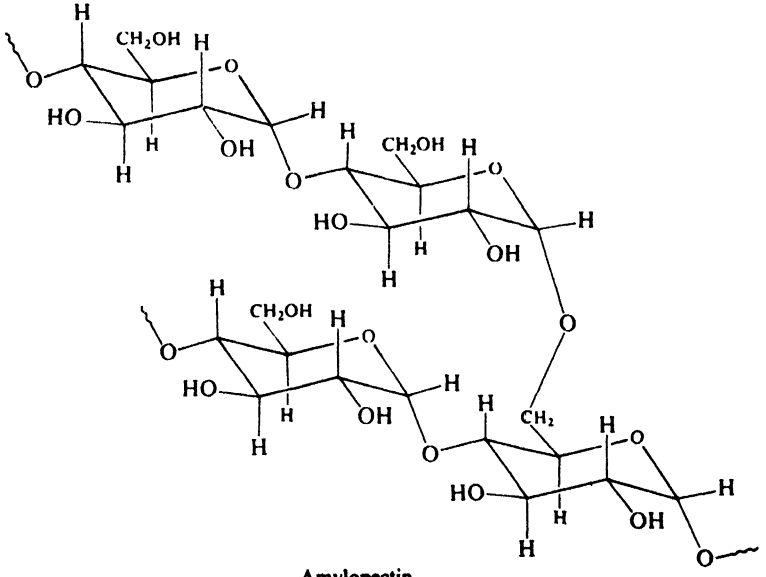
2,3,4,6-Tetra-O-methyl-D-glucose
~5%

and



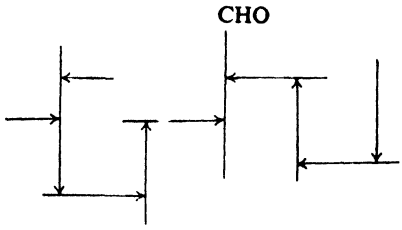
2,3-Di-O-methyl-D-glucose
~5%

Amylopectin has a highly branched structure consisting of several hundred short chains of about 20–25 D-glucose units each. One end of each of these chains is joined through C–1 to a C–6 on the next chain.



Amylopectin
(chair conformations assumed)

Schematically the amylopectin molecule is believed to be something like this:



Amylopectin

Glycogen, the form in which carbohydrate is stored in animals to be released upon metabolic demand, has a structure very similar to that of amylopectin, except that the molecules appear to be more highly branched, and to have shorter chains (12–18 D-glucose units each).

Problem 35.16 Polysaccharides known as *dextrans* have been used as substitutes for blood plasma in transfusions; they are made by the action of certain bacteria on (+)-sucrose. Interpret the following properties of a dextran: Complete hydrolysis by acid yields only D-(+)-glucose. Partial hydrolysis yields only one disaccharide and only one trisaccharide, which contain only α-glycoside linkages. Upon methylation and hydrolysis, there is obtained chiefly 2,3,4-tri-O-methyl-D-glucose, together with smaller amounts of 2,4-di-O-methyl-D-glucose and 2,3,4,6-tetra-O-methyl-D-glucose.

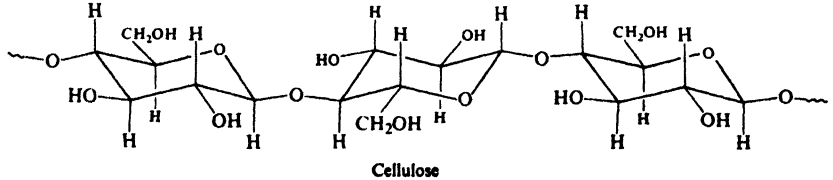
Problem 35.17 Polysaccharides called *xylans* are found along with cellulose in wood and straw. Interpret the following properties of a sample of xylan: Its large negative rotation suggests β-linkages. Complete hydrolysis by acids yields only D-(+)-xylose. Upon methylation and hydrolysis, there is obtained chiefly 2,3-di-O-methyl-D-xylose, together with smaller amounts of 2,3,4-tri-O-methyl-D-xylose and 2-O-methyl-D-xylose.

35.10 Structure of cellulose

Cellulose is the chief component of wood and plant fibers; cotton, for instance, is nearly pure cellulose. It is insoluble in water and tasteless; it is a non-reducing carbohydrate. These properties, in part at least, are due to its extremely high molecular weight.

Cellulose has the formula $(C_6H_{10}O_5)_n$. Complete hydrolysis by acid yields D-(+)-glucose as the only monosaccharide. Hydrolysis of completely methylated cellulose gives a high yield of 2,3,6-tri-O-methyl-D-glucose. Like starch, therefore, cellulose is made up of chains of D-glucose units, each unit joined by a glycoside linkage to C–4 of the next.

Cellulose differs from starch, however, in the configuration of the glycoside



Cellulose

linkage. Upon treatment with acetic anhydride and sulfuric acid, cellulose yields octa-O-acetylcellobiose; there is evidence that all glycoside linkages in cellulose, like the one in (+)-cellobiose, are *beta* linkages.

Physical methods give molecular weights for cellulose ranging from 250,000 to 1,000,000 or more; it seems likely that there are at least 1500 glucose units per molecule. End group analysis by both methylation and periodic acid oxidation gives a chain length of 1000 glucose units or more. X-ray analysis and electron microscopy indicate that these long chains lie side by side in bundles, undoubtedly held together by hydrogen bonds between the numerous neighboring —OH groups. These bundles are twisted together to form rope-like structures, which themselves are grouped to form the fibers we can see. In wood these cellulose "ropes" are embedded in lignin to give a structure that has been likened to reinforced concrete.

35.11 Reactions of cellulose

We have seen that the glycoside linkages of cellulose are broken by the action of acid, each cellulose molecule yielding many molecules of D-(+)-glucose. Now let us look briefly at reactions of cellulose in which the chain remains essentially intact. Each glucose unit in cellulose contains three free —OH groups; these are the positions at which reaction occurs.

These reactions of cellulose, carried out to modify the properties of a cheap, available, ready-made polymer, are of tremendous industrial importance.

35.12 Cellulose nitrate

Like any alcohol, cellulose forms esters. Treatment with a mixture of nitric and sulfuric acids converts cellulose into *cellulose nitrate*. The properties and uses of the product depend upon the extent of nitration.

Guncotton, which is used in making smokeless powder, is very nearly completely nitrated cellulose, and is often called *cellulose trinitrate* (three nitrate groups per glucose unit).

Pyroxylin is less highly nitrated material containing between two and three nitrate groups per glucose unit. It is used in the manufacture of plastics like celluloid and collodion, in photographic film, and in lacquers. It has the disadvantage of being flammable, and forms highly toxic nitrogen oxides upon burning.

35.13 Cellulose acetate

In the presence of acetic anhydride, acetic acid, and a little sulfuric acid, cellulose is converted into the triacetate. Partial hydrolysis removes some of the acetate groups, degrades the chains to smaller fragments (of 200–300 units each), and yields the vastly important commercial *cellulose acetate* (roughly a *di*acetate).

Cellulose acetate is less flammable than cellulose nitrate and has replaced the nitrate in many of its applications, in safety-type photographic film, for example. When a solution of cellulose acetate in acetone is forced through the fine holes of a spinnerette, the solvent evaporates and leaves solid filaments. Threads from these filaments make up the material known as *acetate rayon*.

35.14 Rayon. Cellophane

When an alcohol is treated with carbon disulfide and aqueous sodium hydroxide, there is obtained a compound called a *xanthate.*

$$\text{RONa} + \text{S=C=S} \longrightarrow \underset{\underset{\text{A xanthate}}{\text{S}}}{\text{RO—C—SNa}} \xrightarrow{\text{H}^+} \text{ROH} + \text{CS}_2$$

Cellulose undergoes an analogous reaction to form *cellulose xanthate,* which dissolves in the alkali to form a viscous colloidal dispersion called *viscose.*

When viscose is forced through a spinnerette into an acid bath, cellulose is regenerated in the form of fine filaments which yield threads of the material known as *rayon.* There are other processes for making rayon, but the viscose process is still the principal one used in the United States.

If viscose is forced through a narrow slit, cellulose is regenerated as thin sheets which, when softened by glycerol, are used for protective films (Cellophane).

Although rayon and Cellophane are often spoken of as "regenerated cellulose," they are made up of much shorter chains than the original cellulose because of degradation by the alkali treatment.

35.15 Cellulose ethers

Industrially, cellulose is alkylated by the action of alkyl chlorides (cheaper than sulfates) in the presence of alkali. Considerable degradation of the long chains is unavoidable in these reactions.

Methyl, ethyl, and benzyl ethers of cellulose are important in the production of textiles, films, and various plastic objects.

PROBLEMS

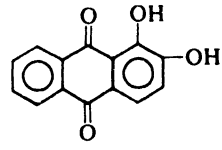
1. (+)-*Gentiobiose,* $C_{12}H_{22}O_{11}$, is found in the roots of gentians. It is a reducing sugar, forms an osazone, undergoes mutarotation, and is hydrolyzed by aqueous acid or by emulsin to D-glucose. Methylation of (+)-gentiobiose, followed by hydrolysis, gives 2,3,4,6-tetra-O-methyl-D-glucose and 2,3,4-tri-O-methyl-D-glucose. What is the structure and systematic name of (+)-gentiobiose?

2. (a) (+)-*Trehalose,* $C_{12}H_{22}O_{11}$, a non-reducing sugar found in young mushrooms, gives only D-glucose when hydrolyzed by aqueous acid or by maltase. Methylation gives an octa-O-methyl derivative that, upon hydrolysis, yields only 2,3,4,6-tetra-O-methyl-D-glucose. What is the structure and systematic name for (+)-trehalose?

(b) (−)-*Isotrehalose* and (+)-*neotrehalose* resemble trehalose in most respects. However, isotrehalose is hydrolyzed by either emulsin or maltase, and neotrehalose is hydrolyzed only by emulsin. What are the structures and systematic names for these two carbohydrates?

3. *Ruberythric acid,* $C_{25}H_{26}O_{13}$, a non-reducing glycoside, is obtained from madder root. Complete hydrolysis gives *alizarin* ($C_{14}H_8O_4$), D-glucose, and D-xylose; graded hydrolysis gives alizarin and *primeverose,* $C_{11}H_{20}O_{10}$. Oxidation of primeverose with bromine water, followed by hydrolysis, gives D-gluconic acid and D-xylose. Methylation

of primeverose, followed by hydrolysis, gives 2,3,4-tri-O-methyl-D-xylose and 2,3,4-tri-O-methyl-D-glucose.



Alizarin

What structure or structures are possible for ruberythric acid? How can any uncertainties be cleared up?

4. (+)-*Raffinose*, a non-reducing sugar found in beet molasses, has the formula $C_{18}H_{32}O_{16}$. Hydrolysis by acid gives D-fructose, D-galactose, and D-glucose; hydrolysis by the enzyme α-galactosidase gives D-galactose and sucrose; hydrolysis by invertase (a sucrose-splitting enzyme) gives D-fructose and the disaccharide *melibiose*.

Methylation of raffinose, followed by hydrolysis, gives 1,3,4,6-tetra-O-methyl-D-fructose, 2,3,4,6-tetra-O-methyl-D-galactose, and 2,3,4-tri-O-methyl-D-glucose.

What is the structure of raffinose? Of melibiose?

5. (+)-*Melezitose*, a non-reducing sugar found in honey, has the formula $C_{18}H_{32}O_{16}$. Hydrolysis by acid gives D-fructose and two moles of D-glucose; partial hydrolysis gives D-glucose and the disaccharide *turanose*. Hydrolysis by maltase gives D-glucose and D-fructose; hydrolysis by another enzyme gives sucrose.

Methylation of melezitose, followed by hydrolysis, gives 1,4,6-tri-O-methyl-D-fructose and two moles of 2,3,4,6-tetra-O-methyl-D-glucose.

(a) What structure of melezitose is consistent with these facts? What is the structure of turanose?

Melezitose reacts with four moles of HIO_4 to give two moles of formic acid but no formaldehyde.

(b) Show that the absence of formaldehyde means either a furanose or pyranose structure for the fructose unit, and either a pyranose or septanose (7-membered ring) structure for the glucose units.

(c) How many moles of HIO_4 would be consumed and how many moles of formic acid would be produced if the two glucose units had septanose rings? (d) Answer (c) for one septanose ring and one pyranose ring. (e) Answer (c) for two pyranose rings. (f) What can you say about the size of the rings in the glucose units?

(g) Answer (c) for a pyranose ring in the fructose unit; for a furanose ring.

(h) What can you say about the size of the ring in the fructose unit?

(i) Are the oxidation data consistent with the structure of melezitose you gave in (a)?

6. The sugar, (+)-*panose*, was first isolated by S. C. Pan and co-workers (at Joseph E. Seagram and Sons, Inc.) from a culture of *Aspergillus niger* on maltose. Panose has a mol. wt. of approximately 475–500. Hydrolysis gives glucose, maltose, and an isomer of maltose called isomaltose. Methylation and hydrolysis of panose gives 2,3,4-tri-, 2,3,6-tri-, and 2,3,4,6-tetra-O-methyl-D-glucose in essentially equimolar amounts. The high positive rotation of panose is considered to exclude the possibility of any β-linkages.

(a) How many monosaccharide units make up a molecule of panose? In how many ways might these be arranged?

(b) Oxidation of panose to the aldonic acid, followed by hydrolysis, gives *no* maltose; reduction of panose to panitol, followed by hydrolysis, gives glucitol and maltitol (the reduction product of maltose). Can you now draw a single structure for panose? What must be the structure of isomaltose?

(c) Panose and isomaltose can be isolated from the partial hydrolysis products of amylopectin. What bearing does this have on the structure of amylopectin?

7. Cellulose can be oxidized by N_2O_4 to $[(C_5H_7O_4)COOH]_n$. (a) What is the structure of this product? (b) What will it give on hydrolysis of the chain? What is the name of this hydrolysis product?

(c) The oxidation product in (a) is readily decarboxylated to $(C_5H_8O_4)_n$. What will this give on hydrolysis of the chain? What is the name of this hydrolysis product? Is it a D or L compound?

8. Suggest structural formulas for the following polysaccharides, neglecting the stereochemistry of the glycoside linkages:

(a) An *araban* from peanut hulls yields only L-arabinose on hydrolysis. Methylation, followed by hydrolysis, yields equimolar amounts of 2,3,5-tri-O-methyl-L-arabinose, 2,3-di-O-methyl-L-arabinose, and 3-O-methyl-L-arabinose.

(b) A *mannan* from yeast yields only D-mannose on hydrolysis. Methylation, followed by hydrolysis, yields 2,3,4,6-tetra-O-methyl-D-mannose, 2,4,6-tri-O-methyl-D-mannose, 3,4,6-tri-O-methyl-D-mannose, and 3,4-di-O-methyl-D-mannose in a molecular ratio of 2:1:1:2, together with small amounts of 2,3,4-tri-O-methyl-D-mannose.

9. When a *xylan* (see Problem 35.17, p. 1127) is boiled with dilute hydrochloric acid, a pleasant-smelling liquid, *furfural*, $C_5H_4O_2$, steam-distills. Furfural gives positive tests with Tollens' and Schiff's reagents; it forms an oxime and a phenylhydrazone but not an osazone. Furfural can be oxidized by $KMnO_4$ to A, $C_5H_4O_3$, which is soluble in aqueous $NaHCO_3$.

Compound A can be readily decarboxylated to B, C_4H_4O, which can be hydrogenated to C, C_4H_8O. C gives no tests for functional groups except solubility in cold concentrated H_2SO_4; it gives negative tests for unsaturation with dilute $KMnO_4$ or Br_2/CCl_4.

Prolonged treatment of C with HCl gives D, $C_4H_8Cl_2$, which on treatment with KCN gives E, $C_6H_8N_2$. E can be hydrolyzed to F, $C_6H_{10}O_4$, identifiable as adipic acid.

What is the structure of furfural? Of compounds A through E?

10. Give a likely structure for each of the following polysaccharides:

(a) *Alginic acid*, from sea weed, is used as a thickening agent in ice cream and other foods. Hydrolysis yields only D-mannuronic acid. Methylation, followed by hydrolysis, yields 2,3-di-O-methyl-D-mannuronic acid. (Mannuronic acid is $HOOC(CHOH)_4CHO$.) The glycoside linkages in alginic acid are thought to be *beta*.

(b) *Pectic acid* is the main constituent of the *pectin* responsible for the formation of jellies from fruits and berries. Methylation of pectic acid, followed by hydrolysis, gives only 2,3-di-O-methyl-D-galacturonic acid. The glycoside linkages in pectic acid are thought to be *alpha*.

(c) *Agar*, from sea weed, is used in the growing of microorganisms. Hydrolysis yields a 9:1:1 molar ratio of D-galactose, L-galactose, and sulfuric acid. Methylation, followed by hydrolysis, yields 2,4,6-tri-O-methyl-D-galactose, 2,3-di-O-methyl-L-galactose, and sulfuric acid in the same 9:1:1 ratio. What uncertainties are there in your proposed structure?

11. The main constituent of the capsule surrounding the Type III pneumonococcus, and the substance responsible for the specificity of its antigen–antibody reactions, is a polysaccharide (mol. wt. about 150,000). Hydrolysis yields equimolar amounts of D-glucose and D-glucuronic acid, $HOOC(CHOH)_4CHO$; careful hydrolysis gives cellobiuronic acid (the uronic acid related to cellobiose). Methylation, followed by hydrolysis, gives equimolar amounts of 2,3,6-tri-O-methyl-D-glucose and 2,4-di-O-methyl-D-glucuronic acid.

What is a likely structure for the polysaccharide?

12. Draw structures of compounds G through J:

amylose + HIO_4 $\longrightarrow$ G + a little HCOOH and HCHO

G + bromine water $\longrightarrow$ H

H + H_2O, H^+ $\longrightarrow$ I ($C_4H_8O_5$) + J ($C_2H_2O_3$)

13. (a) Show what would happen to cellulose when treated with HIO_4. (b) How could this reaction be used to determine chain length? (c) If oxidation by HIO_4 of 203 mg of a sample of cellulose yields 0.0027 millimoles of HCOOH, what is the chain length of the cellulose?

Chapter 36 Amino Acids and Proteins

36.1 Introduction

The name **protein** is taken from the Greek *proteios*, which means *first*. This name is well chosen. Of all chemical compounds, proteins must almost certainly be ranked first, for they are the substance of life.

Proteins make up a large part of the animal body, they hold it together, and they run it. They are found in all living cells. They are the principal material of skin, muscle, tendons, nerves, and blood; of enzymes, antibodies, and many hormones.

(Only the nucleic acids, which control heredity, can challenge the position of proteins; and the nucleic acids are important because they direct the synthesis of proteins.)

Chemically, proteins are high polymers. They are polyamides, and the monomers from which they are derived are the α-amino carboxylic acids. A single protein molecule contains hundreds or even thousands of amino acid units; these units can be of twenty-odd different kinds. The number of different combinations, that is, the number of different protein molecules that are possible, is almost infinite. It is likely that tens of thousands of different proteins are required to make up and run an animal body; and this set of proteins is not identical with the set required by an animal of a different kind.

In this chapter we shall look first at the chemistry of the amino acids, and then briefly at the proteins that they make up. Our chief purpose will be to see the ways in which the structures of these enormously complicated molecules are being worked out, and how, in the last analysis, all this work rests on the basic principles of organic structural theory: on the concepts of bond angle and bond length, group size and shape, hydrogen bonding, resonance, acidity and basicity, optical activity, configuration and conformation.

36.2 Structure of amino acids

Table 36.1 gives the structures and names of 26 amino acids that have been found in proteins. Certain of these (marked *e*) are the *essential* amino acids, which must be fed to young animals if proper growth is to take place; these particular amino acids evidently cannot be synthesized by the animal from the other materials in its diet.

We see that all are *alpha*-amino carboxylic acids; in two cases (proline and hydroxyproline) the amino group forms part of a pyrrolidine ring. This common feature gives the amino acids a common set of chemical properties, one of which is the ability to form the long polyamide chains that make up proteins. It is on these common chemical properties that we shall concentrate.

In other respects, the structures of these compounds vary rather widely. In addition to the carboxyl group and the amino group *alpha* to it, some amino acids contain a second carboxyl group (e.g., aspartic acid or glutamic acid), or a potential carboxyl group in the form of a carboxamide (e.g., asparagine); these are called *acidic amino acids*. Some contain a second basic group, which may be an amino group (e.g., lysine), a guanidino group (arginine), or the imidazole ring (histidine); these are called *basic amino acids*. Some of the amino acids contain benzene or heterocyclic ring systems, phenolic or alcoholic hydroxyl groups, halogen or sulfur atoms. Each of these ring systems or functional groups undergoes its own typical set of reactions.

36.3 Amino acids as dipolar ions

Although the amino acids are commonly shown as containing an amino group and a carboxyl group, $H_2NCHRCOOH$, certain properties, both physical and chemical, are not consistent with this structure:

(a) In contrast to amines and carboxylic acids, the amino acids are non-volatile crystalline solids which melt with decomposition at fairly high temperatures.

(b) They are insoluble in non-polar solvents like petroleum ether, benzene, or ether, and are appreciably soluble in water.

(c) Their aqueous solutions behave like solutions of substances of high dipole moment.

(d) Acidity and basicity constants are ridiculously low for —COOH and $-NH_2$ groups. Glycine, for example, has $K_a = 1.6 \times 10^{-10}$ and $K_b = 2.5 \times 10^{-12}$, whereas most carboxylic acids have K_a's of about 10^{-5} and most aliphatic amines have K_b's of about 10^{-4}.

All these properties are quite consistent with a dipolar ion structure for the amino acids (I).

$$^{+}H_3N-CHR-COO^{-}$$

I

Amino acids: *dipolar ions*

The physical properties—melting point, solubility, high dipole moment—are just what would be expected of such a salt. The acid–base properties also become

Table 36.1 NATURAL AMINO ACIDS

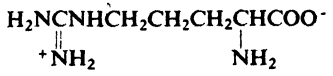
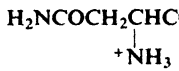
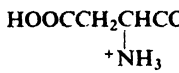
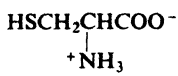
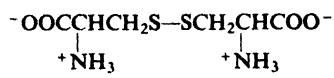
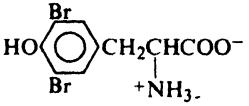
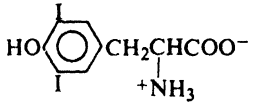
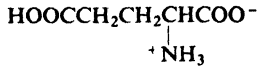
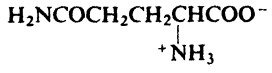
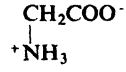
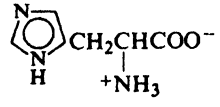
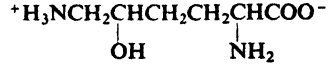
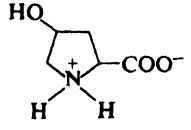
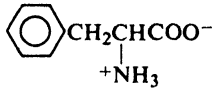
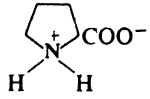
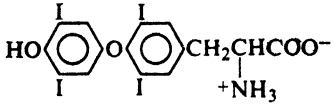
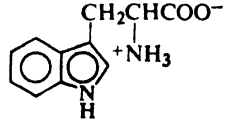
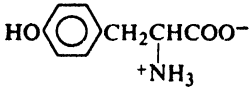
Name	Abbreviation	Formula
(+)-Alanine	Ala	
(+)-Arginine[e]	Arg	
(−)-Asparagine	Asp(NH_2)	
(+)-Aspartic acid	Asp	
(−)-Cysteine	CySH	
(−)-Cystine	CyS—SCy	
(+)-3,5-Dibromotyrosine		
(+)-3,5-Diiodotyrosine		
(+)-Glutamic acid	Glu	
(+)-Glutamine	Glu(NH_2)	
Glycine	Gly	
(−)-Histidine[e]	His	
(−)-Hydroxylysine	Hylys	
(−)-Hydroxyproline	Hypro	

Table 36.1 NATURAL AMINO ACIDS (*continued*)

Name	Abbreviation	Formula
(+)-Isoleucine[e]	Ileu	$CH_3CH_2CH(CH_3)CHCOO^-$ $^+NH_3$
(−)-Leucine[e]	Leu	$(CH_3)_2CHCH_2CHCOO^-$ $^+NH_3$
(+)-Lysine[e]	Lys	$^+H_3NCH_2CH_2CH_2CH_2CHCOO^-$ NH_2
(−)-Methionine[e]	Met	$CH_3SCH_2CH_2CHCOO^-$ $^+NH_3$
(−)-Phenylalanine[e]	Phe	
(−)-Proline	Pro	
(−)-Serine	Ser	$HOCH_2CHCOO^-$ $^+NH_3$
(−)-Threonine[e]	Thr	$CH_3CHOHCHCOO^-$ $^+NH_3$
(+)-Thyroxine		
(−)-Tryptophane[e]	Try	
(−)-Tyrosine	Tyr	
(+)-Valine[e]	Val	$(CH_3)_2CHCHCOO^-$ $^+NH_3$

[e] Essential amino acid

understandable when it is realized that the measured K_a actually refers to the acidity of an ammonium ion, RNH_3^+,

$$\underset{\text{Acid}}{{}^+H_3NCHRCOO^-} + H_2O \rightleftharpoons H_3O^+ + H_2NCHRCOO^-$$

$$K_a = \frac{[H_3O^+][H_2NCHRCOO^-]}{[{}^+H_3NCHRCOO^-]}$$

and K_b actually refers to the basicity of a carboxylate ion, $RCOO^-$.

$$\underset{\text{Base}}{{}^+H_3NCHRCOO^-} + H_2O \rightleftharpoons {}^+H_3NCHRCOOH + OH^-.$$

$$K_b = \frac{[{}^+H_3NCHRCOOH][OH^-]}{[{}^+H_3NCH_2COO^-]}$$

In aqueous solution, the acidity and basicity of an acid and its conjugate base (CH_3COOH and CH_3COO^-, or $CH_3NH_3^+$ and CH_3NH_2, for example) are related by the expression $K_a \times K_b = 10^{-14}$. From this it can be calculated that a K_a of 1.6×10^{-10} for the $-NH_3^+$ of glycine means $K_b = 6.3 \times 10^{-5}$ for $-NH_2$: a quite reasonable value for an aliphatic amine. In the same way, a K_b of 2.5×10^{-12} for the $-COO^-$ of glycine means $K_a = 4 \times 10^{-3}$ for $-COOH$: a quite reasonable value for a carboxylic acid containing the strongly electron-withdrawing (acid-strengthening) $-NH_3^+$ group.

When the solution of an amino acid is made alkaline, the dipolar ion I is converted into the anion II; the stronger base, hydroxide ion, removes a proton from the ammonium ion and displaces the weaker base, the amine.

$$\underset{\substack{\text{I}\\ \text{Stronger acid}}}{{}^+H_3NCHRCOO^-} + \underset{\substack{\\ \text{Stronger base}}}{OH^-} \rightleftharpoons \underset{\substack{\text{II}\\ \text{Weaker base}}}{H_2NCHRCOO^-} + \underset{\substack{\\ \text{Weaker acid}}}{H_2O}$$

When the solution of an amino acid is made acidic, the dipolar ion I is converted into the cation III; the stronger acid, H_3O^+, gives up a proton to the carboxylate ion, and displaces the weaker carboxylic acid.

$$\underset{\substack{\text{I}\\ \text{Stronger base}}}{{}^+H_3NCHRCOO^-} + \underset{\substack{\\ \text{Stronger acid}}}{H_3O^+} \rightleftharpoons \underset{\substack{\text{III}\\ \text{Weaker acid}}}{{}^+H_3NCHRCOOH} + \underset{\substack{\\ \text{Weaker base}}}{H_2O}$$

In summary, the acidic group of a simple amino acid like glycine is $-NH_3^+$ not $-COOH$, and the basic group is $-COO^-$ not $-NH_2$.

Problem 36.1 In quite alkaline solution, an amino acid contains two basic groups, $-NH_2$ and $-COO^-$. Which is the more basic? To which group will a proton preferentially go as acid is added to the solution? What will the product be?

Problem 36.2 In quite acidic solution, an amino acid contains two acidic groups, $-NH_3^+$ and $-COOH$. Which is the more acidic? Which group will more readily give up a proton as base is added to the solution? What will the product be?

Problem 36.3 Account for the fact that *p*-aminobenzoic acid or *o*-aminobenzoic acid does not exist appreciably as the dipolar ion, but *p*-aminobenzenesulfonic acid (*sulfanilic acid*) does. (*Hint:* What is K_b for most aromatic amines?)

We must keep in mind that ions II and III, which contain a free $-NH_2$ or $-COOH$ group, are in equilibrium with dipolar ion I; consequently, amino acids undergo reactions characteristic of amines and carboxylic acids. As ion II is removed, by reaction with benzoyl chloride, for example, the equilibrium shifts to supply more of ion II so that eventually the amino acid is completely benzoylated.

$$\underset{\text{II}}{H_2NCHRCOO^-} \underset{OH^-}{\overset{H^+}{\rightleftharpoons}} \underset{\text{I}}{{}^+H_3NCHRCOO^-} \underset{OH^-}{\overset{H^+}{\rightleftharpoons}} \underset{\text{III}}{{}^+H_3NCHRCOOH}$$

Where feasible we can speed up a desired reaction by adjusting the acidity or basicity of the solution in such a way as to increase the concentration of the reactive species.

Problem 36.4 Suggest a way to speed up (a) esterification of an amino acid; (b) acylation of an amino acid.

36.4 Isoelectric point of amino acids

What happens when a solution of an amino acid is placed in an electric field depends upon the acidity or basicity of the solution. In quite alkaline solution,

$$\underset{\text{II}}{H_2NCHRCOO^-} \underset{OH^-}{\overset{H^+}{\rightleftharpoons}} \underset{\text{I}}{{}^+H_3NCHRCOO^-} \underset{OH^-}{\overset{H^+}{\rightleftharpoons}} \underset{\text{III}}{{}^+H_3NCHRCOOH}$$

anions II exceed cations III, and there is a net migration of amino acid toward the anode. In quite acidic solution, cations III are in excess, and there is a net migration of amino acid toward the cathode. If II and III are exactly balanced, there is no net migration; under such conditions any one molecule exists as a positive ion and as a negative ion for exactly the same amount of time, and any small movement in the direction of one electrode is subsequently canceled by an equal movement back toward the other electrode. The hydrogen ion concentration of the solution in which a particular amino acid does not migrate under the influence of an electric field is called the **isoelectric point** of that amino acid.

A monoamino monocarboxylic acid, ${}^+H_3NCHRCOO^-$, is somewhat more acidic than basic (for example, glycine: $K_a = 1.6 \times 10^{-10}$ and $K_b = 2.5 \times 10^{-12}$). If crystals of such an amino acid are added to water, the resulting solution contains more of the anion II, $H_2NCHRCOO^-$, than of the cation III, ${}^+H_3NCHRCOOH$. This "excess" ionization of ammonium ion to amine ($I \rightleftarrows II + H^+$) must be repressed, by addition of acid, to reach the isoelectric point, which therefore lies somewhat on the acid side of neutrality (pH 7). For glycine, for example, the isoelectric point is at pH 6.1.

Problem 36.5 (a) Will the isoelectric point be on the acid or alkaline side of pH 7 (neutrality) for a monoamino dicarboxylic acid? (b) For a diamino monocarboxylic acid? (c) Compare each of these isoelectric points with that for glycine.

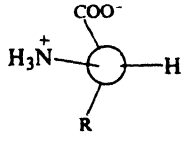
An amino acid usually shows its lowest solubility in a solution at the isoelectric point, since here there is the highest concentration of the dipolar ion. As the solution is made more alkaline or more acidic, the concentration of one of the more soluble ions, II or III, increases.

Problem 36.6 Account for the fact that sulfanilic acid dissolves in alkalies but not in acids.

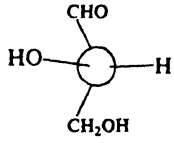
Problem 36.7 Suggest a way to separate a mixture of amino acids into three fractions: monoamino monocarboxylic acids, monoamino dicarboxylic acids (the acidic amino acids), and diamino monocarboxylic acids (the basic amino acids).

36.5 Configuration of natural amino acids

From the structures in Table 36.1, we can see that every amino acid except glycine contains at least one chiral center. As obtained by acidic or enzymatic hydrolysis of proteins, every amino acid except glycine has been found optically active. Stereochemical studies of these naturally occurring amino acids have shown that all have the same configuration about the carbon atom carrying the *alpha*-amino group, and that this configuration is the same as that in L-(−)-glyceraldehyde.



L-Amino acid



L-Glyceraldehyde

Problem 36.8 Draw all possible stereoisomeric formulas for the amino acid threonine. Naturally occurring threonine gets its name from its relationship to the tetrose *threose*; on this basis which is the correct configuration for natural threonine?

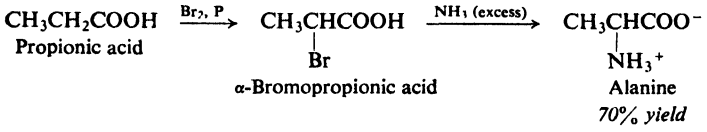
Problem 36.9 Besides threonine, there are four amino acids in Table 36.1 that can exist in more than two stereoisomeric forms. (a) What are they? (b) How many isomers are possible in each case? Indicate enantiomers, diastereomers, any *meso* compounds.

36.6 Preparation of amino acids

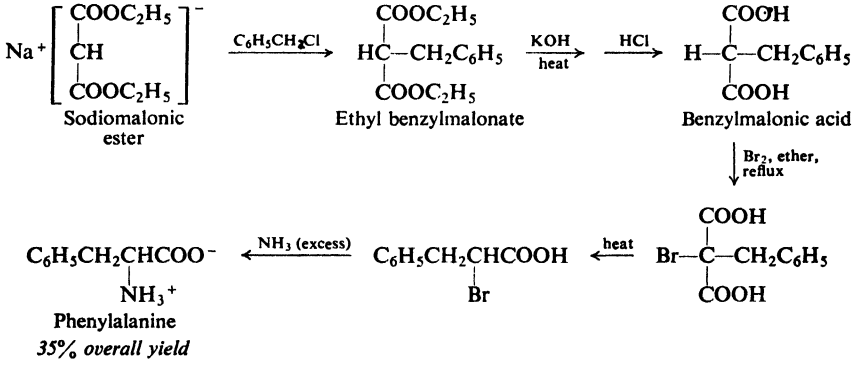
Of the many methods that have been developed for synthesizing amino acids, we shall take up only one: **amination of α-halo acids.** Considered in its various

modifications, this method is probably the most generally useful, although, like any of the methods, it cannot be applied to the synthesis of all the amino acids.

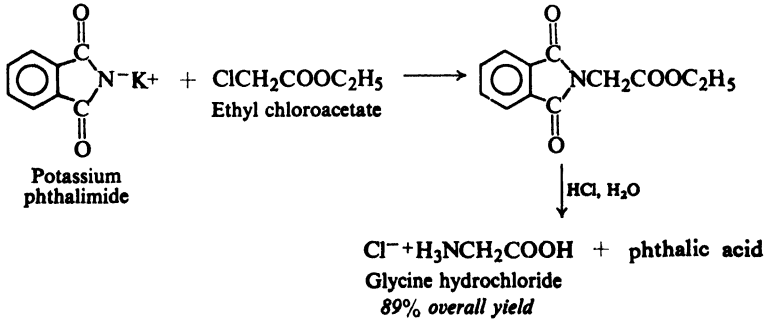
Sometimes an α-chloro or α-bromo acid is subjected to **direct ammonolysis** with a large excess (Why?) of concentrated aqueous ammonia. For example:

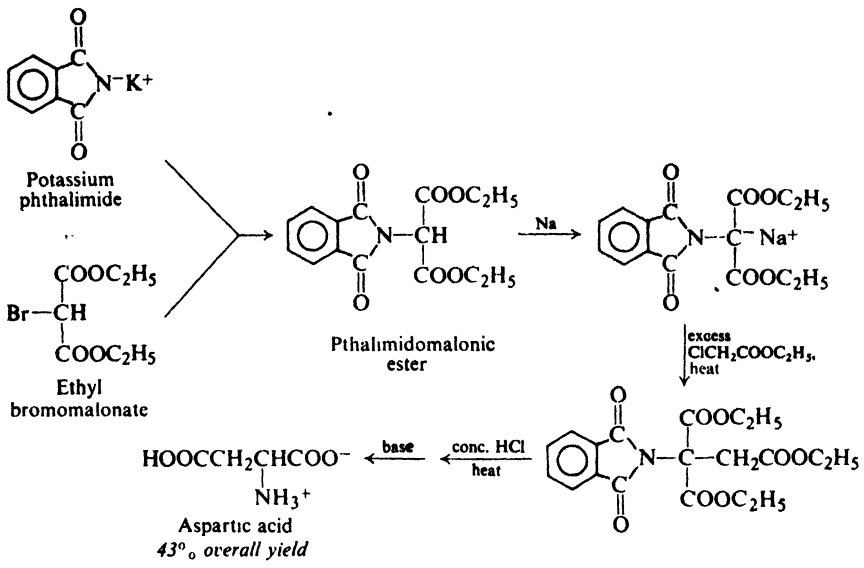


The necessary α-halo acids or esters can be prepared by the Hell-Volhard-Zelinsky halogenation of the unsubstituted acids (Sec. 18.19), or by a modification of the **malonic ester synthesis**, the usual route to the unsubstituted acids. For example:



Better yields are generally obtained by the **Gabriel phthalimide synthesis** (Problem 11, p. 744); the α-halo esters are used instead of α-halo acids (Why?). A further modification, the **phthalimidomalonic ester method**, is a combined malonic ester–Gabriel synthesis.





These synthetic amino acids are, of course, optically inactive, and must be resolved if the active materials are desired for comparison with the naturally occurring acids or for synthesis of peptides (Sec. 36.10).

Problem 36.10 Various amino acids have been made in the following ways:
Direct ammonolysis: glycine, alanine, valine, leucine, aspartic acid
Gabriel synthesis: glycine, leucine
Malonic ester synthesis: valine, isoleucine
Phthalimidomalonic ester method: serine, glutamic acid, aspartic acid
List the necessary starting materials in each case, and outline the entire sequence for one example from each group.

Problem 36.11 Acetaldehyde reacts with a mixture of KCN and NH_4Cl (**Strecker synthesis**) to give a product, $C_3H_6N_2$ (What is its structure?), which upon hydrolysis yields alanine. Show how the Strecker synthesis can be applied to the synthesis of glycine, leucine, isoleucine, valine, and serine (start with $C_2H_5OCH_2CH_2OH$). Make all required carbonyl compounds from readily available materials.

Problem 36.12 (a) Synthesis of amino acids by **reductive amination** (Sec. 22.11) is illustrated by the following synthesis of leucine:

$$\text{ethyl isovalerate} + \text{ethyl oxalate} \xrightarrow{NaOC_2H_5} A\ (C_{11}H_{18}O_5)$$

$$A + 10\%\ H_2SO_4 \xrightarrow{\text{boil}} B\ (C_6H_{10}O_3) + CO_2 + C_2H_5OH$$

$$B + NH_3 + H_2 \xrightarrow{\text{Pd, heat}} \text{leucine}$$

(b) Outline the synthesis by this method of alanine. Of glutamic acid.

36.7 Reactions of amino acids

The reactions of amino acids are in general the ones we would expect of compounds containing amino and carboxyl groups. In addition, any other groups that may be present undergo their own characteristic reactions.

Problem 36.13 Predict the products of the treatment of glycine with:

(a) aqueous NaOH
(b) aqueous HCl
(c) benzoyl chloride + aqueous NaOH
(d) acetic anhydride
(e) $NaNO_2$ + HCl
(f) $C_2H_5OH + H_2SO_4$
(g) benzyl chlorocarbonate (carbobenzoxy chloride), $C_6H_5CH_2OCOCl$

Problem 36.14 Predict the products of the following reactions:

(a) N-benzoylglycine (*hippuric acid*) + $SOCl_2$
(b) product of (a) + NH_3
(c) product of (a) + alanine
(d) product of (a) + C_2H_5OH
(e) tyrosine + Br_2(aq)
(f) asparagine + hot aqueous NaOH
(g) proline + methyl iodide
(h) tyrosine + methyl sulfate + NaOH
(i) glutamic acid + one mole $NaHCO_3$
(j) glutamic acid + excess ethyl alcohol + H_2SO_4 + heat

Problem 36.15 The reaction of primary aliphatic amines with nitrous acid gives a quantitative yield of nitrogen gas, and is the basis of the **Van Slyke determination of amino nitrogen.** What volume of nitrogen gas at S.T.P. would be liberated from 0.001 mole of: (a) leucine, (b) lysine, (c) proline?

Problem 36.16 When a solution of 9.36 mg of an unknown amino acid was treated with excess nitrous acid, there was obtained 2.01 cc of nitrogen at 748 mm and 20°. What is the minimum molecular weight for this compound? Can it be one of the amino acids found in proteins? If so, which one?

36.8 Peptides. Geometry of the peptide linkage

Peptides are amides formed by interaction between amino groups and carboxyl groups of amino acids. The amide group, $-NHCO-$, in such compounds is often referred to as the *peptide linkage.*

Depending upon the number of amino acid residues per molecule, they are known as *dipeptides*, *tripeptides*, and so on, and finally *polypeptides.* (By convention, peptides of molecular weight up to 10,000 are known as polypeptides and above that as proteins.) For example:

$^+H_3NCH_2CONHCH_2COO^-$
Gly.Gly
Glycylglycine
A dipeptide

$^+H_3NCH_2CONHCHCONHCHCOO^-$
CH_3 $CH_2C_6H_5$
Gly.Ala.Phe
Glycylalanylphenylalanine
A tripeptide

$^+H_3NCHCO(NHCHCO)_nNHCHCOO^-$
R R R
A polypeptide

A convenient way of representing peptide structures by use of standard abbreviations (see Table 36.1) is illustrated here. According to convention, the **N-terminal amino acid residue** (having the free amino group) is written at the left end, and the **C-terminal amino acid residue** (having the free carboxyl group) at the right end.

X-ray studies of amino acids and dipeptides indicate that the entire amide group is flat: carbonyl carbon, nitrogen, and the four atoms attached to them all lie in a plane. The short carbon–nitrogen distance (1.32 A as compared with 1.47 A for the usual carbon–nitrogen single bond) indicates that the carbon–nitrogen bond has considerable double-bond character (about 50%); as a result the angles of the bonds to nitrogen are similar to the angles about the trigonal carbon atom (Fig. 36.1).

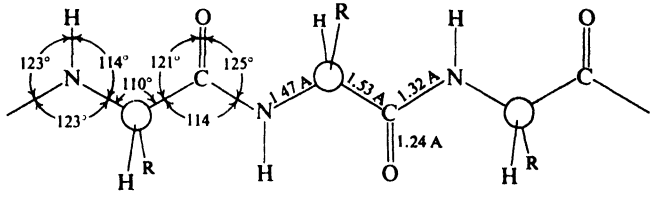


Figure 36.1. Geometry of the peptide link. Carbon–nitrogen bond has much double bond character. Carbonyl carbon, nitrogen, and atoms attached to them lie in a plane.

Problem 36.17 (a) What contributing structure(s) would account for the double-bond character of the carbon–nitrogen bond? (b) What does this resonance mean in terms of orbitals?

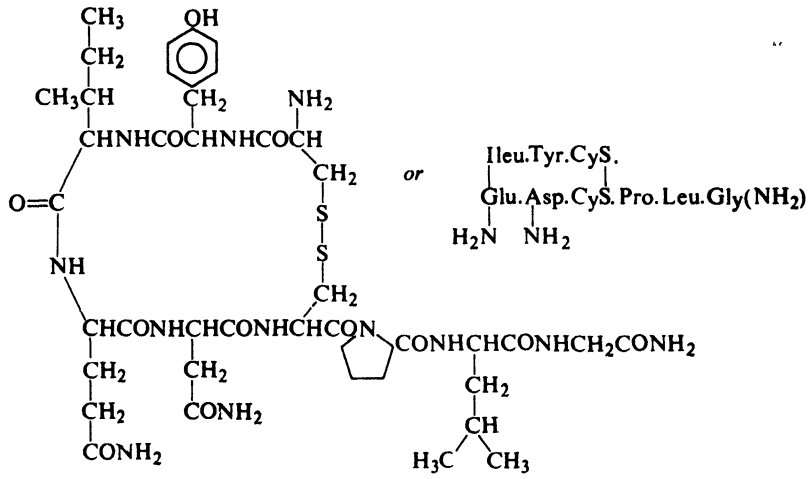
Problem 36.18 At room temperature, N,N-dimethylformamide gives the following nmr spectrum:

a singlet, δ 2.88, 3H
b singlet, δ 2.97, 3H
c singlet, δ 8.02, 1H

As the temperature is raised, signals *a* and *b* broaden and coalesce; finally, at 170°, they are merged into one sharp singlet. (a) How do you account for these observations? (b) What bearing do they have on the structure of the peptide linkage? (*Hint:* See Sec. 13.13.)

Peptides have been studied chiefly as a step toward the understanding of the much more complicated substances, the proteins. However, peptides are extremely important compounds in their own right: the tripeptide *glutathione*, for example, is found in most living cells; the nonapeptide *oxytocin* is a posterior pituitary hormone concerned with contraction of the uterus; *α-corticotropin*, made up of 39 amino acid residues, is one component of the adrenocorticotropic hormone ACTH.

Glutathione
(Glutamylcysteinylglycine)



Oxytocin

Ser.Tyr.Ser.Met.Glu.His.Phe.Arg.Try.Gly.Lys.Pro.Val.
Gly.Lys.Lys.Arg.Arg.Pro.Val.Lys.Val.Tyr.Pro.Ala.Gly.
Glu.Asp.Asp.Glu.Ala.Ser.Glu.Ala.Phe.Pro.Leu.Glu.Phe

α-Corticotropin (sheep)

We shall look at two aspects of the chemistry of peptides: how their structures are determined, and how they can be synthesized in the laboratory.

36.9 Determination of structure of peptides. Terminal residue analysis. Partial hydrolysis

To assign a structure to a particular peptide, one must know (a) what amino acid residues make up the molecule and how many of each there are, and (b) the sequence in which they follow one another along the chain.

To determine the composition of a peptide, one hydrolyzes the peptide (in acidic solution, since alkali causes racemization) and determines the amount of each amino acid thus formed. One of the best ways of analyzing a mixture of amino acids is to separate the mixture into its components by chromatography—sometimes, after conversion into the methyl esters (Why?), by gas chromatography.

From the weight of each amino acid obtained, one can calculate the number of moles of each amino acid, and in this way know the relative numbers of the various amino acid residues in the peptide. At this stage one knows what might be called

the "empirical formula" of the peptide: the relative abundance of each amino acid residue in the peptide.

Problem 36.19 An analysis of the hydrolysis products of *salmine*, a polypeptide from salmon sperm, gave the following results:

	g/100 g salmine
Isoleucine	1.28
Alanine	0.89
Valine	3.68
Glycine	3.01
Serine	7.29
Proline	6.90
Arginine	86.40

What are the relative numbers of the various amino acid residues in salmine; that is, what is its empirical formula? (Why do the weights add up to more than 100 g?)

To calculate the "molecular formula" of the peptide—the actual number of each kind of residue in each peptide molecule—one needs to know the molecular weight. Molecular weights can be determined by chemical methods and by various physical methods: osmotic pressure or light-scattering measurements, behavior in an ultracentrifuge, x-ray diffraction.

Problem 36.20 The molecular weight of salmine (see the preceding problem) is about 10,000. What are the actual numbers of the various amino acid residues in salmine; that is, what is its molecular formula?

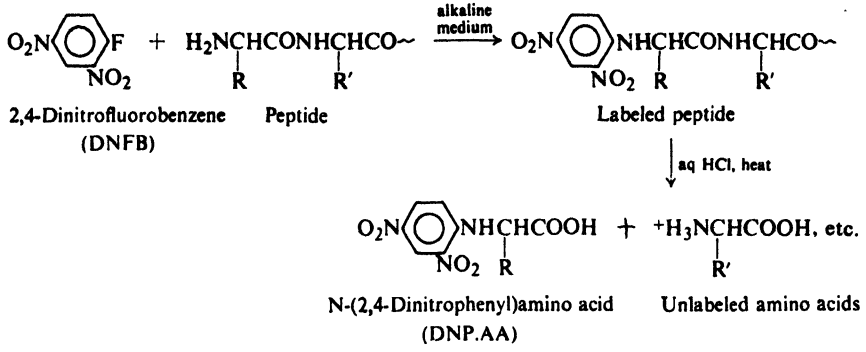
Problem 36.21 A protein was found to contain 0.29% tryptophane (mol. wt. 204). What is the minimum molecular weight of the protein?

Problem 36.22 (a) Horse hemoglobin contains 0.335% Fe. What is the minimum molecular weight of the protein? (b) Osmotic pressure measurements give a molecular weight of about 67,000. How many iron atoms are there per molecule?

There remains the most difficult job of all: to determine the sequence in which these amino acid residues are arranged along the peptide chain, that is, the structural formula of the peptide. This is accomplished by a combination of terminal residue analysis and partial hydrolysis.

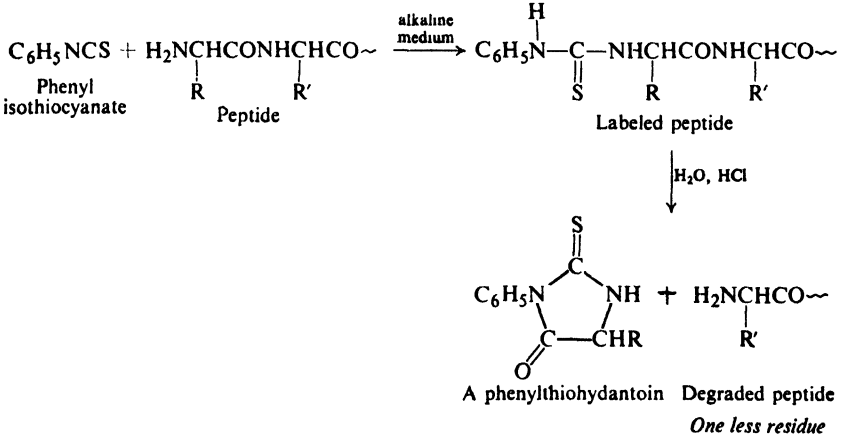
Terminal residue analysis is the identifying of the amino acid residues at the ends of the peptide chain. The procedures used depend upon the fact that the residues at the two ends are different from all the other residues and from each other: one, the *N-terminal residue*, contains a free *alpha* amino group and the other, the *C-terminal residue*, contains a free carboxyl group *alpha* to a peptide linkage.

A very successful method of identifying the N-terminal residue (introduced in 1945 by Frederick Sanger of Cambridge University) makes use of 2,4-dinitrofluorobenzene (DNFB), which undergoes nucleophilic substitution by the free amino group to give an N-dinitrophenyl (DNP) derivative. The substituted peptide



is hydrolyzed to the component amino acids, and the N-terminal residue, labeled by the 2,4-dinitrophenyl group, is separated and identified.

In its various modifications, however, the most widely used method of N-terminal residue analysis is one introduced in 1950 by Pehr Edman (of the University of Lund, Sweden). This is based upon the reaction between an amino group and phenyl isothiocyanate to form a substituted thiourea (compare Sec. 32.7). Mild hydrolysis with hydrochloric acid selectively removes the N-terminal residue as the phenylthiohydantoin, which is then identified. The great advantage of this



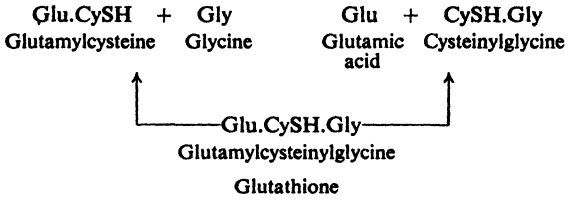
method is that it leaves the rest of the peptide chain intact, so that the analysis can be repeated and the *new* terminal group of the shortened peptide identified. In 1967, Edman reported that this analysis could be carried out *automatically* in his "protein sequenator," which is now available in commercial form. Ideally, residue after residue could be identified until the entire sequence had been determined. In actual practice, this is not feasible; after about the first 40 residues, there is interference from the accumulation of amino acids formed by (slow) hydrolysis during the acid treatment.

Problem 36.23 Edman has also devised the highly sensitive "dansyl" method in which a peptide is treated with 5-dimethylaminonaphthalenesulfonyl chloride, followed by acidic hydrolysis. A derivative of the N-terminal residue is obtained which can be followed during its analysis by virtue of its characteristic fluorescence. What is the derivative? Why does it survive the acid treatment that cleaves the peptide bonds?

The most successful method of determining the C-terminal residue has been enzymatic rather than chemical. The C-terminal residue is removed selectively by the enzyme *carboxypeptidase* (obtained from the pancreas), which cleaves only peptide linkages adjacent to *free alpha*-carboxyl groups in polypeptide chains. The analysis can be repeated on the shortened peptide and the *new* C-terminal residue identified, and so on.

In practice it is not feasible to determine the sequence of all the residues in a long peptide chain by the stepwise removal of terminal residues. Instead, the chain is subjected to partial hydrolysis (acidic or enzymatic), and the fragments formed—dipeptides, tripeptides, and so on—are identified, with the aid of terminal residue analysis. When enough of these smaller fragments have been identified. it is possible to work out the sequence of residues in the entire chain.

To take an extremely simple example, there are six possible ways in which the three amino acids making up glutathione could be arranged; partial hydrolysis to the dipeptides glutamylcysteine (Glu.CySH) and cysteinylglycine (CySH.Gly) makes it clear that the cysteine is in the middle and that the sequence Glu.CySH.Gly is the correct one.



It was by the use of the approach just outlined that structures of such peptides as oxytocin and α-corticotropin (see p. 1143) were worked out. A milestone in protein chemistry was the determination of the entire amino acid sequence in the insulin molecule by a Cambridge University group headed by Frederick Sanger, who received the Nobel Prize in 1958 for this work. (See Problem 12, p. 1162.) Since then the number—and complexity—of completely mapped proteins has grown rapidly: the four chains of hemoglobin, for example, each containing 140-odd amino acid residues; chymotrypsinogen, with a single chain 246 units long; an immunoglobulin (*gamma*-globulin) with two chains of 446 units each and two chains of 214 units each—a total of 1320 amino acid residues.

As usual, final confirmation of the structure assigned to a peptide lies in its synthesis by a method that must unambiguously give a compound of the assigned structure This problem is discussed in the following section.

Problem 36.24 Work out the sequence of amino acid residues in the following peptides:

(a) Asp,Glu,His,Phe,Val (commas indicate unknown sequence) *gives* Val.Asp + Glu.His + Phe.Val + Asp.Glu.
(b) CySH,Gly,His_2,Leu_2,Ser *gives* CySH.Gly.Ser + His.Leu.CySH + Ser.His.Leu.
(c) Arg,CySH,Glu,Gly_2,Leu,Phe_2,Tyr,Val *gives* Val.CySH.Gly + Gly.Phe.Phe + Glu Arg.Gly + Tyr.Leu.Val + Gly.Glu.Arg.

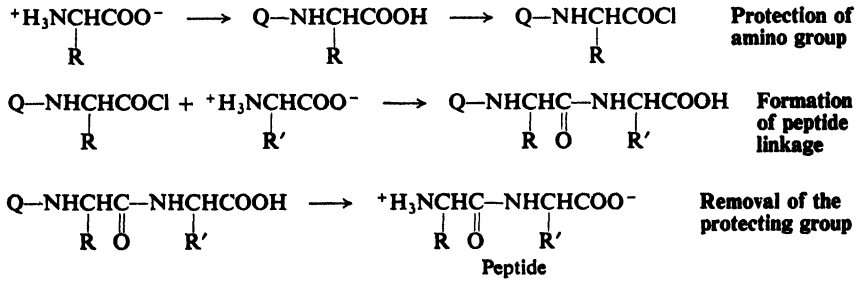
36.10 Synthesis of peptides

Methods have been developed by which a single amino acid (or sometimes a di- or tripeptide) can be polymerized to yield polypeptides of high molecular weight. These products have been extremely useful as model compounds: to show, for example, what kind of x-ray pattern or infrared spectrum is given by a peptide of known, comparatively simple structure.

Most work on peptide synthesis, however, has had as its aim the preparation of compounds identical with naturally occurring ones. For this purpose a method must permit the joining together of optically active amino acids to form chains of predetermined length and with a predetermined sequence of residues. Syntheses of this sort not only have confirmed some of the particular structures assigned to natural peptides, but also—and this is more fundamental—have proved that peptides and proteins are indeed polyamides.

It was Emil Fischer who first prepared peptides (ultimately one containing 18 amino acid residues) and thus offered support for his proposal that proteins contain the amide link. It is evidence of his extraordinary genius that Fischer played the same role in laying the foundations of peptide and protein chemistry as he did in carbohydrate chemistry.

The basic problem of peptide synthesis is one of *protecting the amino group.* In bringing about interaction between the carboxyl group of one amino acid and the amino group of a different amino acid, one must prevent interaction between the carboxyl group and the amino group of the same amino acid. In preparing glycylalanine, for example, one must prevent the simultaneous formation of glycylglycine. Reaction can be forced to take place in the desired way by attaching to one amino acid a group that renders the $-NH_2$ unreactive. There are many such protecting groups; the problem is to find one that can be removed later without destruction of any peptide linkages that may have been built up.



We could, for example, benzoylate glycine (Q = C_6H_5CO), convert this into the acid chloride, allow the acid chloride to react with alanine, and thus obtain benzoylglycylalanine. But if we attempted to remove the benzoyl group by hydrolysis, we would simultaneously hydrolyze the other amide linkage (the peptide linkage) and thus destroy the peptide we were trying to make.

Of the numerous methods developed to protect an amino group, we shall look at just one: **acylation by benzyl chlorocarbonate**, also called **carbobenzoxy chloride.** (This method was introduced in 1932 by Max Bergmann and Leonidas Zervas of the University of Berlin, later of the Rockefeller Institute.) The reagent, $C_6H_5CH_2OCOCl$, is both an ester and an acid chloride of carbonic acid, HOCOOH; it is readily made by reaction between benzyl alcohol and phosgene (carbonyl chloride), $COCl_2$. (In what order should the alcohol and phosgene be mixed?)

$$CO + Cl_2 \xrightarrow{\text{active carbon, 200}} \underset{\text{Phosgene (Carbonyl chloride)}}{Cl{-}\overset{\overset{O}{\|}}{C}{-}Cl} \xrightarrow{C_6H_5CH_2OH} \underset{\text{Carbobenzoxy chloride (Benzyl chlorocarbonate)}}{C_6H_5CH_2O{-}\overset{\overset{O}{\|}}{C}{-}Cl}$$

Like any acid chloride, the reagent can convert an amine into an amide:

$$C_6H_5CH_2O{-}\overset{\overset{O}{\|}}{C}{-}Cl + \underset{\text{Amine}}{H_2NR} \longrightarrow \underset{\text{An amide}}{C_6H_5CH_2O{-}\overset{\overset{O}{\|}}{C}{-}NHR}$$

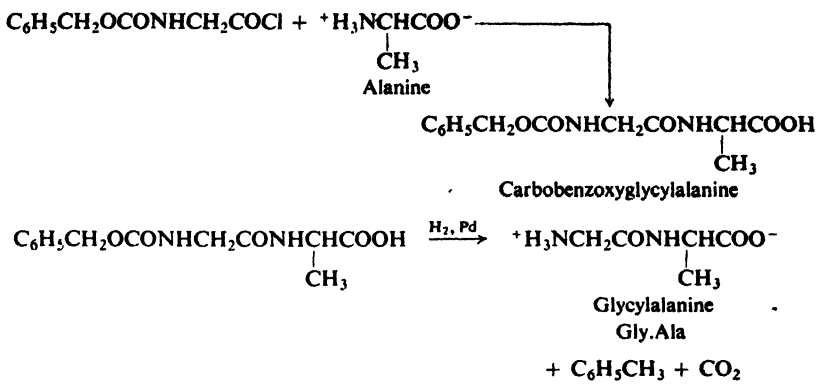
Such amides, $C_6H_5CH_2OCONHR$, differ from most amides, however, in one feature that is significant for peptide synthesis. The carbobenzoxy group can be cleaved by reagents that do not disturb peptide linkages: catalytic hydrogenation or hydrolysis with hydrogen bromide in cold acetic acid.

$$C_6H_5CH_2O{-}\overset{\overset{O}{\|}}{C}{-}NHR{-} \xrightarrow{H_2,\ Pd} C_6H_5CH_3 + \underset{\substack{\text{A carbamic acid}\\ \textit{Unstable}}}{\left[HO{-}\overset{\overset{O}{\|}}{C}{-}NHR\right]} \longrightarrow CO_2 + RNH_2$$

$$C_6H_5CH_2O{-}\overset{\overset{O}{\|}}{C}{-}NHR{-} \xrightarrow[\text{cold HOAc}]{\text{HBr,}} C_6H_5CH_2Br + \left[HO{-}\overset{\overset{O}{\|}}{C}{-}NHR\right] \longrightarrow CO_2 + RNH_2$$

The carbobenzoxy method is illustrated by the synthesis of glycylalanine (Gly.Ala):

$$\underset{\substack{\text{Carbobenzoxy}\\ \text{chloride}}}{C_6H_5CH_2OCOCl} + \underset{\text{Glycine}}{{}^+H_3NCH_2COO^-} \longrightarrow \underset{\text{Carbobenzoxyglycine}}{C_6H_5CH_2OCONHCH_2COOH} \xrightarrow{SOCl_2} \underset{\substack{\text{Acid chloride of}\\ \text{carbobenzoxyglycine}}}{C_6H_5CH_2OCONHCH_2COCl}$$



Problem 36.25 (a) How could the preceding synthesis be extended to the tripeptide glycylalanylphenylalanine (Gly.Ala.Phe)?

(b) How could the carbobenzoxy method be used to prepare alanylglycine (Ala.Gly)?

Methods like this can be repeated over and over with the addition of a new unit each time. In this way the hormone oxytocin (p. 1143) was synthesized by Vincent du Vigneaud of Cornell Medical College, who received the Nobel Prize in 1955 for this and other work. In 1963, the total synthesis of the insulin molecule—with the 51 amino acid residues in the sequence mapped out by Sanger—was reported.

But the bottleneck in such syntheses is the need to isolate and purify the new peptide made in each cycle; the time required is enormous, and the yield of product steadily dwindles. A major break-through came with the development of *solid-phase* peptide synthesis by R. Bruce Merrifield at Rockefeller University. Synthesis is carried out with the growing peptide *attached* chemically to polystyrene beads; as each new unit is added, the reagents and by-products are simply washed away, leaving the growing peptide behind, ready for another cycle. The method was automated, and in 1969 Merrifield announced that, using his "protein-making machine," he had synthesized—in *six weeks*—the enzyme ribonuclease, made up of 124 amino acid residues.

Problem 36.26 Give formulas for compounds A–G, and tell what is happening in each reaction.

polystyrene + $CH_3OCH_2Cl \xrightarrow{SnCl_4}$ A + CH_3OH

A + $C_6H_5CH_2OCONHCH_2COO^-{}^+NHEt_3 \longrightarrow$ B + Et_3NHCl

B + dil. HBr $\longrightarrow$ C + $C_6H_5CH_2Br$ + CO_2

C + carbobenzoxyalanyl chloride $\longrightarrow$ D

D + dil. HBr $\longrightarrow$ E + $C_6H_5CH_2Br$ + CO_2

E + HBr $\xrightarrow{CF_3COOH}$ F ($C_5H_{10}O_3N_2$) + G

36.11 Proteins. Classification and function. Denaturation

Proteins are divided into two broad classes: **fibrous proteins**, which are insoluble in water, and **globular proteins**, which are soluble in water or aqueous

solutions of acids, bases, or salts. (Because of the large size of protein molecules, these solutions are colloidal.) The difference in solubility between the two classes is related to a difference in molecular shape, which is indicated in a rough way by their names.

Molecules of fibrous proteins are long and thread-like, and tend to lie side by side to form fibers; in some cases they are held together at many points by hydrogen bonds. As a result, the intermolecular forces that must be overcome by a solvent are very strong.

Molecules of globular proteins are folded into compact units that often approach spheroidal shapes. The folding takes place in such a way that the hydrophobic parts are turned inward, toward each other, and away from water; hydrophilic parts—charged groups, for example—tend to stud the surface where they are near water. Hydrogen bonding is chiefly intramolecular. Areas of contact between molecules are small, and intermolecular forces are comparatively weak.

Molecular and intermolecular structure determines not only the solubility of a protein but also the general kind of function it performs.

Fibrous proteins serve as the chief structural materials of animal tissues, a function to which their insolubility and fiber-forming tendency suit them. They make up: *keratin*, in skin, hair, nails, wool, horn, and feathers; *collagen*, in tendons; *myosin*, in muscle; *fibroin*, in silk.

Globular proteins serve a variety of functions related to the maintenance and regulation of the life process, functions that require mobility and hence solubility. They make up: all enzymes; many hormones, as, for example, *insulin* (from the pancreas), *thyroglobulin* (from the thyroid gland), *ACTH* (from the pituitary gland); antibodies, responsible for allergies and for defense against foreign organisms; *albumin* in eggs; *hemoglobin*, which transports oxygen from the lungs to the tissues; *fibrinogen*, which is converted into the insoluble, fibrous protein *fibrin*, and thus causes the clotting of blood.

Within the two broad classes, proteins are subdivided on the basis of physical properties, especially solubility: for example, albumins (soluble in water, coagulated by heat), globulins (insoluble in water, soluble in dilute salt solutions), etc.

Irreversible precipitation of proteins, called **denaturation**, is caused by heat, strong acids or bases, or various other agents. Coagulation of egg white by heat, for example, is denaturation of the protein egg albumin. The extreme ease with which many proteins are denatured makes their study difficult. Denaturation causes a fundamental change in a protein, in particular destroying any physiological activity. (Denaturation appears to involve changes in the secondary structure of proteins, Sec. 36.16.)

Only one other class of compounds, the *nucleic acids* (Sec. 37.7), shows the phenomenon of denaturation. Although closely related to the proteins, polypeptides do not undergo denaturation, presumably because their molecules are smaller and less complex.

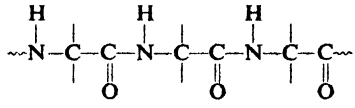
36.12 Structure of proteins

We can look at the structure of proteins on a number of levels. At the lowest level, there is the *primary* structure: the way in which the atoms of protein molecules are joined to one another by covalent bonds to form chains. Next, there is

the *secondary* structure: the way in which these chains are arranged in space to form coils, sheets, or compact spheroids, with hydrogen bonds holding together different chains or different parts of the same chain. Even higher levels of structure are gradually becoming understood: the weaving together of coiled chains to form ropes, for example, or the clumping together of individual molecules to form larger aggregates. Let us look first at the primary structure of proteins.

36.13 Peptide chain

Proteins are made up of peptide chains, that is, of amino acid residues joined by amide linkages. They differ from polypeptides in having higher molecular

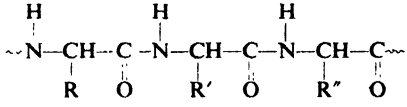


weights (by convention over 10,000) and more complex structures.

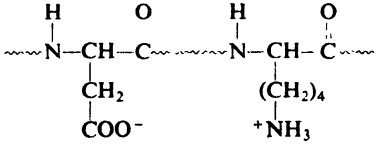
The peptide structure of proteins is indicated by many lines of evidence: hydrolysis of proteins by acids, bases, or enzymes yields peptides and finally amino acids; there are bands in their infrared spectra characteristic of the amide group; secondary structures based on the peptide linkage can be devised that exactly fit x-ray data.

36.14 Side chains. Isoelectric point. Electrophoresis

To every third atom of the peptide chain is attached a side chain. Its structure depends upon the particular amino acid residue involved: —H for glycine, $-CH_3$ for alanine, $-CH(CH_3)_2$ for valine, $-CH_2C_6H_5$ for phenylalanine, etc.



Some of these side chains contain basic groups: $-NH_2$ in lysine, or the imidazole ring in histidine. Some side chains contain acidic groups: —COOH in aspartic acid or glutamic acid. Because of these acidic and basic side chains, there are positively and negatively charged groups along the peptide chain. The behavior



of a protein in an electric field is determined by the relative numbers of these positive and negative charges, which in turn are affected by the acidity of the solution. At the isoelectric point, the positive and negative charges are exactly balanced and the protein shows no net migration; as with amino acids, solubility is usually at a minimum here. On the acid side of the isoelectric point, positive charges exceed

negative charges and the protein moves to the cathode; on the basic side of the isoelectric point, negative charges exceed positive charges and the protein moves to the anode.

While all proteins contain the peptide backbone, each protein has its own characteristic sequence of side chains, which gives it its characteristic properties. Different proteins have different proportions of acidic and basic side chains, and hence have different isoelectric points. In a solution of a particular hydrogen ion concentration, some proteins move toward a cathode and others toward an anode; depending upon the size of the charge as well as upon molecular size and shape, different proteins move at different speeds. This difference in behavior in an electric field is the basis of one method of separation and analysis of protein mixtures: **electrophoresis.**

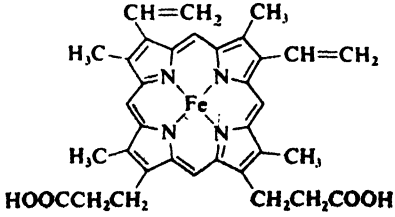
Side chains affect the properties of proteins not only by their acidity or basicity, but also by their other chemical properties and even by their sizes and shapes. Hydroxyl and sulfhydryl (—SH) groups can form esters; amino nitrogen is not only basic but nucleophilic. It seems likely that the "permanent" waving of hair depends upon changes in disulfide (—S—S—) cross-linkages provided by cysteine side chains; that much of the difference between silk and wool is related to the small side chains, —H and $-CH_3$, that predominate in silk fibroin; that the toughness of tendon is due to the flatness of the pyrrolidine ring and the ability of the —OH group of hydroxyproline to form hydrogen bonds. Replacement of *one* glutamic acid side chain in the hemoglobin molecule (300 side chains in all) by a valine unit is the cause of the fatal sickle-cell anemia.

The sequence of amino acids in hemoglobin has been used to study evolution, in the new science called *chemical paleogenetics.* In the *beta*-chain of hemoglobin, for example, the horse differs from man at 26 of the 146 sites; a pig, at 10 sites; and the gorilla at just *one* site. It has been estimated that, on the average, it takes roughly ten million years for one successful amino acid substitution to occur—that is, a substitution that improves the chances of survival. (Such a change is due to a change in the base sequence in a molecule of nucleic acid, Sec. 37.8.)

36.15 Conjugated proteins. Prosthetic groups. Coenzymes

Some protein molecules contain a non-peptide portion called a **prosthetic group**; such proteins are called *conjugated proteins.* The prosthetic group is intimately concerned with the specific biological action of the protein.

The prosthetic group of hemoglobin, for example, is *hemin.* As we see, hemin

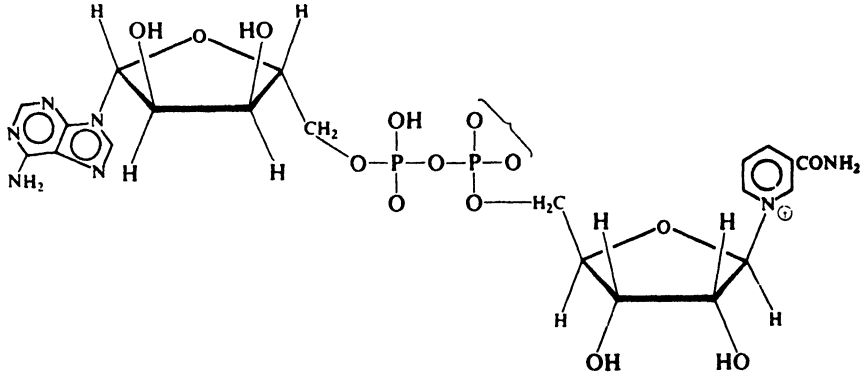


Hemin

contains iron bound to the pyrrole system known as *porphin* (compare with the structure of chlorophyll, p. 1004). It is the formation of a reversible oxygen–hemin complex that enables hemoglobin to carry oxygen from the lungs to the tissues. Carbon monoxide forms a similar, but more stable, complex; it thus ties up hemoglobin, prevents oxygen transport, and causes death. Hemin is separated from the peptide portion (*globin*) of the protein by mild hydrolysis; the two units are presumably held together by an amide linkage between a carboxyl group of hemin and an amino group of the polypeptide.

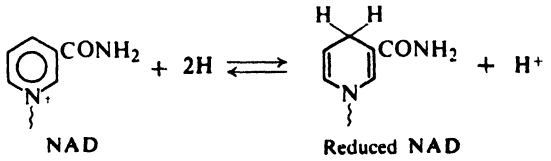
Many enzymes require *cofactors* if they are to exert their catalytic effects: metal ions, for example. Organic cofactors are called **coenzymes** and, if they are covalently bonded to the enzyme, these too are prosthetic groups.

The coenzyme *nicotinamide adenine dinucleotide* (NAD), for example, is associated with a number of dehydrogenation enzymes. This coenzyme we see, is



Nicotinamide adenine dinucleotide (NAD)
(Diphosphopyridinenucleotide)

made up of two molecules of D-ribose linked as phosphate esters, the fused heterocyclic system known as *adenine*, and nicotinamide in the form of a quaternary ammonium salt. In some systems one encounters *nicotinamide adenine dinucleotide phosphate* (NADP), in which the —OH on C-2 of the left-hand ribose unit of NAD has been phosphorylated. The characteristic biological function of these dehydrogenation enzymes (see, for example, Sec. 37.5) involves conversion of the nicotinamide portion of NAD or NADP into the dihydro structure.



NAD Reduced NAD

Like nicotinamide, many molecules making up coenzymes are **vitamins**, that is, substances that must be supplied in the diet to permit proper growth or maintenance of structure. Undoubtedly it is for their coenzyme activity that these substances are needed.

36.16 Secondary structure of proteins

It seems clear that proteins are made up of polypeptide chains. How are these chains arranged in space and in relationship to each other? Are they stretched out side by side, looped and coiled about one another, or folded into independent spheroids?

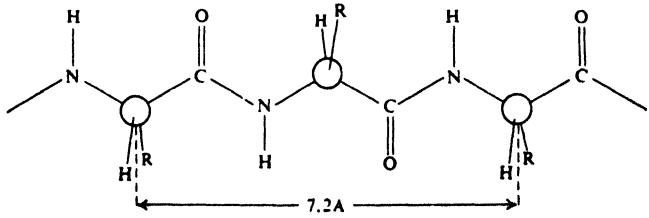
Much of our understanding of the secondary structure of proteins is the result of x-ray analysis. For many proteins the x-ray diffraction pattern indicates a regular repetition of certain structural units. For example, there are *repeat distances* of 7.0 A in silk fibroin, and of 1.5 A and 5.1 A in α-keratin of unstretched wool.

The problem is to devise structures that account for the characteristic x-ray diffraction patterns, and are at the same time consistent with what is known about the primary structure: bond lengths and bond angles, planarity of the amide group, similarity of configuration about chiral centers (all L-family), size and sequence of side chains. Of key importance in this problem has been recognition of the stabilizing effect of hydrogen bonds (5–10 kcal per mole per hydrogen bond), and the principle that the most stable structure is one that permits formation of the maximum number of hydrogen bonds. On the basis of the study of simpler compounds, it has been further assumed that the N—H···O bond is very nearly linear, hydrogen lying on, or within 20° of, the line between nitrogen and oxygen. In all this work the simultaneous study of simpler, synthetic polypeptides containing only a single kind of amino acid residue has been of great help.

The progress made on a problem of this size and difficulty has necessarily been the work of many people. Among them is Linus Pauling, of the California Institute of Technology, who received the Nobel Prize in 1954. In 1951 Pauling wrote: "Fourteen years ago Professor Robert B. Corey and I, after we had made a vigorous but unsuccessful attack on the problem of formulating satisfactory configurations of polypeptide chains in proteins, decided to attempt to solve the problem by an indirect method—the method of investigating with great thoroughness crystals of amino acids, simple peptides, and related substances, in order to obtain completely reliable and detailed information about the structural characteristics of substances of this sort, and ultimately to permit the confident prediction of precisely described configurations of polypeptide chains in proteins." (Record Chem. Prog., *12*, 156–7 (1951).). This work on simple substances, carried on for more than 14 years, gave information about the geometry of the amide group that eventually led Pauling and his co-workers to propose what may well be the most important secondary structure in protein chemistry: the α-helix.

Let us look at some of the secondary structures that have been proposed.

As a point of departure, it is convenient to consider a structure (perhaps hypothetical) in which peptide chains are fully extended to form flat zig-zags:



Extended peptide chain

These chains lie side by side to form a *flat sheet*. Each chain is held by hydrogen bonds to the two neighboring chains (Fig. 36.2). This structure has a repeat distance of 7.2 A, the distance between *alternate* amino acid residues. (Notice that alternate side chains lie on the same side of the sheet.) However, crowding between side chains makes this idealized flat structure impossible, except perhaps for synthetic polyglycine.

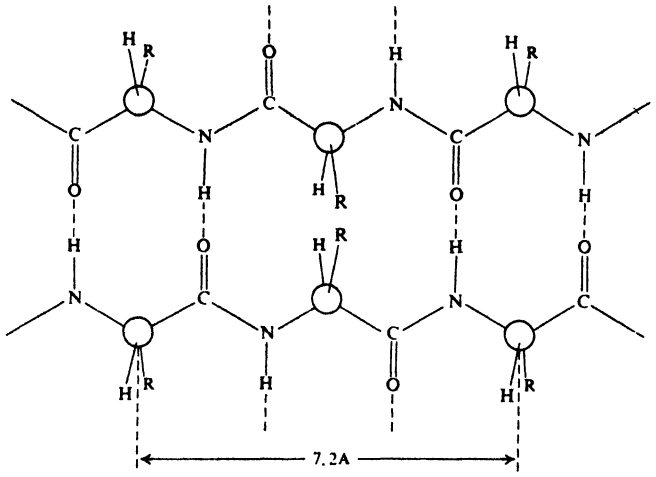
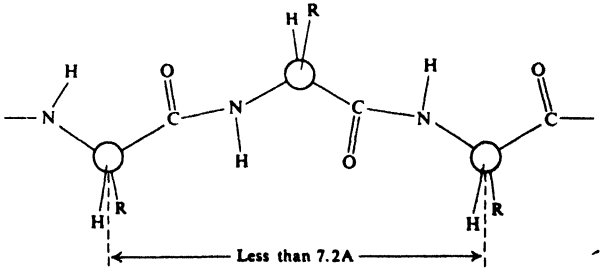


Figure 36.2. Hypothetical flat sheet structure for a protein. Chains fully extended; adjacent chains head in opposite directions; hydrogen bonding between adjacent chains. Side chains (R) are crowded.

Room can be made for small or medium-sized side chains by a slight contraction of the peptide chains:



Contracted peptide chain

The chains still lie side by side, held to each other by hydrogen bonds. The contraction results in a *pleated sheet*, with a somewhat shorter distance between alternate amino acid residues (see Fig. 36.3). Such a structure, called the **beta**

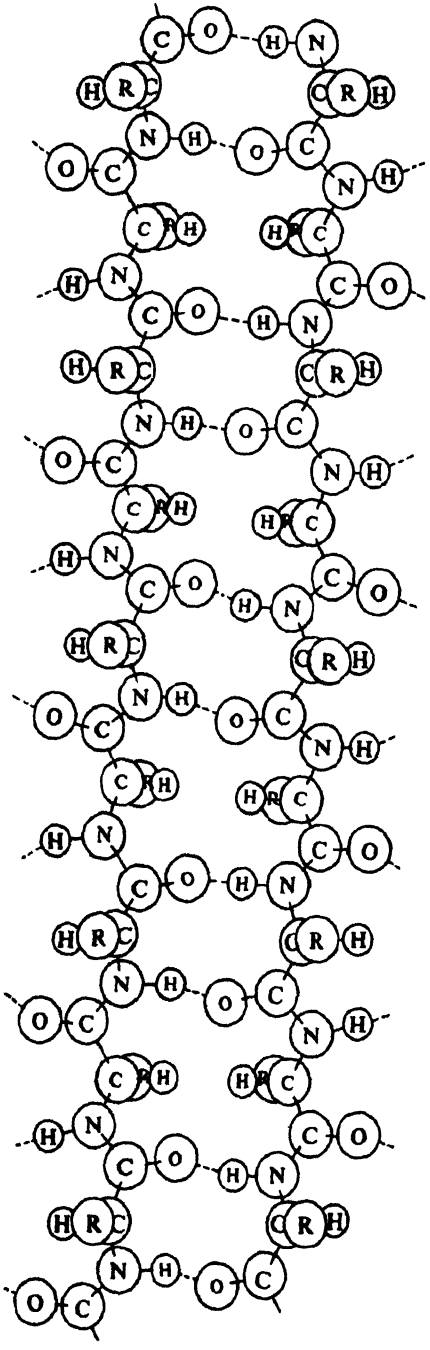
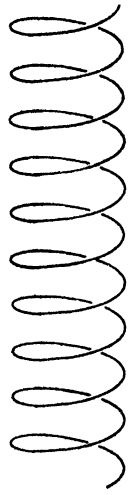


Figure 36.3. Pleated sheet structure (*beta arrangement*) proposed by Pauling for silk fibroin. Chains contracted to make room for small side chains. Adjacent chains head in opposite directions; hydrogen bonding between adjacent chains.

arrangement, has been proposed for silk fibroin, which has a repeat distance of 7.0 A and most closely approaches the fully extended, flat-sheet structure. It is significant that, although 15 kinds of amino acid residue are found in silk fibroin, 46% of the residues are glycine, which has no side chain, and another 38% are alanine and serine with the small side chains $-CH_3$ and $-CH_2OH$.

When the side chains are quite large, they are best accommodated by a quite different kind of structure. Each chain is coiled to form a *helix* (like a spiral



A helix
(right-handed)

staircase). Hydrogen bonding occurs between different parts of the *same* chain, and holds the helix together. For α-keratin (unstretched wool, hair, horn, nails) Pauling has proposed a helix in which there are 3.6 amino acid residues per turn (Fig. 36.4). Models show that this 3.6-helix provides room for the side chains and allows all possible hydrogen bonds to form. It accounts for the repeat distance of 1.5 A, which is the distance between amino acid residues measured along the axis of the helix. To fit into this helix, all the amino acid residues must be of the same configuration, as, of course, they are; furthermore, their L-configuration requires the helix to be *right-handed*, as shown. It is becoming increasingly clear that the **alpha helix**, as it is called, is of fundamental importance in the chemistry of proteins.

(To account for the second repeat distance of 5.1 A for α-keratin, we must go to what is properly the *tertiary structure*. Pauling has suggested that each helix can itself be coiled into a superhelix which has one turn for every 35 turns of the *alpha* helix. Six of these superhelixes are woven about a seventh, straight helix to form a seven-strand cable.)

When wool is stretched, α-keratin is converted into β-keratin, with a change in the x-ray diffraction pattern. It is believed that the helixes are uncoiled and the chains stretched side by side to give a sheet structure of the *beta* type. The hydrogen

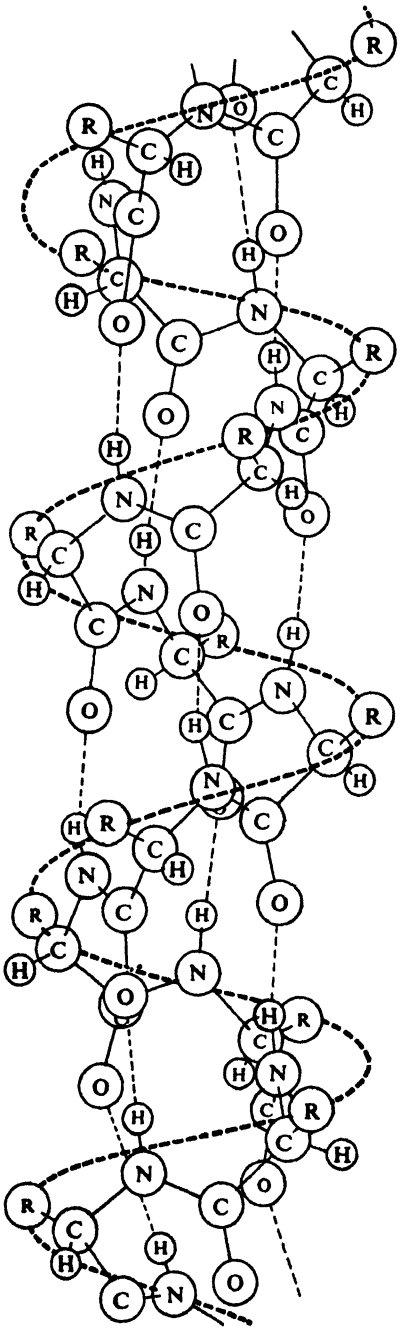
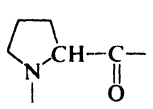


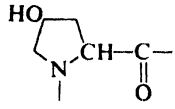
Figure 36.4. Alpha helix structure proposed by Pauling for α-keratin. Makes room for large side chains. Right-handed helix with 3.6 residues per turn; hydrogen bonding within a chain.

bonds within the helical chain are broken, and are replaced by hydrogen bonds between adjacent chains. Because of the larger side chains, the peptide chains are less extended (repeat distance 6.4 A) than in silk fibroin (repeat distance 7.0 A).

Besides the x-ray diffraction patterns characteristic of the *alpha-* and *beta-*type proteins, there is a third kind: that of *collagen*, the protein of tendon and skin. On the primary level, collagen is characterized by a high proportion of proline and hydroxyproline residues, and by frequent repetitions of the sequence Gly.Pro.Hypro. The pyrrolidine ring of proline and hydroxyproline can affect



Proline residue



Hydroxyproline residue

the secondary structure in several ways. The amido nitrogen carries no hydrogen for hydrogen bonding. The flatness of the five-membered ring, in conjunction with the flatness of the amide group, prevents extension of the peptide chain as in the *beta* arrangement, and interferes with the compact coiling of the *alpha* helix.

The structure of collagen combines the helical nature of the *alpha*-type proteins with the inter-chain hydrogen bonding of the *beta*-type proteins. Three peptide chains—each in the form of a left-handed helix—are twisted about one another to form a three-strand right-handed superhelix. A small glycine residue at every third position of each chain makes room for the bulky pyrrolidine rings on the other two chains. The three chains are held strongly to each other by hydrogen bonding between glycine residues and between the —OH groups of hydroxyproline.

When collagen is boiled with water, it is converted into the familiar water-soluble protein *gelatin*; when cooled, the solution does not revert to collagen but sets to a gel. Gelatin has a molecular weight one-third that of collagen. Evidently the treatment separates the strands of the helix, breaking inter-chain hydrogen bonds and replacing them with hydrogen bonds to water molecules.

Turning from the insoluble, fibrous proteins to the soluble, globular proteins (e.g., hemoglobin, insulin, *gamma*-globulin, egg albumin), we find that the matter of secondary structure can be even more complex. Evidence is accumulating that here, too, the *alpha* helix often plays a key role. These long peptide chains are not uniform: certain segments may be coiled into helixes or folded into sheets; other segments are looped and coiled into complicated, irregular arrangements. Look, for example, at α-chymotrypsin in Fig. 37.1 (p. 1166).

This looping and coiling appears to be random, but it definitely is *not*. The sequence of amino acids is determined genetically (Sec. 37.7) but, once formed, the chain *naturally* falls into the arrangement that is *most stable* for that particular sequence.

We find all our kinds of "intermolecular" forces at work here—but acting between different parts of the same molecule: van der Waals forces, hydrogen bonds, interionic attraction (or repulsion) between charged groups. There is chemical cross-linking by disulfide bonds. The characteristic feature of these

globular proteins is that hydrophobic parts are turned inward, toward each other and away from water—like the hydrophobic tails in a soap micelle.

In their physiological functions, proteins are highly specific. We have encountered, for example, an enzyme that will cleave α-glucosides but not β-glucosides, and an enzyme that will cleave only C-terminal amino acid residues in polypeptides. It seems clear that the biological activity of a protein depends not only upon its prosthetic group (if any) and its particular amino acid sequence, but also upon its molecular shape. As Emil Fischer said in 1894: "... enzyme and glucoside must fit together like a lock and key. ..." In Sec. 37.2 we shall see how one enzyme is believed to exert its effect, and how that effect depends, in a very definite and specific way, on the shape of the enzyme molecule.

Denaturation uncoils the protein, destroys the characteristic shape, and with it the characteristic biological activity.

In 1962, M. F. Perutz and J. C. Kendrew of Cambridge University were awarded the Nobel Prize in chemistry for the elucidation of the structure of hemoglobin and the closely related oxygen-storing molecule, myoglobin. Using x-ray analysis, and knowing the amino acid sequence (p. 1146), they determined the shape—in three dimensions—of these enormously complicated molecules: precisely for myoglobin, and very nearly so for hemoglobin. They can say, for example, that the molecule is coiled in an *alpha* helix for sixteen residues from the N-terminal unit, and then turns through a right angle. They can even say *why*: at the corner there is an aspartic acid residue; its carboxyl group interferes with the hydrogen bonding required to continue the helix, and the chain changes its course. The four folded chains of hemoglobin fit together to make a spheroidal molecule, 64 A × 55 A × 50 A. Four flat heme groups, each of which contains an iron atom that can bind an oxygen molecule, fit into separate pockets in this sphere. When oxygen is being carried, the chains move to make the pockets slightly smaller; Perutz has described hemoglobin as "a breathing molecule." These pockets are lined with the hydrocarbon portions of the amino acids; such a non-polar environment prevents electron transfer between oxygen and ferrous iron, and permits the complexing necessary for oxygen transport.

PROBLEMS

1. Outline all steps in the synthesis of phenylalanine from toluene and any needed aliphatic and inorganic reagents by each of the following methods:

(a) direct ammonolysis
(b) Gabriel synthesis
(c) malonic ester synthesis
(d) phthalimidomalonic ester method
(e) Strecker synthesis
(f) reductive amination

2. (a) Give structures of all intermediates in the following synthesis of proline:

potassium phthalimide + bromomalonic ester $\longrightarrow$ A

A + $Br(CH_2)_3Br \xrightarrow{NaOC_2H_5}$ B ($C_{18}H_{20}O_6NBr$)

B + potassium acetate $\longrightarrow$ C ($C_{20}H_{23}O_8N$)

C + $NaOH$, heat; then H^+, heat $\longrightarrow$ D ($C_5H_{11}O_3N$)

D + HCl $\longrightarrow$ [E ($C_5H_{10}O_2NCl$)] $\longrightarrow$ proline

(b) Outline a possible synthesis of lysine by the phthalimidomalonic ester method.

3. Give structures of all intermediates in the following syntheses of amino acids:

(a) ethyl acetamidomalonate [$CH_3CONHCH(COOC_2H_5)_2$] + acrolein $\xrightarrow{\text{Michael}}$ F ($C_{12}H_{19}O_6N$)

F + KCN + acetic acid $\longrightarrow$ G ($C_{13}H_{20}O_6N_2$)
G + acid + heat $\longrightarrow$ H ($C_{13}H_{18}O_5N_2$)
H + H_2, catalyst, in acetic anhydride $\longrightarrow$ [I ($C_{13}H_{24}O_5N_2$)]
I $\xrightarrow{\text{acetic anhydride}}$ J ($C_{15}H_{26}O_6N_2$)
J + OH^-, heat; then H^+; then heat $\longrightarrow$ (±)-lysine

(b) acrylonitrile + ethyl malonate $\xrightarrow{\text{Michael}}$ K ($C_{10}H_{15}O_4N$)
K + H_2, catalyst $\longrightarrow$ [L ($C_{10}H_{19}O_4N$)] $\longrightarrow$ M ($C_8H_{13}O_3N$)
M + SO_2Cl_2 in $CHCl_3$ $\longrightarrow$ N ($C_8H_{12}O_3NCl$)
N + HCl, heat $\longrightarrow$ O ($C_5H_{10}O_2NCl$)
O $\xrightarrow{\text{base}}$ (±)-proline

(c) Glutamic acid has been made from acrolein via a Strecker synthesis. Show how this might have been done. (*Hint:* See Sec. 27.5.)

4. Using the behavior of hydroxy acids (Sec. 20.15) as a pattern, predict structures for the products obtained when the following amino acids are heated:

(a) the α-amino acid, glycine $\longrightarrow$ $C_4H_6O_2N_2$ (*diketopiperazine*)
(b) the β-amino acid, $CH_3CH(NH_2)CH_2COOH$ $\longrightarrow$ $C_4H_6O_2$
(c) the γ-amino acid, $CH_3CH(NH_2)CH_2CH_2COOH$ $\longrightarrow$ C_5H_9ON (*a lactam*)
(d) the δ-amino acid, $H_2NCH_2CH_2CH_2CH_2COOH$ $\longrightarrow$ C_5H_9ON (*a lactam*)

5. (a) Show how the particular dipolar structure given for histidine in Table 36.1 is related to the answer to Problem 20 (b), p. 1026.

(b) Draw the two possible dipolar structures for lysine. Justify the choice of structure given in Table 36.1. (c) Answer (b) for aspartic acid. (d) Answer (b) for arginine. (*Hint:* See Problem 20.24, p. 686.) (e) Answer (b) for tyrosine.

6. (a) *Betaine*, $C_5H_{11}O_2N$, occurs in beet sugar molasses. It is a water-soluble solid that melts with decomposition at 300°. It is unaffected by base but reacts with hydrochloric acid to form a crystalline product, $C_5H_{12}O_2NCl$. It can be made in either of two ways: treatment of glycine with methyl iodide, or treatment of chloroacetic acid with trimethylamine.

Draw a structure for betaine that accounts for its properties.

(b) *Trigonelline*, $C_7H_7O_2N$, is an alkaloid found in coffee beans; it is also excreted from the body as a metabolic product of nicotinic acid. It is insoluble in benzene or ether, and dissolves in water to give a neutral solution. It is unaffected by boiling aqueous acid or base. It has been synthesized as follows:

nicotinic acid + CH_3I + KOH $\longrightarrow$ P ($C_8H_{10}O_2NI$)
P + Ag_2O + H_2O, warm $\longrightarrow$ trigonelline + AgI + CH_3OH

What structure for trigonelline is consistent with these properties?

7. Addition of ethanol or other organic solvents to an aqueous "solution" of a globular protein brings about denaturation. Such treatment also tends to break up micelles of, say, soap (Sec. 33.3). What basic process is at work in both cases?

8. An amino group can be protected by acylation with phthalic anhydride to form an N-substituted phthalimide. The protecting group can be removed by treatment with hydrazine, $H_2N—NH_2$ without disturbing any peptide linkages. Write equations to show how this procedure (exploited by John C. Sheehan of the Massachusetts Institute of Technology) could be applied to the synthesis of glycylalanine (Gly.Ala) and alanylglycine (Ala.Gly).

9. An elemental analysis of *Cytochrome c*, an enzyme involved in oxidation–reduction processes, gave 0.43% Fe and 1.48% S. What is the minimum molecular weight of the enzyme? What is the minimum number of iron atoms per molecule? Of sulfur atoms?

10. A protein, *β-lactoglobulin*, from cheese whey, has a molecular weight of 42020 ± 105. When a 100-mg sample was hydrolyzed by acid and the mixture was made alkaline, 1.31 mg of ammonia was evolved. (a) Where did the ammonia come from, and approximately how many such groups are there in the protein?

Complete hydrolysis of a 100-mg sample of the protein used up approximately 17 mg of water. (b) How many amide linkages per molecule were cleaved?

(c) Combining the results of (a) and (b), and adding the fact that there are four N-terminal groups (four peptide chains in the molecule), how many amino acid residues are there in the protein?

11. The complete structure of *Gramicidin S*, a polypeptide with antibiotic properties, has been worked out as follows:

(a) Analysis of the hydrolysis products gave an empirical formula of Leu,Orn.Phe,Pro,Val. (*Ornithine*, Orn, is a rare amino acid of formula $^{+}H_3NCH_2CH_2CH_2CH(NH_2)COO^{-}$.) It is interesting that the phenylalanine has the unusual D-configuration.

Measurement of the molecular weight gave an approximate value of 1300. On this basis, what is the molecular formula of Gramicidin S?

(b) Analysis for the C-terminal residue was negative; analysis for the N-terminal residue using DNFB yielded only $DNP{-}NHCH_2CH_2CH_2CH(\overset{+}{N}H_3)COO^{-}$. What structural feature must the peptide chain possess?

(c) Partial hydrolysis of Gramicidin S gave the following di- and tripeptides:

Leu.Phe	Phe.Pro	Phe.Pro.Val	Val.Orn.Leu
Orn.Leu	Val.Orn	Pro.Val.Orn.	

What is the structure of Gramicidin S?

12. The structure of beef insulin was determined by Sanger (see Sec. 36.9) on the basis of the following information. Work out for yourself the sequence of amino acid residues in the protein.

Beef insulin appears to have a molecular weight of about 6000 and to consist of two polypeptide chains linked by disulfide bridges of cystine residues. The chains can be separated by oxidation, which changes any CyS—SCy or CySH residues to sulfonic acids ($CySO_3H$).

One chain, A, of 21 amino acid residues, is acidic and has the empirical formula

$$GlyAlaVal_2Leu_2Ileu(CySH)_4Asp_2Glu_4Ser_2Tyr_2$$

The other chain, B, of 30 amino acid residues, is basic and has the empirical formula

$$Gly_3Ala_2Val_3Leu_4ProPhe_3(CySH)_2ArgHis_2LysAspGlu_3SerThrTyr_2$$

(Chain A has four simple side-chain amide groups, and chain B has two, but these will be ignored for the time being.)

Treatment of chain B with 2,4-dinitrofluorobenzene (DNFB) followed by hydrolysis gave DNP.Phe and DNP.Phe.Val; chain B lost alanine (Ala) when treated with carboxypeptidase.

Acidic hydrolysis of chain B gave the following tripeptides:

Glu.His.Leu	Leu.Val.CySH	Tyr.Leu.Val
Gly.Glu.Arg	Leu.Val.Glu	Val.Asp.Glu
His.Leu.CySH	Phe.Val.Asp	Val.CySH.Gly
Leu.CySH.Gly	Pro.Lys.Ala	Val.Glu.Ala
	Ser.His.Leu	

Many dipeptides were isolated and identified; two important ones were Arg.Gly and Thr.Pro.

(a) At this point construct as much of the B chain as the data will allow.

Among the numerous tetrapeptides and pentapeptides from chain B were found:

His.Leu.Val.Glu	Tyr.Leu.Val.CySH
Ser.His.Leu.Val	Phe.Val.Asp.Glu.His

(b) How much more of the chain can you reconstruct now? What amino acid residues are still missing?

Enzymatic hydrolysis of chain B gave the necessary final pieces:

Val.Glu.Ala.Leu	His.Leu.CySH.Gly.Ser.His.Leu
Tyr.Thr.Pro.Lys.Ala	Tyr.Leu.Val.CySH.Gly.Glu.Arg.Gly.Phe.Phe

(c) What is the complete sequence in the B chain of beef insulin?

Treatment of chain A with DNFB followed by hydrolysis gave DNP.Gly; the C-terminal group was shown to be aspartic acid (Asp).

Acidic hydrolysis of chain A gave the following tripeptides:

CySH.CySH.Ala	Glu.Leu.Glu
Glu.Asp.Tyr	Leu.Tyr.Glu
Glu.CySH.CySH	Ser.Leu.Tyr
Glu.Glu.CySH	Ser.Val.CySH

Among other peptides isolated from acidic hydrolysis of chain A were:

CySH.Asp Tyr.CySH Gly.Ileu.Val.Glu.Glu

(d) Construct as much of chain A as the data will allow. Are there any amino acid residues missing?

Up to this point it is possible to arrive at the sequences of four parts of chain A, but it is still uncertain which of the two center fragments, Ser.Val.CySH or Ser.Leu.Tyr, etc., comes first. This was settled by digestion of chain A with pepsin, which gave a peptide that contained no aspartic acid (Asp) or tyrosine (Tyr). Hydrolysis of this peptide gave Ser.Val.CySH and Ser.Leu.

(e) Now what is the complete structure of chain A of beef insulin?

In insulin the cysteine units (CySH) are involved in cystine disulfide links (CyS –SCy). Residue 7 of chain A (numbering from the N-terminal residue) is linked to residue 7 of chain B, residue 20 of chain A to residue 19 of chain B, and there is a link between residues 6 and 11 of chain A.

There are amide groups on residues 5, 15, 18, and 21 of chain A, and on residues 3 and 4 of chain B.

(f) Draw a structure of the complete insulin molecule. (*Note:* The disulfide loop in chain A is a 20-atom, pentapeptide ring, of the same size as the one in oxytocin.)

In the analysis for the N-terminal group in chain B of insulin, equal amounts of *two* different DNP derivatives of single amino acids actually were found. One was DNP.Phe; what could the other have been?

(g) What would have been obtained if that second amino acid had been N-terminal?

Chapter 37

Biochemical Processes

Molecular Biology

37.1 Biochemistry, molecular biology, and organic chemistry

In the past four chapters, we have learned something about fats, carbohydrates, and proteins: their structures and how these are determined, and the kind of reactions they undergo in the test tube. These, we said, are biomolecules: they are participants in the chemical process we call life. But just what do they *do*? What reactions do they undergo, not in the test tube, but in a living organism?

Even a vastly simplified answer to that question would fill—and does—a book as big as this one. Having come this far, though, we cannot help being curious. And so, in this chapter, we shall take a brief glance at the answer—or, rather, at the kind of thing the answer entails.

We shall look at just a few examples of biochemical processes: how one enzyme—of the thousands in our bodies—may work; what happens in one of the dozens of reactions by which carbohydrates are oxidized to furnish energy; how one kind of chemical compound—fatty acids—is synthesized. Finally, we shall learn a little about another class of biomolecules, the nucleic acids, and how they are involved in the most fascinating biochemical process of all—heredity.

The study of nucleic acids has become known as "molecular biology." Actually, of course, all of these processes are a part of molecular biology—biology on the molecular level—and they are, in the final analysis, organic chemistry. And it is as organic chemistry that we shall treat them. We shall see how all these vital processes—even the mysterious powers of enzymes—come down to a matter of molecular structure as we know it: to molecular size and shape; to intermolecular and intramolecular forces; to the chemistry of functional groups; to acidity and basicity, oxidation and reduction; to energy changes and rate of reaction.

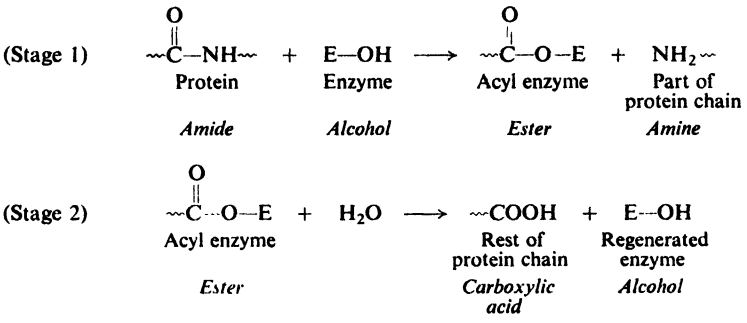
Since catalysis by enzymes is fundamental to everything else, let us begin there.

37.2 Mechanism of enzyme action. Chymotrypsin

Enzymes, we have said, are proteins that act as enormously effective catalysts for biological reactions. To get some idea of how they work, let us examine the action of just one: *chymotrypsin*, a digestive enzyme whose job is to promote hydrolysis of certain peptide links in proteins. The sequence of the 245 amino acid residues in chymotrypsin has been determined and, through x-ray analysis, the conformation of the molecule is known (Fig. 37.1). It is, like all enzymes, a soluble globular protein coiled in the way that turns its hydrophobic parts inward, away from water, and that permits maximum intramolecular hydrogen bonding.

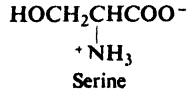
The action of chymotrypsin has been more widely explored than that of any other enzyme. In crystalline form, it is available for studies in the test tube under a variety of conditions. It catalyzes hydrolysis not only of proteins but of ordinary amides and esters, and much has been learned by use of these simpler substrates. Compounds modeled after portions of the chymotrypsin molecule have been made, and their catalytic effects measured.

To begin with, it seems very likely that chymotrypsin acts in two stages. In the first stage, acting as an alcohol, it breaks the peptide chain. We recognize this as alcoholysis of a substituted amide: nucleophilic acyl substitution. The products are an amine—the liberated portion of the substrate molecule—and, as we shall



see shortly, an ester of the enzyme. In the second stage, the enzyme ester is hydrolyzed. This yields a carboxylic acid—the other portion of the substrate molecule—and the regenerated enzyme, ready to go to work again.

What is the structure of this intermediate ester formed from the enzyme? The answer has been found by use of simple esters as substrates, *p*-nitrophenyl acetate, for example. An appreciable steady-state concentration of the intermediate ester builds up and, by quenching of the reaction mixture in acid, it can be isolated. Sequence analysis of the enzyme ester showed that the acetyl group from the substrate was linked to *serine-195*. It is, then, at the —OH group of this particular amino acid residue that the enzyme reacts.



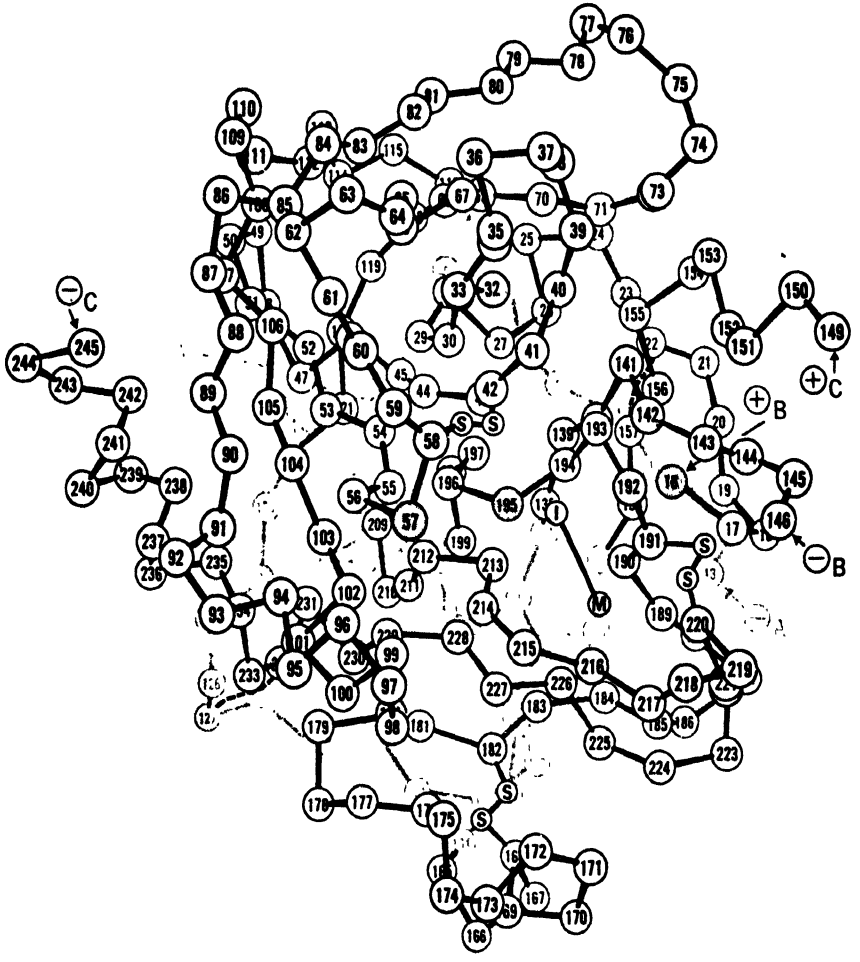


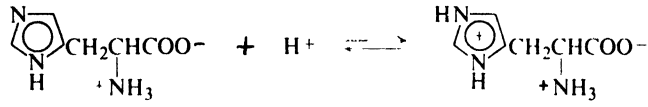
Figure 37.1. Three-dimensional structure of α-chymotrypsin. Histidine-57, serine-195, and isoleucine-16 are shaded. The hydrophobic pocket lies to the right of histidine-57 and serine-196, where M is marked; it is bounded by residues 184–191 and 214–227.

The ⊕ and ⊖ signs show the N-terminal and C-terminal ends of chains A, B, and C. The M and I stand for the methyl and sulfonyl parts of the inhibitor, a tosyl group held as an ester of serine-195.

We can see one short segment of α-helix at residues 234–245; another (mostly hidden) lies at 164–170. There is a hint of a twisted sheet beginning with residues 91–86 and 103–108, and extending to their right.

But evidence shows that certain other amino acid residues are also vital to enzyme activity. The rate of enzyme-catalyzed hydrolysis changes as the acidity of the reaction medium is changed. If one plots the rate of hydrolysis against the pH of the solution, one gets a bell-shaped curve: as the pH is increased, the rate rises to a maximum and then falls off. The rate is fastest at about pH 7.4 (fittingly, the physiological pH) and slower in either more acidic or more basic solution. Analysis of the data shows the following. Hydrolysis requires the presence of a free base, of K_b about 10^{-7}, and a protonated base, of K_b about 3×10^{-5}. At low pH (acid solution), both bases are protonated; at high pH (alkaline solution), both bases are free. Hydrolysis is fastest at the intermediate pH where the weaker base is mostly free and the stronger base is mostly protonated.

The K_b of the weaker base fits that of the imidazole ring of histidine, and there is additional evidence indicating that this is indeed the base: studies involving

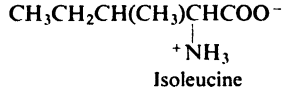


Histidine
Base

Protonated histidine
Acid

catalysis by imidazole itself, for example. Now, examination of the conformation of chymotrypsin (Fig 37.1) shows that very close to serine-195 there *is* a histidine residue. This is *histidine-57*, and it is believed to be the one involved in enzyme activity.

What about the stronger base which, according to the kinetics, is involved in its protonated form? Its K_b fits the α-amino group of most amino acids—an α-amino group, that is, which is not tied up in a peptide link. But all the (free) amino groups in chymotrypsin—*except one*—may be acetylated without complete loss of activity. The exception is *isoleucine-16*, the N-terminal unit of chain B.



Isoleucine

Presumably, then, this amino group cannot be acetylated, but must be free to be protonated and do its part of the job.

Now, what is the job of each of these key units in the enzyme molecule? It is clear what serine-195 does: it provides the —OH for ester formation. What does isoleucine-16 do? The descending leg of the bell-shaped rate curve was attributed to protonation of this unit. But something else happens as the pH is raised above 7.4: the optical activity of the solution decreases—evidently due to a change in conformation of the enzyme molecule—and in a way that parallels the decrease in rate of hydrolysis. It is believed, then, that, through hydrogen bonding or electrostatic attraction, the $-NH_3^+$ on isoleucine-16 helps hold the enzyme chain in the proper shape for it to act as a catalyst: to keep histidine-57 near serine-195, among other things. At higher pH the $-NH_3^+$ is converted into $-NH_2$, and the chain changes its shape; with the change in shape goes loss of catalytic power and a change in optical rotation.

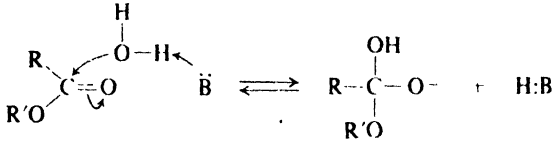
Next, we come to the question: what is the role of histidine-57? We are observing an example of *general acid–base catalysis:* catalysis not just by hydroxide ions and oxonium ions, but by all the bases and conjugate acids that are present, each contributing according to its concentration and its acid or base strength.

Let us look at this concept first with a simple example: hydrolysis of an ester catalyzed by the simple heterocyclic base, imidazole. Catalysis by hydroxide ions

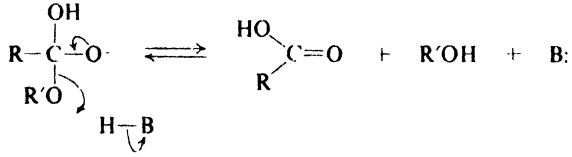
$$RCOOR' + H_2O \xrightarrow{\text{imidazole}} RCOOH + R'OH$$

we understand: these highly nucleophilic ions are more effective than water at attacking acyl carbon. Imidazole generates some hydroxide ions by reaction with water, but these are already taken into account. We are talking now about hydrolysis that is directly proportional to the concentration of the base itself: imidazole. What seems to be involved in such reactions is something like the following. In step (1), water adds to acyl carbon with *simultaneous loss of a proton to the base*;

(1)



(2)



reaction is fast because, in effect, the attacking nucleophile is not just water, but an incipient hydroxide ion. In step (2), transfer of the proton from the protonated base is simultaneous with loss of the ethoxy group; again reaction is fast, this time because the leaving group is not the strongly basic ethoxide ion, but an incipient alcohol molecule.

Reactions like (1) and (2) need not involve unlikely three-body collisions among the reactive molecules. Instead, there is prior hydrogen bonding between the base and water or between the protonated base and ester; it is these double molecules that collide with the third reagent and undergo reaction, with the dipole–dipole attraction of the hydrogen bonding being replaced by a covalent bond.

Figure 37.2 depicts the action of chymotrypsin, with the imidazole group of histidine-57 playing the same role of general base as that just described—and with protonated imidazole necessarily acting as general acid. There is general acid–base catalysis of both reactions involved: first, in the formation of the acyl enzyme, and then in its hydrolysis.

Chymotrypsin is not, as enzymes go, very specific in its action; it hydrolyzes proteins, peptides, simple amides, and esters alike. There is one structural requirement, nevertheless; a relatively non-polar group in the acyl moiety of the substrate,

typically an aromatic ring. Now, turning once more to Fig. 37.1, we find that at the reactive site in the enzyme there is a pocket; this pocket is lined with hydrophobic substituents to receive the non-polar group of the substrate and thus hold the molecule in position for hydrolysis. It is the size of this pocket and the nature of its lining that gives the enzyme its specificity; here we find, in a very real sense, Emil Fischer's lock into which the substrate key must fit.

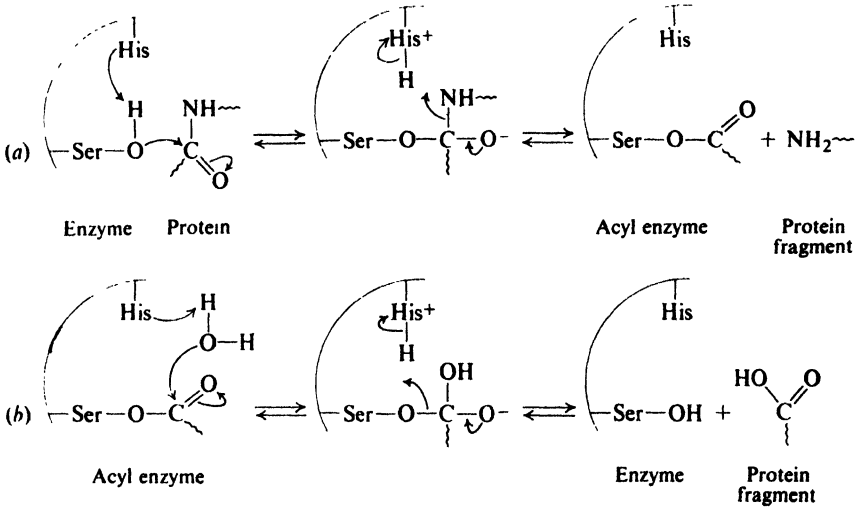


Figure 37.2. Catalysis by the enzyme chymotrypsin of the cleavage of one peptide bond in a protein: a proposed mechanism. Histidine and protonated histidine act as general base and acid in two successive nucleophilic substitution reactions: (*a*) cleavage of protein with formation of acyl enzyme and liberation of one protein fragment; (*b*) hydrolysis of acyl enzyme with regeneration of the enzyme and liberation of the other protein fragment.

We see, then, some of the factors that give enzymes their catalytic powers. The substrate is bound to a particular site in the enzyme, where the necessary functional groups are gathered: here, hydroxyl of serine and imidazole of histidine. In most cases, there are other functional groups as well, in molecules of cofactors—reagents, really—bound by the enzyme near the reactive site. In the enzyme–substrate complex, these functional groups are *neighboring groups*, and in their reactions enjoy all the advantages we listed (Sec. 28.1) for such groups. *They are there*, poised in just the right position for attack on the substrate. They need not wait for the lucky accident of a molecular collision; in effect, concentration of reagents is very high. Orientation of reacting groups is exactly right. There are no clinging solvent molecules to be stripped away as reaction occurs.

All these factors are important, and can be shown independently to speed up the rate of reactions, and very powerfully, too—but they do not seem to be nearly enough to account for the enormous activity of enzymes. Perhaps other factors are involved. It has been suggested, for example, that the pocket in which reaction

occurs fits the transition state better than it fits the reactants, so that relief of strain or an increase in van der Waals attractions provides a driving force. Perhaps the correct factors are being considered but, in extrapolation to enzyme systems, their power has been underestimated.

37.3 The source of biological energy. The role of ATP

In petroleum we have a fuel reserve on which we can draw for energy—as long as it lasts. We burn it, and either use the heat produced directly to warm ourselves or convert it into other kinds of energy: mechanical energy to move things about; electrical energy, which is itself transformed—at a more convenient place than where the original burning happened—into light, or mechanical energy, or back into heat.

In the same way, the energy our bodies need to keep warm, move about, and build new tissue comes from a food reserve: carbohydrates, chiefly in the form of starch. (We eat other animals, too, but ultimately the chain goes back to a carbohydrate-eater.) In the final analysis, we get energy from food just as we do from petroleum: we oxidize it to carbon dioxide and water.

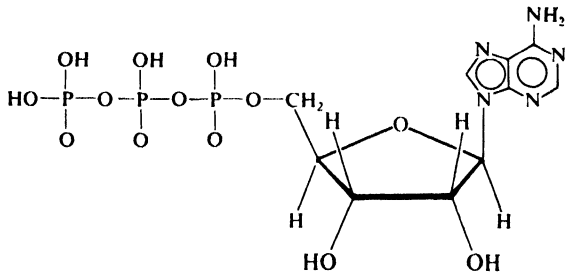
This food reserve is not, however, a limited one that we steadily deplete. Our store of carbohydrates—and the oxygen to go with it—is constantly replenished by the recombining, in plants, of carbon dioxide and water. The energy for recombination comes, of course, from the sun.

We speak of both petroleum and carbohydrates as sources of energy; we could speak of them as "energy-rich molecules." But the oxygen that is also consumed in oxidation is equally a source of energy. What we really mean is that the energy content of carbohydrates (or petroleum) plus oxygen is greater than that of carbon dioxide plus water. (In total, the bonds that are to be broken are weaker—contain more energy—than the bonds that are to be formed.) These reactants are, of course, energy-rich only in relation to the particular products we want to convert them into. But this is quite sufficient; in our particular kind of world, these *are* our sources of energy.

The body takes in carbohydrates and oxygen, then, and eventually gives off carbon dioxide and water. In the process considerable energy is generated. But in what form? And how is it used to move muscles, transport solutes, and build new molecules? Certainly each of our cells does not contain a tiny fire in which carbohydrates burn merrily, running a tiny steam engine and over which a tiny organic chemist stews up his reaction mixtures. Nor do we contain a central power plant where, again, carbohydrates are burned, and the energy sent about in little steam pipes or electric cables to run muscle-machines and protein-and-fat factories.

In a living organism, virtually the whole energy system is a chemical one. Energy is generated, transported, and consumed by way of chemical reactions and chemical compounds. Instead of a single reaction with a long plunge from the energy level of carbohydrates and oxygen to that of carbon dioxide and water—as in the burning of a log, say—there are long series of chemical reactions in which the energy level descends in gentle cascades. Energy resides, ultimately, in the molecules involved; as they move through the organism, they carry energy with them.

Constantly appearing in these reactions is one compound, *adenosine triphosphate* (ATP). It is called by biochemists an "energy-rich" molecule, but there



Adenosine triphosphate
ATP

is nothing magical about this. ATP does not carry about a little bag of energy which it sprinkles on molecules to make them react. Nor does it undergo hydrolysis alongside other molecules and in some mystical way make this energy available to them. ATP simply undergoes reactions—only one reaction, really. It *phosphorylates*, that is, transfers a phosphoryl group, $-PO_3H_2$, to some other molecule. For example:

$$\underset{\text{Adenosine triphosphate}}{\text{ATP}} + \underset{\text{An alcohol}}{\text{R—OH}} \longrightarrow \underset{\text{Adenosine diphosphate}}{\text{ADP}} + \underset{\text{A phosphate ester}}{\text{R—OPO}_3\text{H}_2}$$

ATP is called a "high-energy phosphate" compound, but this simply means that it is a fairly reactive phosphorylating agent. It is exactly as though we were to call acetic anhydride "high-energy acetate" because it is a better acetylating agent than acetic acid. And, indeed, there is a true parallel here: ATP is an anhydride, too, an anhydride of a substituted phosphoric acid, and it is a good phosphorylating agent for much the same reasons that acetic anhydride is a good acetylating agent.

When ATP loses a phosphoryl group to another molecule, it is converted into ADP, *adenosine diphosphate*. If ATP is to be regenerated, ADP must itself be phosphorylated, and it is: by certain other compounds that are good enough phosphorylating agents to do this. The important thing in all this is not really the energy level of these various phosphorylating agents—so long as they are reactive enough to do the job they must—but the fact *that the energy level of the carbohydrates and their oxidation products is gradually sinking to the level of carbon dioxide and water*. These compounds—and oxygen—are where the energy is, and ATP is simply a chemical reagent that helps to make it available.

We have seen that very often factors that stabilize products also stabilize the transition state leading to those products, that is, that often there is a parallel between ΔH and E_{act}. To that extent, the energy level of the various phosphorylating agents may enter in, too: less stable phosphorylating agents—less stable, let us say, relative to phosphate anion—may in general tend to transfer phosphate to more stable phosphorylating agents. In addition, of course, if any of the phosphate transfers should be too highly endothermic, this would require a prohibitively high E_{act} for reaction (see Sec. 2.17).

In following sections, we shall see some of the specific reactions in which ATP is involved.

37.4 Biological oxidation of carbohydrates

Next, let us take a look at the overall picture of the biological oxidation of carbohydrates. We start with glycogen ("sugar-former"), the form in which carbohydrates are stored in the animal body. This, we have seen (Sec. 35.9), is a starch-like polymer of D-glucose.

The trip from glycogen to carbon dioxide and water is a long one. It is made up of dozens of reactions, each of which is catalyzed by its own enzyme system. Each of these reactions must, in turn, take place in several steps, most of them unknown. (Consider what is involved in the "reaction" catalyzed by chymotrypsin.) We can divide the trip into three stages. (a) First, glycogen is broken down into its component D-glucose molecules. (b) Then, in *glycolysis* ("sugar-splitting"), D-glucose is itself broken down, into three-carbon compounds. (c) These, in *respiration*, are converted into carbon dioxide and water. Oxygen appears in only the third stage; the first two are anaerobic ("without-air") processes.

The first stage, **cleavage of glycogen**, is simply the hydrolytic cleavage of acetal linkages (Sec. 34.16), this time enzyme-catalyzed.

$$\underset{\text{Glycogen}}{(C_6H_{10}O_5)_n} + nH_2O \xrightarrow{\text{enzyme}} \underset{\text{D-Glucose}}{nC_6H_{12}O_6}$$

The second stage, **glycolysis**, takes eleven reactions and eleven enzymes. The sum of these reactions is:

$$\underset{\text{D-glucose}}{} + \underset{\text{Phosphate}}{2HPO_4^{--}} + 2ADP^{3-} \longrightarrow \underset{\text{Lactate}}{2CH_3CHOHCOO^-} + 2H_2O + 2ATP^{4-}$$

No oxygen is consumed, and we move only a little way down the energy hill toward carbon dioxide and water. What is important is that a start has been made in breaking the five carbon–carbon bonds of glucose, and that two molecules of ADP are converted into ATP. (ATP is required for some of the steps of glycolysis, but there is a *net* production of two molecules of ATP for each molecule of glucose consumed.)

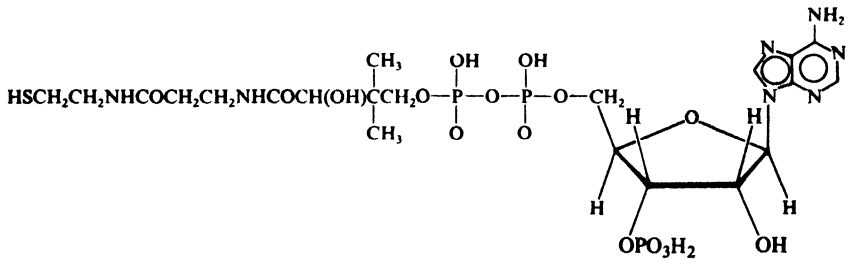
The third stage, **respiration**, is a complex system of reactions in which molecules provided by glycolysis are oxidized. Oxygen is consumed, carbon dioxide and water are formed, and energy is produced.

Let us look at the linking-up between glycolysis and respiration. Ordinarily, the energy needs of working muscles are met by respiration. But, during short periods of vigorous exercise, the blood cannot supply oxygen enough for respiration to carry the entire load; when this happens, glycolysis is called upon to supply the energy difference. The end-product of glycolysis, lactic acid, collects in the muscle, and the muscle feels tired. The lactic acid is removed by the blood and rebuilt into glycogen, which is ready for glycolysis again.

The last step of glycolysis is reduction of pyruvic acid to lactic acid. (The reducing agent is, incidentally, an old acquaintance, reduced nicotinamide adenine

$$\underset{\text{Pyruvate}}{CH_3COCOO^-} + \underset{\substack{\text{Reduced nicotinamide}\\ \text{adenine}\\ \text{dinucleotide}}}{NADH} + H^+ \longrightarrow \underset{\text{Lactate}}{CH_3CHOHCOO^-} + \underset{\substack{\text{Nicotinamide}\\ \text{adenine}\\ \text{dinucleotide}}}{NAD^+}$$

dinucleotide, Sec. 36.15.) Most of the time, however, glycolysis does not proceed to the very end. Instead, pyruvic acid is diverted, and oxidized to acetic acid in the form of a thiol ester, $CH_3CO–S–CoA$, derived from *coenzyme A* and called "acetyl CoA."



Coenzyme A
CoA

It is as acetyl CoA that the products of glycolysis are fed into the respiration cycle.

The acetyl CoA that is fuel for respiration comes not only from carbohydrates but also from the breakdown of amino acids and fats. It is thus the common link between all three kinds of food and the energy-producing process. (Acetyl CoA is even more than that: as we shall see, it is the building block from which the long chains of fatty acids are synthesized.)

Thiols are sulfur analogs of alcohols. They contain the sulfhydryl group, —SH, which plays many parts in the chemistry of biomolecules. Easily oxidized, two —SH groups are converted into disulfide links, —S—S—, which hold together different peptide chains or different parts of the same chain. (See, for example, oxytocin on p. 1143.) Thiols form the same kinds of derivatives as alcohols: *thio*ethers, *thio*acetals, *thiol* esters. Thiol ester groups show the chemical behavior we would expect—they undergo nucleophilic acyl substitution and they make α-hydrogens acidic—this last more effectively than their oxygen counterparts.

37.5 Mechanism of a biological oxidation

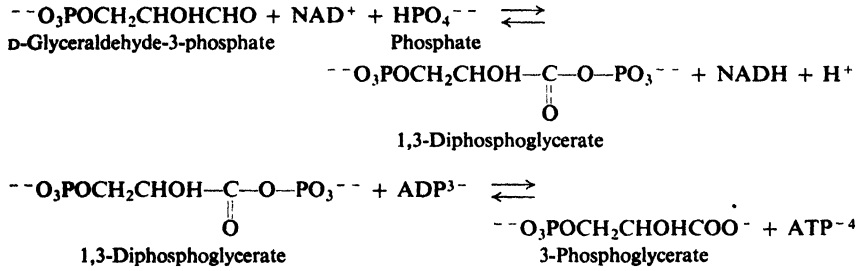
Now let us take just one of the many steps in carbohydrate oxidation and look at it in some detail.

Although there is no *net* oxidation in glycolysis, certain individual reactions do involve oxidation and reduction. About mid-way in the eleven steps we arrive

$H_2O_3P–O–CH_2CHOHCHO$	$H_2O_3P–O–CH_2CHOHCOOH$
D-Glyceraldehyde-3-phosphate	3-Phosphoglyceric acid

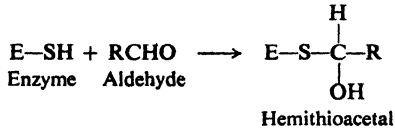
at D-glyceraldehyde-3-phosphate and its oxidation to 3-phosphoglyceric acid. In the course of this conversion, a phosphate ion becomes attached to ADP to generate a molecule of ATP.

Two reactions are actually involved. First, D-glyceraldehyde-3-phosphate is oxidized, but not directly to the corresponding acid, 3-phosphoglyceric acid.



Instead, a phosphate ion is picked up to give the mixed anhydride, 1,3-diphosphoglycerate. This is a highly reactive phosphorylating agent and, in the second reaction, transfers a phosphoryl group to ADP to form ATP.

Now, how does all this happen? The enzyme required for the first reaction is *glyceraldehyde-3-phosphate dehydrogenase* ("enzyme-that-dehydrogenates-glyceraldehyde-3-phosphate"). Its action is by no means as well understood as that of chymotrypsin, but let us look at the kind of thing that is believed to happen. A sulfhydryl group (—SH) of the enzyme adds to the carbonyl group of glyceraldehyde-3-phosphate. Thiols are sulfur analogs of alcohols, and the product is a



hemiacetal: more precisely, a hemi*thio*acetal. Like other acetals, this is both an ether (a *thio* ether) and an alcohol. Such an alcohol group is especially easily oxidized to a carbonyl group (Problem 11, p. 649).

The oxidizing agent is a compound that, like ATP, constantly appears in these reactions: *nicotinamide adenine dinucleotide* (NAD). The functional group here, we remember (Sec. 36.15), is the pyridine ring, which can accept a hydride ion to form NADH. Like the hemiacetal moiety, NAD is bound to the enzyme, and in a position for easy reaction (Fig. 37.3).

Oxidation converts the hemithioacetal into a thiol ester—an acyl enzyme. Like other esters, this one is prone to nucleophilic acyl substitution. It is cleaved, with phosphate ion as nucleophile, to regenerate the sulfhydryl group in the enzyme. The other product is 1,3-diphosphoglycerate. The molecule is (still) a phosphate ester at the 3-position, and has become a mixed anhydride at the 1-position.

The anhydride phosphoryl group is easily transferred; in another enzyme-catalyzed reaction, 1,3-diphosphoglycerate reacts with ADP to yield 3-phosphoglycerate and ATP. The 3-phosphoglycerate goes on in the glycolysis process.

The ATP is available to act as a phosphorylating agent: to convert a molecule

of D-glucose into D-glucose-6-phosphate, for example, and help start another molecule through glycolysis; to assist in the synthesis of fatty acids; to change the cross-linking between molecules of *actin* and *myosin*, and thus cause muscular contraction.

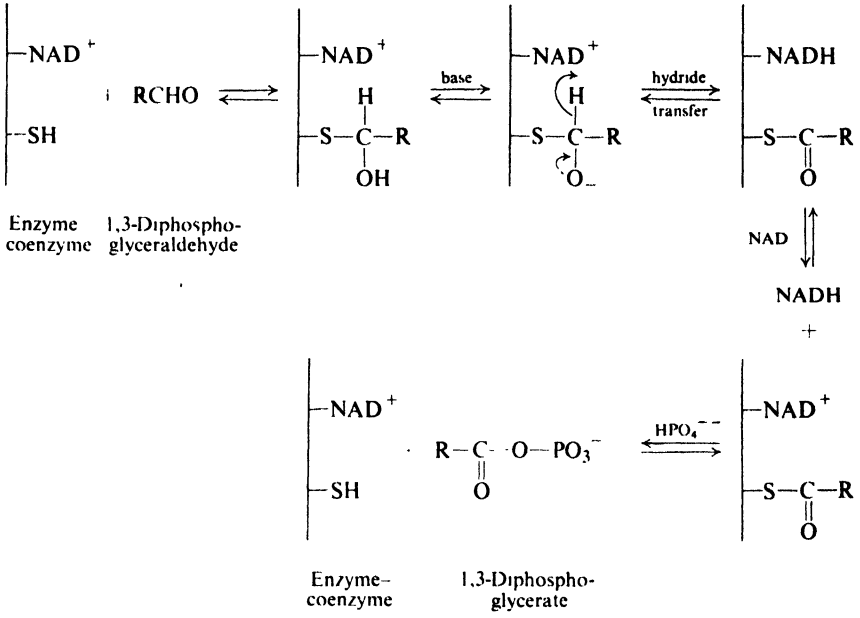


Figure 37.3. Enzymatic conversion of glyceraldehyde-3-phosphate into 1,3-diphosphoglycerate.

The NADH produced is also available to do *its* job, that of reducing agent. It may, for example, reduce pyruvate to lactate in the last step of glycolysis. The extra electrons that make it a reducing agent are passed along, and ultimately are accepted by molecular oxygen.

We are in a strange, complex chemical environment here, but in it we recognize familiar kinds of compounds—hemiacetals, esters, anhydrides, carboxylic acids—and familiar kinds of reactions—nucleophilic carbonyl addition, hydride transfer, nucleophilic acyl substitution.

37.6 Biosynthesis of fatty acids

When an animal eats more carbohydrate than it uses up, it stores the excess: some as the polysaccharide glycogen (Sec. 35.9), but most of it as fats. Fats, we know (Sec. 33.2), are triacylglycerols, esters derived (in most cases) from long straight-chain carboxylic acids containing an *even number* of carbon atoms. These even numbers, we said, are a natural consequence of the way fats are synthesized in biological systems.

There are even numbers of carbons in fatty acids because the acids are built

up, two carbons at a time, from acetic acid units. These units come from acetyl CoA: the thiol ester derived from acetic acid and coenzyme A (Sec. 37.4). The acetyl CoA itself is formed either in glycolysis, as we have seen, or by oxidation of fatty acids.

Let us see how fatty acids are formed from acetyl CoA units. As before, we must realize that every reaction is catalyzed by a specific enzyme and proceeds by several steps—steps that in some direct, honest-to-goodness chemical way, involved the enzyme.

First, acetyl CoA takes up carbon dioxide (1) to form malonyl CoA. (To illustrate the point made above: this does not happen directly; carbon dioxide combines

(1) $\underset{\text{Acetyl CoA}}{CH_3CO\text{-}S\text{-}CoA} + CO_2 + ATP \rightleftharpoons \underset{\text{Malonyl CoA}}{HOOCCH_2CO\text{-}S\text{-}CoA} + ADP + \text{phosphate}$

with the prosthetic group of the enzyme—*acetyl CoA carboxylase*—and is then transferred to acetyl CoA.) Just as in the carbonation of a Grignard reagent, the *carbanionoid* character of the α-carbon of acetyl CoA must in some way be involved.

In the remaining steps, acetic and malonic acids react, not as CoA esters, but as thiol esters of *acyl carrier protein* (ACP), a small protein with a prosthetic group quite similar to CoA. These esters are formed by (2) and (3), which we recognize as examples of transesterification.

(2) $CH_3CO\text{-}S\text{-}CoA + ACP\text{-}SH \rightleftharpoons \underset{\text{Acetyl-S-ACP}}{CH_3CO\text{-}S\text{-}ACP} + CoA\text{-}SH$

(3) $HOOCCH_2CO\text{-}S\text{-}CoA + ACP\text{-}SH \rightleftharpoons \underset{\text{Malonyl-S-ACP}}{HOOCCH_2CO\text{-}S\text{-}ACP} + CoA\text{-}SH$

Now starts the first of many similar cycles. Acetyl-S-ACP condenses (4) with malonyl-S-ACP to give a four-carbon chain.

(4) $CH_3CO\text{-}S\text{-}ACP + HOOCCH_2CO\text{-}S\text{-}ACP \rightleftharpoons \underset{\text{Acetoacetyl-S-ACP}}{CH_3COCH_2CO\text{-}S\text{-}ACP} + CO_2 + ACP\text{-}SH$

At this point we see a strong parallel to the malonic ester synthesis (Sec. 26.2). The carbon dioxide taken up in reaction (1) is lost here; its function was to generate malonate, with its highly acidic α-hydrogens, its carbanionoid α-carbon. Here, as in test tube syntheses, the formation of carbon–carbon bonds is all-important; here, as in test tube syntheses (Sec. 26.1), carbanionoid carbon plays a key role. In the malonic ester synthesis, decarboxylation follows the condensation step; here, it seems, the steps are concerted, with loss of carbon dioxide providing driving force for the reaction.

The next steps are exact counterparts of what we would do in the laboratory: reduction to an alcohol (5), dehydration (6), and hydrogenation (7). The reducing agent for both (5) and (7) is reduced nicotinamide adenine dinucleotide phosphate, NADPH (Sec. 36.15).

(5) $CH_3COCH_2CO\text{–}S\text{–}ACP + NADPH + H^+ \rightleftharpoons$
$D\text{-}CH_3CHOHCH_2CO\text{–}S\text{–}ACP + NADP^+$
D-β-Hydroxybutyryl-S-ACP

(6) $D\text{-}CH_3CHOHCH_2CO\text{–}S\text{–}ACP \rightleftharpoons trans\text{-}CH_3CH{=}CHCO\text{–}S\text{–}ACP + H_2O$
Crotonyl-S-ACP

(7) $trans\text{-}CH_3CH{=}CHCO\text{–}S\text{–}ACP + NADPH + H^+ \rightleftharpoons$
$CH_3CH_2CH_2CO\text{–}S\text{–}ACP + NADP^+$
n-Butyryl-S-ACP

We now have a straight-chain saturated fatty acid, and with this the cycle begins again: reaction of it with malonyl-S-ACP, decarboxylation, reduction, dehydration, hydrogenation. After seven such cycles we arrive at the 16-carbon acid, palmitic acid—and here, for some reason, the process stops. Additional carbons can be added, but by a different process. Double bonds can be introduced, to produce unsaturated acids. Finally, glycerol esters are formed: triacylglycerols, to be stored and, when needed, oxidized to provide energy; and phosphoglycerides (Sec. 33.8) to help make up cell walls.

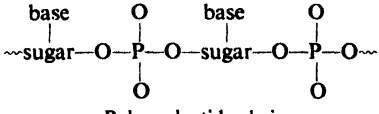
Enzymes are marvelous catalysts. Yet, even with their powerful help, these biological reactions seek the easiest path. In doing this, they take advantage of the same structural effects that the organic chemist does: the acidity of α-hydrogens, the leaving ability of a particular group, the ease of decarboxylation of β-keto acids.

37.7 Nucleoproteins and nucleic acids

In every living cell there are found **nucleoproteins**: substances made up of proteins combined with natural polymers of another kind, the **nucleic acids**. Of all fields of chemistry, the study of the nucleic acids is perhaps the most exciting, for these compounds are the substance of heredity. Let us look very briefly at the structure of nucleic acids and, then, in the next section, see how this structure may be related to their literally vital role in heredity.

Although chemically quite different, nucleic acids resemble proteins in a fundamental way: there is a long chain—a backbone—that is the same (except for length) in all nucleic acid molecules; and attached to this backbone are various groups, which by their nature and sequence characterize each individual nucleic acid.

Where the backbone of the protein molecule is a polyamide chain (a polypeptide chain), the backbone of the nucleic acid molecule is a polyester chain (called a *polynucleotide* chain). The ester is derived from phosphoric acid (the acid portion) and a sugar (the alcohol portion).



Polynucleotide chain

The sugar is D-ribose (p. 1086) in the group of nucleic acids known as ribonucleic

acids (RNA), and D-2-deoxyribose in the group known as deoxyribonucleic acids (DNA). (The prefix *2-deoxy* simply indicates the lack of an —OH group at the 2-position.) The sugar units are in the furanose form, and are joined to phosphate through the C–3 and C–5 hydroxyl groups (Fig. 37.4).

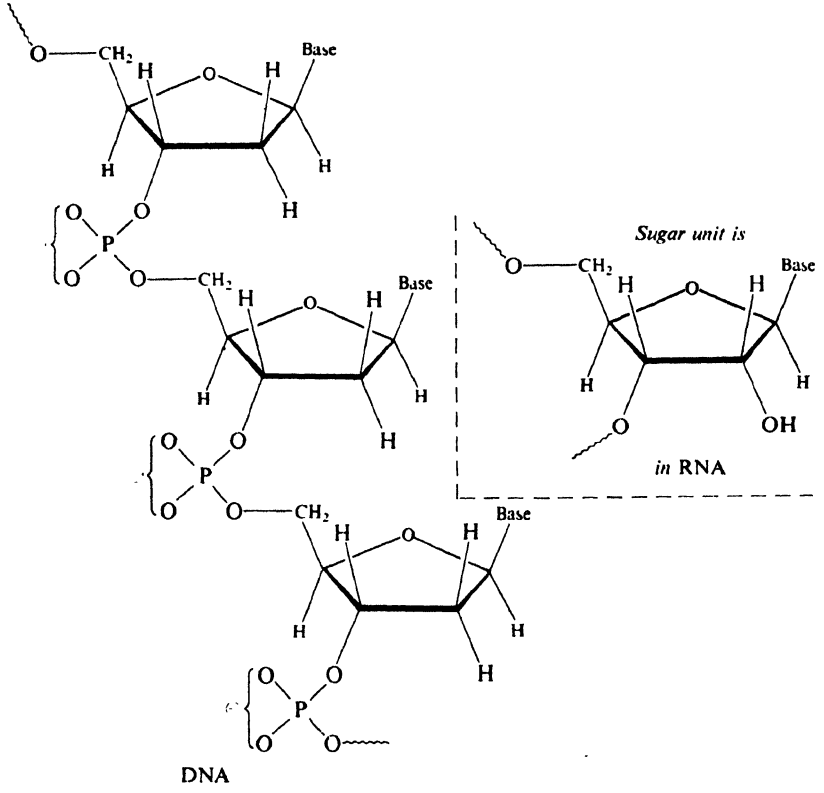


Figure 37.4. Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).

Attached to C–1 of each sugar, through a β-linkage, is one of a number of heterocyclic bases. A base–sugar unit is called a *nucleoside*; a base–sugar–phosphoric acid unit is called a *nucleotide*. An example of a nucleotide is shown in Fig. 37.5.

The bases found in DNA are *adenine* and *guanine*, which contain the purine ring system, and *cytosine*, *thymine*, and *5-methylcytosine*, which contain the pyrimidine ring system. RNA contains adenine, guanine, cytosine, and *uracil*. (See Fig. 37.6.)

The proportions of these bases and the sequence in which they follow each other along the polynucleotide chain differ from one kind of nucleic acid to another. This primary structure is studied in essentially the same way as the structure of proteins: by hydrolytic degradation and identification of the fragments. In this way, and after *seven years* of work, Robert W. Holley and his collaborators at

Cornell University. determined the exact sequence of the 77 nucleotides in the molecule of one kind of transport RNA (p. 1181).

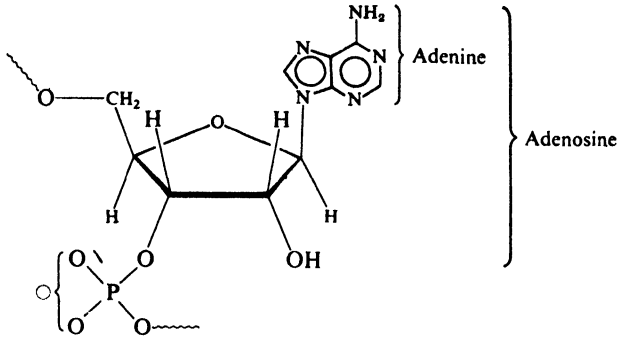


Figure 37.5. A nucleotide: an adenylic acid unit of RNA. Here, the nucleoside is adenosine, and the heterocyclic base is adenine.

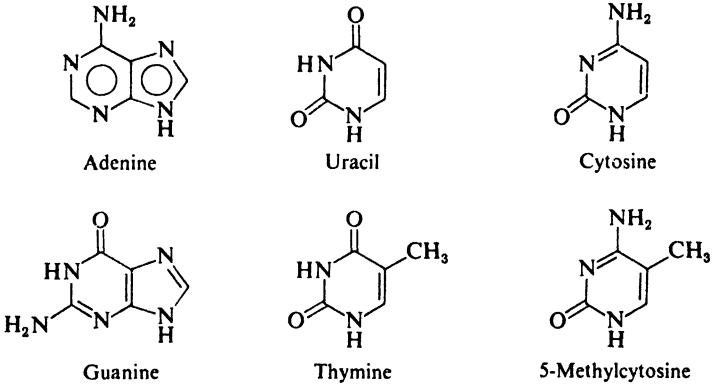


Figure 37.6. The heterocyclic bases of DNA and RNA.

What can we say about the secondary structure of nucleic acids? The following picture of DNA fits both chemical and x-ray evidence. Two polynucleotide chains, identical but heading in opposite directions, are wound about each other to form a double helix 18 A in diameter (shown schematically in Fig. 37.7). Both helixes are right-handed and have ten nucleotide residues per turn.

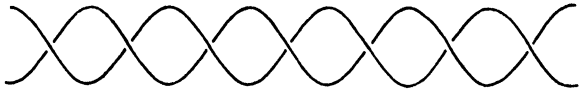


Figure 37.7. Schematic representation of the double helix structure proposed for DNA. Both helixes are right-handed and head in opposite directions; ten residues per turn. Hydrogen bonding between the helixes.

The two helixes in DNA are held to each other at intervals by hydrogen bonding between bases. From study of molecular models, it is believed that these hydrogen bonds can form only between adenine and thymine and between guanine and cytosine; hydrogen bonding between other pairs of bases would not allow them to fit into the double helical structure. In agreement with this idea, the adenine:thymine and guanine:cytosine ratios are found to be 1:1.

Less is known about structures of the various kinds of RNA, although here, too, helixes are involved. In 1973, the precise shape of one RNA molecule—the transport RNA that delivers phenylalanine—was reported: two short segments of double helix at right angles to each other and held together by two loops, the whole making a sort of four-leaf-clover pattern.

So far we have discussed only the nucleic acid portion of nucleoproteins. There is evidence that in one nucleoprotein (found in fish sperm), a polyarginine chain lies in one of the grooves of the double helix, held by electrostatic forces between the negative phosphate groups of the polynucleotide (which face the outside of the helix) and the positive guanidium groups of the arginine residues.

37.8 Chemistry and heredity. The genetic code

Just how is the structure of nucleic acids related to their function in heredity? Nucleic acids control heredity *on the molecular level.* The double helix of DNA is the repository of the hereditary information of the organism. The information is stored as the sequence of bases along the polynucleotide chain; it is a message "written" in a language that has only four letters, A, G, T, C (adenine, guanine, thymine, cytosine).

DNA must both *preserve* this information and *use* it. It does these things through two properties: (a) DNA molecules can duplicate themselves, that is, can bring about the synthesis of other DNA molecules identical with the originals; and (b) DNA molecules can control the synthesis, in an exact and specific way, of the proteins that are characteristic of each kind of organism.

First, there is the matter of self-duplication. The sequence of bases in one chain of the double helix controls the sequence in the other chain. The two chains fit together (as F. H. C. Crick of Cambridge University puts it) like a hand and a glove. They separate, and about the hand is formed a new glove, and inside the glove is formed a new hand. Thus, the pattern is preserved, to be handed down to the next generation.

Next, there is the matter of guiding the synthesis of proteins. A particular sequence of bases along a polynucleotide chain leads to a particular sequence of amino acid residues along a polypeptide chain. A protein has been likened to a long sentence written in a language of 20 letters: the 20 different amino acid residues. But the hereditary message is written in a language of only four letters; it is written in a *code*, with each word standing for a particular amino acid.

The genetic code has been broken, but research continues, aimed at tracking down the lines of communication. DNA serves as a template on which molecules of RNA are formed. It has been suggested that the double helix of DNA partially uncoils, and about the individual strands are formed chains of RNA; the process thus resembles self-duplication of DNA, except that these new chains contain

ribose instead of deoxyribose. The base sequence along the RNA chain is different from that along the DNA template, but is determined *by it*: opposite each adenine of DNA, there appears on RNA a uracil; opposite guanine, cytosine; opposite thymine, adenine; opposite cytosine, guanine. Thus, AATCAGTT on DNA becomes UUAGUCAA on RNA.

One kind of RNA—called, fittingly, *messenger RNA*—carries a message to the ribosome, where protein synthesis actually takes place. At the ribosome, messenger RNA calls up a series of *transport RNA* molecules, each of which is loaded with a particular amino acid. The order in which the transport RNA molecules are called up—the sequence in which the amino acids are built into the protein chain—depends upon the sequence of bases along the messenger RNA chain. Thus, GAU is the code for aspartic acid; UUU, phenylalanine; GUG, valine. There are 64 three-letter code words (*codons*) and only 20-odd amino acids, so that more than one codon can call up the same amino acids: CUU and CUC, leucine; GAA and GAG, glutamic acid.

A difference of a single base in the DNA molecule, or a single error in the "reading" of the code can cause a change in the amino acid sequence. The tiny defect in the hemoglobin molecule that results in sickle-cell anemia (p. 1152) has been traced to a single gene—a segment of the DNA chain—where, perhaps, the codon GUG appears instead of GAG. There is evidence that antibiotics, by altering the ribosome, cause misreading of the code and death to the organism.

Thus, the structure of nucleic acid molecules determines the structure of protein molecules. The structure of protein molecules, we have seen, determines the way in which they control living processes. Biology is becoming more and more a matter of shapes and sizes of molecules.

At the beginning of this book, we said that the structural theory is the basis of the science of organic chemistry. It is much more than that: the structural theory is the basis of our understanding of life.

PROBLEMS

1. Carbon dioxide is required for the conversion of acetyl CoA into fatty acids. Yet when carbon dioxide labeled with ^{14}C is used, none of the labeled carbon appears in the fatty acids that are formed. How do you account for these facts?

2. Taken together, what do these two facts show about chymotrypsin action? (a) The two esters, *p*-nitrophenyl acetate and *p*-nitrophenyl thiolacetate, $p\text{-}NO_2C_6H_4SCOCH_3$, undergo chymotrypsin-catalyzed hydrolysis at the same rate and with the same pH-dependence of rate, despite the fact that —SR is a much better leaving group than —OR. (b) There is no oxygen exchange (Sec. 20.17) in chymotrypsin-catalyzed hydrolysis of an ester RCOOR .

3. In DNA, the bases are bonded to deoxyribose at the following positions (that is, a hydrogen in Fig. 37.6, p. 1179, is replaced by C–1 of the sugar): adenine and guanine, NH in the five-membered ring; cytosine and thymine, NH.

(a) Draw structures to show likely hydrogen bonding between adenine and thymine; between guanine and cytosine. (b) Can you account for the fact that guanine and cytosine pairs hold the chains together more strongly than do adenine and thymine pairs?

4. For each enzyme-catalyzed reaction shown in the following equations, tell what fundamental organic chemistry is involved.

(a) So that acetyl CoA can get through the membrane from the mitochondria where it is formed to the cytoplasm where fatty acids are made, it is converted into citric acid.

$$CH_3CO\text{–}S\text{–}CoA + \underset{\text{Oxaloacetic acid}}{HOOCCOCH_2COOH} \rightleftharpoons \underset{\text{Citric acid}}{HOOCCH_2\overset{\overset{OH}{|}}{C}CH_2COOH} + CoA\text{–}SH$$

(b) Cholesterol is made up of isoprene units derived from isopentenyl pyrophosphate (Sec. 8.26), which is, in turn, formed from mevalonic acid.

$$CH_3CO\text{–}S\text{–}CoA + CH_3COCH_2CO\text{–}S\text{–}CoA \rightleftharpoons$$

$$HOOCCH_2\overset{\overset{CH_3}{|}}{\underset{\underset{OH}{|}}{C}}CH_2CO\text{–}S\text{–}CoA + CoA\text{–}SH$$

$$HOOCCH_2\overset{\overset{CH_3}{|}}{\underset{\underset{OH}{|}}{C}}CH_2CO\text{–}S\text{–}CoA + 2NADPH + 2H^+ \rightleftharpoons$$

$$\underset{\text{Mevalonic acid}}{HOOCCH_2\overset{\overset{CH_3}{|}}{\underset{\underset{OH}{|}}{C}}CH_2CH_2OH} + 2NADP^+ + CoA\text{–}SH$$

5. Three of the bases found in nucleic acids are *uracil*, *thymine*, and *cytosine*. (See p. 1179 for their structures.) They have been synthesized as follows:

(a) urea + ethyl acrylate $\xrightarrow{\text{Michael}}$ [A ($C_6H_{12}O_3N_2$)] $\longrightarrow$
B ($C_4H_6O_2N_2$) + C_2H_5OH

B + Br_2 in acetic acid $\longrightarrow$ C ($C_4H_5O_2N_2Br$)
C + boiling pyridine $\longrightarrow$ uracil ($C_4H_4O_2N_2$)

Give structures of A, B, and C.

(b) Thymine ($C_5H_6O_2N_2$) has been made in the same way, except that ethyl methacrylate, $CH_2{=}C(CH_3)COOC_2H_5$, is used instead of ethyl acrylate. Write equations for all the steps.

(c) uracil + POCl, heat $\longrightarrow$
D ($C_4H_2N_2Cl_2$), chlorine atoms on different carbon atoms
D + NH_3(alc), 100° $\longrightarrow$ E ($C_4H_4N_3Cl$) and F ($C_4H_4N_3Cl$)
E + $NaOCH_3$ $\longrightarrow$ G ($C_5H_7ON_3$)
G + HCl(aq) $\longrightarrow$ cytosine ($C_4H_5ON_3$)

Give structures of D through G.

(d) Six tautomeric structures for uracil have been considered. What are they?

6. In 1904, Franz Knoop outlined a scheme for the biological oxidation of fatty acids that was shown—50 years later—to be correct. In his key experiments, he fed rabbits fatty acids of formula $C_6H_5(CH_2)_nCOOH$. When the side chain (n + 1) contained an even number of carbons, a derivative of phenylacetic acid, $C_6H_5CH_2COOH$, was excreted in the urine; an odd number, and a derivative of benzoic acid was excreted. What general hypothesis can you formulate from these results?

7. In the actual cleavage reaction of glycolysis, D-fructose-1,6-diphosphate is converted into D-glyceraldehyde-3-phosphate and dihydroxyacetone, $CH_2OHCOCH_2OH$. What kind of reaction is this, basically? Sketch out a possible mechanism, neglecting, of course, the all-important role of the enzyme. (*Hints:* The enzyme required is called *aldolase*. See Problem 21.14, p. 711.)

8. Interpret each of the following facts.

(a) In the presence of the proper enzyme, nicotinamide adenine dinucleotide (Sec. 36.15) can oxidize ethanol reversibly to acetaldehyde. When D_2O is the solvent, the re-

$$CH_3CH_2OH + NAD^+ \rightleftharpoons CH_3CHO + NADH + H^+$$

duced NAD formed (NADH) contains no deuterium. When CH_3CD_2OH is oxidized, the reduced NAD formed (NADD) contains one atom of deuterium per molecule.

(b) Enzymatic reoxidation by acetaldehyde of the NADD of part (a) gives NAD^+ that contains no deuterium.

(c) If the NADD of part (a) is oxidized enzymatically by D-glucose, all of the original deuterium remains in the NAD^+.

(d) NADD can also be prepared non-enzymatically by chemical reduction ($Na_2S_2O_4$ in D_2O) of NAD^+. This, too, contains one atom of deuterium per molecule. When it is oxidized enzymatically by acetaldehyde, the NAD^+ formed still contains 0.44 atom of deuterium per molecule.

(e) Acetaldehyde is reduced enzymatically by the NADD of part (a) to give ethanol X. Labeled acetaldehyde, CH_3CDO, is reduced enzymatically by NADH to give ethanol Y. Both X and Y contain one atom of deuterium per molecule. On enzymatic oxidation by (unlabeled) NAD^+, ethanol X gives NADD and unlabeled CH_3CHO, whereas ethanol Y gives unlabeled NADH and CH_3CDO.