an increase in the number of carbon atoms in the molecule. The change in physical state is accompanied by a decrease in solubility. Stearic acid, which melts at 70°C., is practically insoluble in water.

The fatty acids up to and including capric are easily removed from solutions by distillation with steam and hence are known as volatile fatty acids. The determination of the volatile fatty acids is a matter of considerable importance in the analysis of fats, as it aids in distinguishing one type of fat from another. Butter fat, for example, gves a higher proportion of volatile acids than any other fat or oil.

The volatile fatty acids likewise have a decided odor. Butyric acid has a strong odor similar to that of rancid butter. Caproic, caprylic, and capric acids have a pronounced animal odor and are sometimes spoken of as the goat acids. When fats become rancid, a small amount of these volatile fatty acids is formed and gives to the fats a particularly objectionable odor and taste. Thus the objectionable odor of rancid butter is due largely to the presence of free butyric acid.

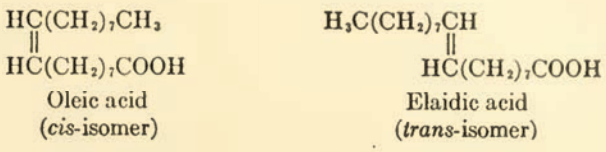
Unsaturated fatty acids

If a fatty acid contains a pair of carbon atoms that are joined together by two bonds instead of one, that is, a *double bond,* it is said to be unsaturated. Such an acid can take up hydrogen, iodine, bromine, oxygen, or other elements by the breaking open of one of these *two* bonds. This leaves an open position on each carbon atom. Therefore, for each double bond, two hydrogens can combine with the compound, which would then be regarded as a saturated compound. This condition of unsaturation has a unique relation to the physical state of the fatty acid and likewise to the glycerides of the fatty acid. For example, oleic acid ($C_{17}H_{33}COOH$), which contains one double bond, and therefore two fewer hydrogens than stearic acid, is a liquid, whereas stearic acid ($C_{17}H_{35}COOH$) is a solid. Oleic acid melts at 14°, whereas stearic acid melts at 70°. If the fatty acid contains more than one double bond, it will have a correspondingly lower melting point; linoleic acid, which contains two double bonds, has a melting point of —18°.

The physical state of the fatty acids is carried over to the glycerides of these acids. Oleic acid and olein are liquids, and stearic acid and stearin are solids. Linolein and linolenin, as expected, are liquid glycerides. Since fats are mixtures of glycerides, a fat will be soft or hard depending upon the proportion of liquid glycerides it contains. Oils differ from fats in that they contain a larger proportion of liquid glycerides. As a general statement, it might be said that oils contain about 80 per cent of unsaturated glycerides, whereas solid fats do not contain more than 40 to 50 per cent. Unsaturation is a fundamental property and, in most fats, is the key to the whole question of their physical state.

If this is kept in mind, an understanding of many physical and chemical properties of fats is easily acquired.

Unsaturated acids have the ability to exist in different isomeric forms, which are called *geometric* isomers. These are designated by the prefixes *cis-* and *trans-*. This type of isomerism, a consequence of the presence of carbon-to-carbon double bonds, may be illustrated by the formulas of oleic acid and its *trans*-isomer, elaidic acid:



Oleic acid (*cis*-isomer) — Elaidic acid (*trans*-isomer)

Acids like linoleic with two double bonds can exist in four geometric isomers, corresponding to the *cis-trans* arrangement about each; in general, the number of isomers possible is $(2)^n$, where n is the number of double bonds present. Generally the natural fatty acids occur in the *cis* form, although vaccenic appears to be a *trans* acid. Where *trans* forms do not occur naturally, they may readily be produced by treating the *cis* acids with nitrous acid or certain other reagents. This reaction has come to be spoken of as "elaidinization" from the circumstance that oleic acid is thus partially converted into elaidic. The *trans* acids are higher melting and less soluble than the corresponding *cis* forms.

The most important unsaturated fatty acids, together with their formulas and occurrence, are listed in Table 4–4. Many other acids with varying numbers of carbon atoms and different degrees of unsaturation have been reportedly obtained from brain, liver, and other tissues. The chemistry of these acids and their function in the animal organism are not yet clearly defined. Their presence in some of the most important organs of the body leaves little room for doubt that their role is an important one.

Although it is rather well established that the animal body can desaturate fats, certain limitations to this process apparently exist in many, if not all, species. Rats kept on diets devoid of unsaturated fats develop a scaliness of the skin, lesions in the kidneys, sterility, and loss of weight, and eventually die. This nutritional deficiency can be prevented by including either linoleic or arachidonic acid in the diet. These particular unsaturated fatty acids have therefore come to be called "essential fatty acids." No one has demonstrated a need of the human body for these acids, but even though they may be required, their widespread occurrence in foodstuffs renders any disease in man resulting from their deficiency quite unlikely.

Quantitative Relations of the Fatty Acids. Many of the statements made in the preceding pages regarding the fatty acids become more

Table 4–4

The chief unsaturated fatty acids

Name	*Formula*	*Occurrence*
I. OLEIC SERIES	Members contain one carbon-to-carbon double bond.	
General formula	$C_nH_{2n-1}COOH$	
Palmitoleic	$CH_3(CH_2)_5CH{=}CH(CH_2)_7COOH$	Liver and other body fats
Oleic *	$CH_3(CH_2)_7CH{=}CH(CH_2)_7COOH$	All animal and vegetable fats. Probably most abundant fatty acid in nature
Vaccenic	$CH_3(CH_2)_5CH{=}CH(CH_2)_9COOH$	Animal fats
Erucic	$CH_3(CH_2)_7CH{=}CH(CH_2)_{11}COOH$	Rapeseed and similar oils
II. LINOLEIC SERIES	Members contain two carbon-to-carbon double bonds.	
General formula	$C_nH_{2n-3}COOH$	
Linoleic *	$CH_3(CH_2)_4CH{=}CHCH_2CH{=}CH(CH_2)_7COOH$	Animal fats, but more particularly vegetable oils, *e.g.*, cottonseed, linseed
III. LINOLENIC SERIES	Members contain three carbon-to-carbon double bonds.	
General formula	$C_nH_{2n-5}COOH$	
Linolenic *	$CH_3CH_2CH{=}CHCH_2CH{=}CHCH_2CH{=}CH(CH_2)_7COOH$	Chief acid of linseed oil
Eleostearic	$CH_3(CH_2)_3(CH{=}CH)_3(CH_2)_7COOH$	Tung nut oil
IV. OTHER UNSATURATED ACIDS		
Ricinoleic	$CH_3(CH_2)_5CHOHCH_2CH{=}CH(CH_2)_7COOH$	Castor oil
Arachidonic (four double bonds between carbons 5-6, 8-9, 11-12, and 14-15) †	$C_{19}H_{31}COOH$	Brain, liver, egg yolk, butter
Clupanodonic (five double bonds between carbons 5-6, 8-9, 12-13, 15-16, and 19-20)	$C_{21}H_{33}COOH$	Fish and marine oils

* Found abundantly in fats and oils. See Table 4–5.

† Counting the carboxyl C as 1.

clear if a study is made of their quantitative distribution. The composition of the mixture of fatty acids obtained by hydrolysis of some common fats and oils is given in Table 4–5.

Butterfat and coconut oil are unique in that they yield such a large number of fatty acids, many of which are lower members of the saturated fatty acid series. Note the small number of fatty acids obtained from lard and the large percentage of unsaturated acids given by the oils.

Nonsaponifiable matter

In addition to glycerol and fatty acids, natural fats contain another type of material called nonsaponifiable matter or "nonsap." This is customarily separated after saponification by extracting the alkaline soap solution with ether. The "nonsap" left after evaporation of the ether consists of fat-soluble pigments, sterols, vitamins, antioxidants, and other miscellaneous substances. Although the nonsaponifiable components constitute only a small part (1–2 per cent) of most natural fats, they are often of great importance in relation to the flavor, color, keeping qualities, and nutritional value of the fat.

GLYCERIDES OF COMMON FATS

Although the percentages of different fatty acids given by hydrolysis of natural fats are fairly accurately known, much less information exists as to the particular glycerides from which these fatty acids are obtained. By crystallizing the fats from acetone and other solvents, a partial separation of the individual glycerides in a number of fats has been made. The separation is a long and laborious procedure, and in no sense complete. All the data accumulated show that the number of glycerides is very great and that they are more complex than was previously supposed.

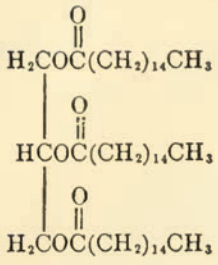
Since glycerol, $C_3H_5(OH)_3$, contains three hydroxyl groups, it can be esterified with one, two, or three molecules of acid to give monoglycerides, diglycerides, and triglycerides, respectively. It is this last type which is found in fats. The three acid radicals in a triglyceride may be all alike, in which case the substance is called a simple glyceride; if more than one kind of radical is present, the compound is called a *mixed glyceride*. The glycerides are named according to the fatty acids involved in their formation.

Thus the ester formed from glycerol and three molecules of palmitic acid is called tripalmitin, or simply palmitin. Its structural formula is written below. Other typical simple glycerides are triolein, tristearin, and tributyrin:

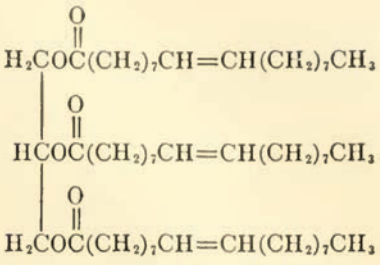
Table 4–5

Fatty acids obtained by hydrolysis of some common fats

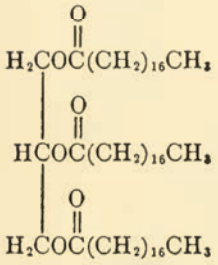
Fatty acid	Percentage of total fatty acids from: *Butter fat*	*Lard*	*Beef fat*	*Olive oil*	*Cottonseed oil*	*Linseed oil*	*Soybean oil*	*Corn oil*	*Coconut oil*	*Peanut oil*
Butyric	4.6									
Caproic	1.8								0.5	
Caprylic	1.3								8.6	
Capric	1.3								6.3	
Lauric	5.0								45.0	
Myristic	17.7	0.7	2.7		0.3				18.0	
Palmitic	16.0	25.2	27.0	6.0	21.0	10.0	9.4	7.8	8.8	6.3
Stearic	3.7	12.8	23.9	4.0	2.0		4.1	3.5	2.1	4.9
Arachidic			0.8		0.6			0.4		3.3
Oleic	48.0	54.2	40.7	83.0	28.6	11.0	28.5	46.3	6.6	60.6
Linoleic		7.1	1.8	7.0	47.2	35.6	55.0	41.8	1.7	21.6
Linolenic						43.4	3.0			
Lignoceric								0.2		2.6
Miscellaneous	0.4									
Undetermined	0.2		3.1		0.3				2.4	0.7
Total:	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0



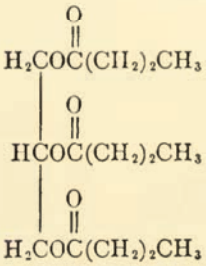
Tripalmitin



Triolein

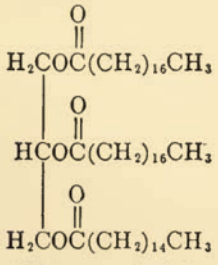


Tristearin

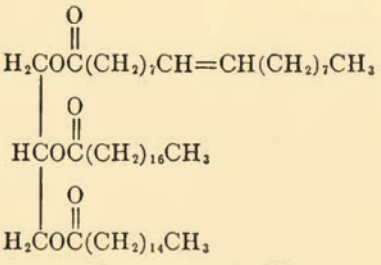


Tributyrin

The ester formed from glycerol combined with two molecules of stearic acid and one of palmitic acid is called distearo-palmitin. This is a typical mixed glyceride. Another is oleo-stearo-palmitin:

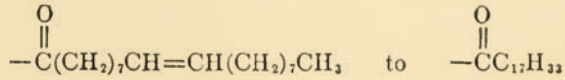


Distearo-palmitin

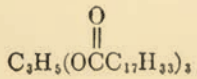


Oleo-stearo-palmitin

More condensed formulas are also frequently written. Thus, for example, the oleic acid radical (oleyl radical) may be abbreviated from



and triolein may be written



However, the more detailed structural formulas should be used by the beginning student until familiarity with them has been gained. It is evident that a great many different individual glycerides might be formed by suitably combining glycerol with the various fatty acids. It has been calculated that ten fatty acids can produce 550 possible combinations.

Formerly it was customary to regard the natural fats as consisting chiefly of simple glycerides, but more recent work shows that they consist largely of mixed glycerides. It is impossible to say with certainty just how much of any given fatty acid goes to make up simple or mixed glycerides in a fat. In the case of butyric acid, it is known that the acid does not exist as a simple glyceride in butter, but is present in a mixed combination. Tributyrin is a bitter substance; obviously, it cannot be present in butter. It is probably more nearly correct to say that natural fats consist essentially of mixtures of mixed glycerides.

PHYSICAL PROPERTIES OF FATS

As already explained, fats may be either *solids* or *liquids* at room temperature (20°C.). The common animal fats are solids at this temperature, and the majority of vegetable fats (oils) are liquids. Fats and oils are lighter than water and, as a rule, have a *specific gravity* of about 0.8. As usually seen they are *noncrystalline,* although many fats can be made to crystallize under suitable conditions. (See Fig. 4–1.) They are *poor conductors of heat,* therefore serving a useful purpose in the insulation of the body.

Fats are *colorless* when they are obtained in a pure state. The more or less yellow color common to many fats is due to the presence of a pigment, and not to the fat itself. Butter, for example, varies in color with the season. In June, when cows feed on grass, the butter is highly colored, while in January, when the animals receive a dry ration, the butter is paler in color. The yellow pigment, therefore, is contained in the feed of the animals. This is largely carotene, $C_{40}H_{56}$, which occurs in all green plants, in the petals of many flowers, such as the narcissus, and in many vegetables such as carrots, squash, etc. In grass it is not evident because the green pigment, chlorophyll, masks the carotene. Xanthophyll, $C_{40}H_{56}O_2$, is another yellow pigment which is widely distributed in plant materials. This is the chief coloring pigment found in egg yolk. It is also contained in butterfat, but in a much smaller percentage than is carotene. The egg yolk likewise varies in color with the feed of the poultry. Lard contains no coloring matter, probably because swine are fed on rations largely free from green material.

The coloring of fats, which is of importance commercially, has been brought into considerable prominence in connection with the vitamin A

content of food. In 1919 Steenbock called attention to the occurrence of vitamin A in close association with the yellow pigmentation of fats and foods; for example, yellow corn was found to be rich in vitamin A, while white corn was deficient in this vitamin. The same correlation between vitamin and pigment was found with respect to carrots, squash,

From Hawk and Bergeim, *Practical Physiological Chemistry.*
Courtesy of P. Blakiston's Son & Co., Inc.

Fig. 4–1. Crystals of beef fat.

cabbage, and other vegetables. The work of von Euler, Moore, and others has demonstrated that carotene of plants is the precursor of vitamin A in the animal body. Carotene is transformed in the body, apparently in the liver, from an intensely yellow pigment into an almost colorless compound. This transformation is accompanied by changes in structure and other properties of the pigment (see p. 207).

In the pure state, fats have no *taste,* but by a curious anomaly *natural* fats are the chief materials that give flavor to food. Food prepared without the addition of fat is considered by many people unpalatable; liberal additions of butter make food particularly appetizing. Butter owes its taste largely to diacetyl ($CH_3\,CO\,CO\,CH_3$) and acetylmethyl-carbinol ($CH_3\,CO\,CHOH\,CH_3$), compounds produced by bacteria in the ripening of the cream.

The same condition obtains with respect to *odor.* When purified, fats have no odor; yet natural fats frequently have marked odors. This apparent contradiction is explained by the readiness with which fats take up odors. The housewife carefully avoids putting onions and butter together in the refrigerator. The absorption of odors by fats is

used to advantage in the extraction of delicate perfumes from flowers. Rose petals are spread on glass plates covered with a thin layer of lard and tallow and left in contact with the fat for 24–72 hours. At the end of this time the old flowers are removed and a new lot is added. Finally the fat is extracted with cold alcohol to remove the essence, and the alcoholic solution is concentrated or bottled directly. This process yields perfumes of the finest quality.

The substances that often are present in natural fats and, as explained above, impart to them characteristic colors, flavors, and odors are not chemically related to the fats themselves, but are merely associated with them on account of being "fat-soluble," that is, easily soluble in fat. Thus carotene is a fat-soluble pigment. When the cream is churned nearly all of the carotene remains in the butter rather than in the buttermilk, which of course is largely an aqueous medium. Substances that are fat-soluble are usually insoluble in water, and vice versa.

CHEMICAL PROPERTIES OF FATS

Hydrolysis and saponification

The most important chemical reaction of fats is *hydrolysis,* with the production of glycerol and fatty acids. This process may be brought about by means of acids, superheated steam, or enzymes. The fat-splitting enzymes are known as *lipases.* They occur in many tissues of the body and in plant material, especially in oily seeds. A well known fat-splitting enzyme is *steapsin,* which is secreted by the pancreas and is involved in the digestion of fats. Many bacteria, molds, and other microorganisms produce fat-splitting enzymes. An equation illustrating the hydrolysis of a fat, for example, palmitin, is:

$$\underset{\text{Palmitin}}{C_3H_5(OCC_{15}H_{31})_3} + 3H_2O \xrightarrow{\text{steapsin}} \underset{\text{Glycerol}}{C_3H_5(OH)_3} + 3\underset{\text{Palmitic acid}}{HOCC_{15}H_{31}}$$

If the hydrolysis is brought about by means of alkali, a soap is formed instead of a fatty acid, and the process is then called *saponification:*

$$\underset{\text{Palmitin}}{C_3H_5(OCC_{15}H_{31})_3} + 3NaOH \longrightarrow \underset{\text{Glycerol}}{C_3H_5(OH)_3} + 3\underset{\text{Sodium palmitate (soap)}}{NaOCC_{15}H_{31}}$$

In this equation the formula of sodium palmitate is written in such a manner as to show how it is produced by the action of the NaOH on the palmitin. It might also be written $C_{15}H_{31}COONa$. This substance

is a typical soap, and the above equation illustrates the commercial manufacture of soaps.

Soaps

A soap is defined chemically as a metallic salt of a fatty acid containing ten or more carbon atoms. All commercial soaps, however, are mixtures of several individual "soaps" because they are made from natural fats which are mixtures of glycerides. The glycerides are all saponified at once, and each fatty acid radical is converted into the corresponding soap. Thus the product is a mixture corresponding in composition to the fatty acid make-up of the original fat.

Sodium and potassium soaps, being fairly soluble in water, are useful washing agents. Soaps of other metals, although too insoluble in water to form a lather, are very valuable for other purposes, as described below. The consistency and washing qualities of soaps depend partly on the metal and on the fatty acid radicals of which they are composed. Thus from a given fat, sodium hydroxide will tend to produce the harder and potassium hydroxide, the softer soap. On the other hand, if the same alkali is used throughout, a liquid fat, containing unsaturated or low molecular weight fatty acid radicals, will tend to produce a soft or liquid soap, whereas a hard fat like tallow will make a hard soap. Other things being equal, the soaps of capric, lauric, and myristic acids, that is the saturated fatty acids containing 10, 12, and 14 carbon atoms, lather better, and of these the lauric soaps are the best. It is for this reason that fats such as palm and coconut oils, yielding a large amount of these particular fatty acids on hydrolysis, are so valuable in soap making.

Water-soluble soaps are classified as *detergents*, substances which, when dissolved in water, lower the surface tension of the water and help to loosen and wash away particles of grease and dirt. Other kinds of detergents are also produced more or less directly from fats and have become so popular that over 1.2 million lb. were produced in 1950. These so-called "synthetic detergents," which should not be called soaps, are of many types, but all of them consist of a water-soluble, salt-like group attached to a long-chain, fat-like residue. A typical example is sodium alkyl sulfate, $ROSO_3Na$, where R represents alkyl groups corresponding to various fatty acids such as lauric, myristic, palmitic, and stearic. These synthetic detergents differ mainly from ordinary soaps by having a sulfate in place of the carboxyl group. Since they form soluble calcium and magnesium compounds, which are not precipitated by the minerals in hard water, they are as effective washing agents in hard water as in soft water. Their aqueous solutions are also practically neutral, whereas those of ordinary soaps are quite strongly alkaline and have a pH of about 9.

Various water-insoluble soaps of metals other than sodium or potassium have important industrial uses. Soaps of the higher saturated or slightly unsaturated fatty acids such as stearic, palmitic, and oleic acids, with such metals as aluminum, calcium, lead, barium, lithium and others, are used in making lubricating greases. When combined with lubricating oils, these soaps produce semisolid gels, or greases. More highly unsaturated acids such as linoleic or eleostearic are combined with lead, manganese, or cobalt to produce "driers" for use in paints, varnishes, and enamels. These soaps catalyze the oxidation processes, which cause the films to "dry" or harden. Zinc oleate and stearate are used as antiseptics and astringents in medicinal preparations. Of the above soaps, aluminum and zinc stearates are quantitatively the most important, being produced annually to the extent of some 10 million lb. each.

Acrolein test

When glycerol is heated strongly, and especially if a dehydrating agent such as potassium bisulfate is present, it decomposes into water and acrolein:

$$C_3H_5(OH)_3 \xrightarrow[KHSO_4]{\Delta} CH_2{=}CHCHO + 2H_2O$$

The unsaturated aldehyde, acrolein, has a characteristic sharp, irritating odor and is partially responsible for the smell of burnt fat.

Fats likewise give the acrolein test since on heating to a sufficiently high temperature the glycerides in the fat are partially broken down with the eventual formation of acrolein.

Iodine number

The unsaturation of a fat is determined by means of iodine, which gives the so-called iodine number of a fat. The iodine number is the percentage of iodine by weight that the fat will absorb; for example, if a fat has an iodine number of 100, one gram of the fat will absorb one gram of iodine. The following table shows how the iodine number generally varies with the physical state of the fats. It will be noted that the hard fat, tallow, has a low iodine number, 35–45, whereas lard, a soft fat, has an iodine number of 50–70, and the oils have iodine numbers ranging from 80 to 200. If judged by the low iodine number, butter, and especially coconut oil, should be hard fats. The low melting point of coconut oil (about 25°) as compared with that of tallow (about 45°) is due to the presence of large quantities of glycerides of the lower saturated fatty acids. The softness of butter is caused by two factors, unsaturation and glycerides of low molecular weight. In most fats and oils only the first of these two factors plays a part.

Table 4–6

Iodine number of some common oils and fats

Coconut oil	5– 10
Butterfat	26– 38
Beef tallow	35– 45
Oleo oil from beef tallow	40– 55
Lard	50– 70
Olive oil	79– 90
Peanut oil	87–100
Cottonseed oil	104–116
Corn oil	111–124
Soybean oil	137–143
Linseed oil	170–200

Linseed oil has the highest iodine number of any known fat. It is a highly unsaturated oil and takes up atmospheric oxygen very readily to form a hard tough film. For this reason it is peculiarly well adapted for paint purposes. When paint is spread over a surface, the linseed oil takes up oxygen from the air and forms a thin, hard, watertight coat. No other oil has ever been found which is equal to linseed oil in this respect. Tung and soybean oils come nearer to it than any other oils and are used to supplement linseed oil in the paint industry. At the present time the demand for linseed oil is far greater than can be met by the supply. Were it possible to desaturate other oils such as cottonseed and peanut oil so that they would have the same drying capacity as linseed oil, it would be of enormous benefit to the paint industry. Unfortunately no method for doing this has yet been discovered, although the reverse process of saturating an unsaturated oil can be easily accomplished. The term "drying oil" is applied to liquid fats, like linseed oil, which have iodine numbers in the range 150 to 200 and form hard, dry films when spread over a surface and exposed to the air.

The drying oils are unsuited for lubricating purposes because they tend to become gummy and sticky. The same property of unsaturation operates, but in lubricating oils it is an undesirable property rather than a desirable one.

Hydrogenation of oils

The saturating or hardening of fats has become an important commercial process. If an oil or soft fat is exposed to the action of hydrogen in the presence of finely divided nickel (a catalyst) at a moderately high temperature (150° to 190°C.) and pressure (25 lb. per sq. in.), it combines with the hydrogen and is converted into a solid fat. This process is applied annually to several hundred thousand tons of unsaturated fats in the United States. Although not all of this hydrogenated

material is converted into edible fats, great quantities of well known commercial products, *e.g.*, "Crisco" and "Snowdrift," are prepared in this manner from peanut, cottonseed, and other oils. These fats are more stable to heat than natural fats, such as lard, and are, therefore, peculiarly well adapted to certain cooking operations, such as deep-fat frying. The natural fats tend to decompose at higher temperatures, owing, it is assumed, to the presence of small amounts of free fatty acid. The more free acid present in a fat, the more readily it is decomposed by heat.

Rancidity of fats

When fats are kept for a long time, they develop objectionable odors and tastes, a condition which is known as rancidity. Many different factors such as heat, light, moisture, air, enzymes, bacteria, and metals are involved in the decomposition of fats. The principal chemical changes are hydrolysis and oxidation. The former is particularly important in the case of butter and other dairy products because the free fatty acids, produced when butterfat is hydrolyzed, include several of the lower, saturated series (*e.g.*, butyric, caproic, etc., see Table 4–5), which have very sharp, unpleasant odors. Rancidity due to oxidation develops particularly in moderately unsaturated fats and oils, a group which includes the bulk of the common food fats. Atmospheric oxygen slowly reacts to produce hydroperoxides, fats or fatty acids having an —OOH group attached to a carbon atom next to a double bond. Once formed, the hydroperoxides serve as catalysts for further oxidation. As a result, lower fatty acids, ketones, peroxides, and other substances are formed, and the glycerol disappears. The unpleasant odors and flavors of these products make the fat rancid. Rancidity is a term which applies to any objectionable odor or taste in fats, no matter how brought about.

There exist a number of different substances, not themselves fats, which have a remarkable power of slowing down the development of oxidative rancidity. Such substances are called *antioxidants*. They are present naturally in many fats, which have not been too extensively refined, and have a large influence on the keeping qualities of fats. Examples of antioxidants are crude lecithin, hydroquinone (HOC_6H_4OH), vitamin C, and the tocopherols (vitamins E, p. 216). Only small amounts, of the order of 1 per cent of the fat, or less, are sufficient to delay the onset of oxidative rancidity for extended periods. Antioxidants appear to function by interfering with the catalytic effect of the hydroperoxides mentioned above. In so doing they are themselves slowly oxidized, so that their effect eventually wears off. Antioxidants are deliberately added to many food fats to improve their keeping qualities. They are also thought, by many investigators, to play an important biological role in preventing unwanted oxidations from occurring *in vivo*.

Determination of fat

The percentage of fat in a foodstuff is determined by extracting the fat with ether and weighing it. In tables of analyses this is generally spoken of as fat, or more correctly, ether extract. It is not necessarily all fat, since ether will dissolve many other substances such as waxes, resins, fatty acids, and coloring matter, all of which may be contained in natural fats. The ether extract of cereals is mainly fat, whereas a large proportion of that obtained from vegetables consists of fatty acids, phospholipides, nonsaponifiable matter, etc. The following table shows how variable is the composition of ether extract:

Table 4–7

Composition of ether extract

Material extracted	*Neutral fats* (per cent)	*Free fatty acids* (per cent)	*"Lecithin"* (per cent)	*Nonsaponifiable matter* (per cent)
Potatoes	16.3	56.9	3.1	10.9
Beets	23.0	35.3		10.7
Corn	88.7	6.7		3.7
Barley	73.0	14.0		6.1
Oats	59.2	35.4	0.8	2.7
Peas	58.6	11.2	27.4	7.4
Soybeans	95.5	1.2	1.8	1.5

A special test for the determination of fat in certain dairy products, particularly milk, was introduced in 1890 by Stephan Moulton Babcock. The "Babcock test," as it is universally called, has been of decisive importance to the growth of dairying in this country, since it made possible a quick, practical means of judging the butter fat production of individual cows, permitting selection of the best producers for breeding. The test is made by treating a definite amount of milk with an equal volume of 90 per cent sulfuric acid and warming the mixture gently. On centrifuging this mixture, the fat separates as a distinct layer, which is measured in the neck of a special flask calibrated to read directly the percentage of butter fat.

WAXES

Definition

Waxes are classified as simple lipides, together with the true fats, but unlike fats they contain higher monohydroxy (sometimes dihydroxy) alcohols in place of glycerol. These alcohols exist in the wax in com-

bination with fatty acids, that is, as esters. An example of an individual wax ester is cetyl palmitate:

$$CH_3(CH_2)_{14}CH_2OC(=O)(CH_2)_{14}CH_3$$

The natural waxes are mixtures of many such esters and often contain hydrocarbons as well.

Occurrence and importance

Waxes are widely distributed in nature, both in plant and animal material. They are generally found on the external surfaces, where they serve as a protective coating and prevent undue evaporation of moisture. They keep the feathers of birds and the wool and hair of animals soft and pliable. Some of the important natural waxes are as follows: beeswax, secreted by the honey bee; lanolin (wool wax or degras), a waxy material obtained from wool; spermaceti and sperm oil, solid and liquid waxes, respectively, which are separated from the oily liquid in the head of the sperm whale; Chinese wax, the secretion of an insect; and carnauba wax, a coating found on the leaves of the carnauba (Brazil) palm. Waxes are important commercial materials and are extensively used in the manufacture of candles, floor polishes, and varnishes. Many industrial waxes, however, are synthetic products made from higher chlorinated hydrocarbons, etc.

Properties

The waxes are, as a rule, solid materials that have a melting point between 60° and 80°C. They are insoluble in water and are poor conductors of heat. Waxes are much more resistant to saponification than fats or oils. Light, air, bacteria, molds, and other agents that readily bring about changes in the fats have little or no effect upon the waxes. Because of these properties it is evident that they are particularly suitable for the protection of the surfaces of animals and plants.

Composition

Waxes can be separated into their constituents by saponification with alkali, which slowly forms soaps of the acidic components, and extraction of the aqueous soap solution with ether. The ether dissolves the nonsaponifiable material (higher alcohols and hydrocarbons), which in the case of waxes amounts to about 50 per cent of the original weight (note difference from fats, p. 81). A list of the chief components of several natural waxes is given in Table 4–8.

Table 4–8

The chief components of natural waxes

Common name	*Number of carbons*	*Formula*	*Source*
I. ALCOHOLS			
Cetyl	16	$CH_3(CH_2)_{11}CH_2OH$	Spermaceti
Oleyl	18	$CH_3(CH_2)_7CH{=}CH(CH_2)_7CH_2OH$	Spermaceti, blubber, fish oils
Ceryl	26	$CH_3(CH_2)_{24}CH_2OH$	Chinese insect wax, beeswax
Montanyl	28	$CH_3(CH_2)_{26}CH_2OH$	Beeswax, carnauba wax, many fruit coat and leaf waxes
Melissyl	30	$CH_3(CH_2)_{28}CH_2OH$	
*	32	$CH_3(CH_2)_{30}CH_2OH$	
II. FATTY ACIDS			
Palmitic	16	$CH_3(CH_2)_{14}COOH$	Beeswax, spermaceti
Cerotic	26	$CH_3(CH_2)_{24}COOH$	Beeswax, Chinese insect wax, carnauba wax, fruit coat and leaf waxes
Montanic	28	$CH_3(CH_2)_{26}COOH$	
Melissic	30	$CH_3(CH_2)_{28}COOH$	
*	32	$CH_3(CH_2)_{30}COOH$	Beeswax, carnauba wax
*	34	$CH_3(CH_2)_{32}COOH$	
III. PARAFFIN HYDROCARBONS			
*	27	$CH_3(CH_2)_{25}CH_3$	Beeswax
*	29	$CH_3(CH_2)_{27}CH_3$	Beeswax, carnauba wax
*	31	$CH_3(CH_2)_{29}CH_3$	Beeswax, leaf waxes
*	33	$CH_3(CH_2)_{31}CH_3$	

* No common name.

Note particularly the high carbon number of most of the components listed, and the fact that the hydrocarbons all contain *odd* rather than even numbers of carbon atoms. Although lanolin is one of the most common and important waxes, it is not included in Table 4–8 because of its unique composition. It contains over 20 unusual branched chain acids and hydroxyacids combined as esters with a variety of alcohols, among which are ceryl alcohol, cholesterol, other sterols, and at least two alcohols of the triterpene type. Another unusual type of wax constituent, reportedly present in Chinese urushi wax, is represented by two of the higher dibasic acids, $HOOC(CH_2)_{18}COOH$, and $HOOC(CH_2)_{20}COOH$.

STEROLS

Sterols are solid, cyclic alcohols usually containing 27–29 carbon atoms, 17 of which are arranged in a characteristic ring system consisting of three six-membered rings and one five-membered ring (see Fig. 4–2). Over thirty sterols have been found in nature. They occur in the tissues of animals, in plants, abundantly in yeasts and molds, but apparently not at all in bacterial cells. Some are saturated; others contain one, two or three double bonds. Various side chains and one or more hydroxyl groups complete the sterol structure (see formulas of cholesterol below, and ergosterol, p. 213). The sterols may occur free, or combined with fatty acids as esters, or with carbohydrates as glycosides.

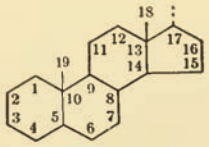


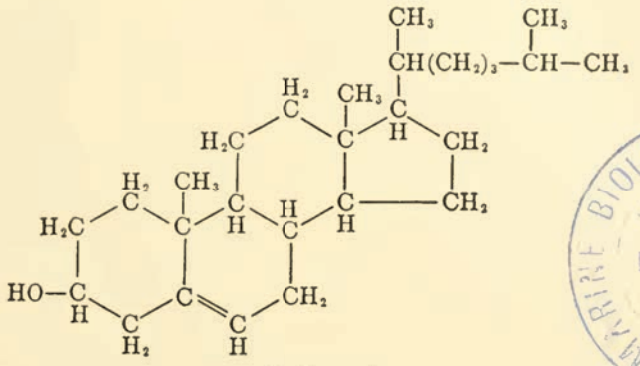
Fig. 4–2. Diagram and numbering of the steroid ring system.

Several types of substance closely related to the sterols are of particular biological importance. They include the bile acids, sex hormones (p. 292), adrenal cortical hormones (p. 290), vitamins D (p. 210), several heart-stimulating drugs (*e.g.*, digitalis), saponins, and others. The sterols and their related substances are collectively designated as *steroids*. All contain the characteristic steroid ring system (Fig. 4–2).

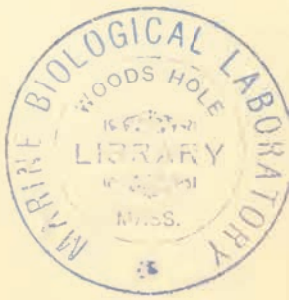
Cholesterol

This substance is the characteristic sterol of higher animals. Although present in every cell of the body, it is most abundant in egg yolk, animal fats like cream and butter, cod liver oil, and especially in nerve and brain tissue. In fact, the concentration in the human brain may reach the

surprisingly high value of 17 per cent of the dry weight. Cholesterol is also the chief component of human gallstones, deriving its name from this circumstance (Greek: *chole,* bile + *stereos,* solid). It has the formula:



Cholesterol



The general distribution of cholesterol in the body indicates that it must perform some important function there. It has been found that cholesterol protects the blood corpuscles from the dissolving action of certain poisons (saponins). It also checks the tendency of the bile salts to dissolve the blood corpuscles and inhibits the action of the fat-splitting enzymes (lipases). A closely related compound, 7-dehydrocholesterol, is the precursor of one of the important D vitamins.

In some types of heart disease (atherosclerosis) so much cholesterol and other lipides are deposited on the inner walls of the heart arteries that the blood supply to the heart muscle is seriously impaired. Since this condition is associated with a high level of cholesterol in the blood, it has been suggested that diets rich in cholesterol should be avoided. Unfortunately this would eliminate the use of eggs, milk, butter, pork, liver, and other highly nutritious foods. Furthermore, even if no cholesterol is ingested, large amounts are synthesized in the body from other substances. It has been shown that the blood cholesterol level is more definitely associated with other factors such as age and obesity than it is with dietary intake.

Sitosterol

Vegetable oils such as corn, wheat, cottonseed, and linseed oils contain sitosterol and other sterols. Sitosterol is found in the nonsaponifiable fraction of the oil. Some investigators have maintained that the plant sterols are converted into cholesterol and other zoosterols in the animal body. There is abundant evidence, however, that cholesterol is syn-

thesized in the body from two-carbon fragments produced during carbohydrate and fat metabolism.

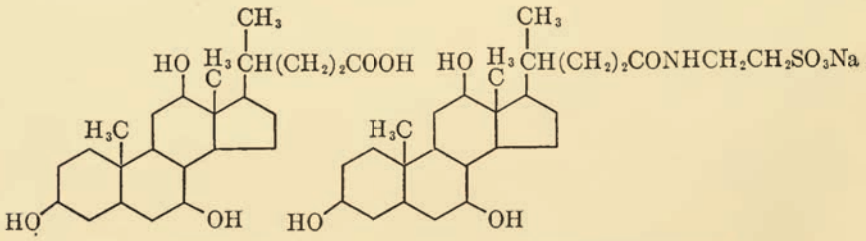
Ergosterol

This sterol has acquired great importance because of the discovery that it is the parent substance of one of the D vitamins. When ergosterol is exposed to ultra violet light for a few minutes, several new compounds are formed. One of these, calciferol, has, besides chemical properties that differentiate it from its parent substance, marked antirachitic potency. (See discussion on vitamin D, p. 212, for details regarding these changes.)

Ergosterol is a white, crystalline compound, which has a melting point of 160–161°C. It is obtained commercially from yeast and mold cells, where it occurs in concentrations up to 2 per cent of the dry weight of the fungal tissue. The dry yeast or mold is extracted with alcohol and benzene, and on concentration of the extract, crystals of crude ergosterol separate. Its chief commercial use is in the manufacture of vitamin D preparations.

BILE SALTS

Bile aids in fat digestion (p. 317) because most of its solid matter consists of *bile salts,* strong emulsifying agents which form water-soluble complexes with fatty materials. These substances are sodium salts of peptide-like combinations of glycine or taurine ($NH_2CH_2CH_2SO_3H$) with *bile acids,* steroids bearing a carboxyl and one or more hydroxyl groups. The commonest and most abundant bile acid is *cholic acid,* and the glycine and taurine conjugates of it are *glycocholic* and *taurocholic* acids, respectively.



Cholic acid

Sodium salt of taurocholic acid
(a typical bile salt)

PHOSPHOLIPIDES

Phospholipides, or phosphatides, are fat-like substances that contain phosphorus and nitrogen. Upon hydrolysis they are broken down into

their components, which include fatty acids or aldehydes, phosphoric acid, a nitrogenous base, and a polyhydroxy alcohol. Phospholipides may be divided into three subgroups according to the alcoholic component, as follows:

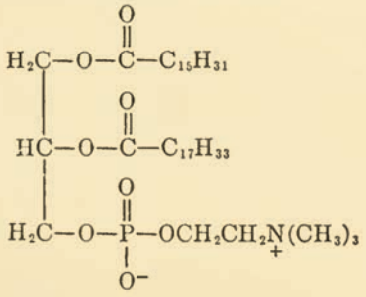
glycerol:	lecithin, cephalin, plasmalogen
inositol:	lipositol, mono- and diphosphoinositides
sphingosine:	sphingomyelin

The lecithins and cephalines are the best known and most important members of the entire group.

Phospholipides are believed to occur in every cell and are particularly abundant in some of the most important and active tissues of the body: brain, liver, and mammary gland. They appear to be an essential part of the actual cell structure, and not merely stored-up food as are the true fats in adipose tissue. This essential character is indicated by the fact that the amount of phospholipides in the tissues is not materially reduced during extreme starvation. For this reason these substances are often called "essential lipides," or the "nonvariable component" of tissue lipides, in contrast to the true fat, or "variable component."

Lecithins

The structure of a lecithin may be illustrated by the formula:



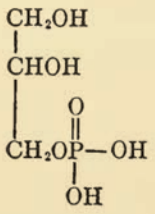
This formula shows that the substance is a glyceride in which one fatty acid radical has been replaced by a group consisting of phosphoric acid combined as an ester with choline (p. 162). Since the phosphate residue is strongly acidic and the choline nitrogen strongly basic, the two neutralize each other and form an inner salt, or *zwitterion,* which is indicated in the formula by the + and — signs. The fatty acids in lecithins tend to be rather highly unsaturated, although some saturated acids are usually present. A *hydrolecithin,* in which both fatty acids (palmitic) are saturated, has also been found in brain and lung tissue. It is much less soluble in the usual fat solvents than unsaturated lecithins. It is theoretically possible to have lecithins in which the phosphoric acid-choline

group is attached to the center, or *beta*, carbon atom of glycerol. However, it seems very doubtful if such *beta* lecithins exist in nature.

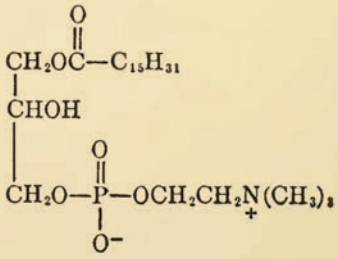
Many different lecithins have been found in brain, nerves, liver, pancreas, heart, blood, and other active tissues of the body. Egg yolk and soybeans are especially rich (2 to 5 per cent) in lecithins and serve as starting materials for commercial preparation. Soybean lecithin, produced in ton quantities, is used as an emulsifying agent in many food products; for example, the addition of 0.2 per cent soybean lecithin to oleomargarine gives a consistency that closely resembles that of butter. To obtain the crude lecithin, the fat extracted from the original source material is treated with acetone. True fats are soluble; lecithins and cephaline, being insoluble, precipitate out. The precipitation can be made more complete by adding magnesium chloride or cadmium chloride, which form sparingly soluble addition compounds with lecithins and cephalins. The cephalins are separated from lecithins by virtue of their lower solubility in alcohol.

Highly purified lecithins are colorless, greasy, paraffin-like substances. Colorless preparations, however, are seldom seen, since they are difficult to obtain and quickly darken on exposure to air. The outstanding characteristic of the lecithins is their high chemical reactivity. They are easily oxidized, easily hydrolyzed, and have a great capacity for combining with other substances such as water, salts, proteins, and carbohydrates. As indicated above they are also excellent emulsifying agents, probably because their structure includes both the water-attracting salt group and the long carbon chains of the fatty acid residues (compare soaps and detergents, p. 87).

Hydrolysis of lecithins, with moderately strong acid or alkali, readily breaks them down into the constituent fatty acids, choline, and α-glycerophosphoric acid. The attachment between the glycerol and the phosphoric acid is very strong. It is only broken by long boiling with strong acid.



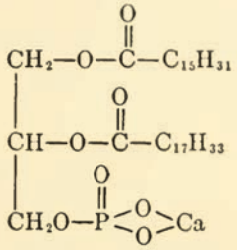
α-Glycerophosphoric acid



A lysolecithin

Enzymatic hydrolysis also occurs in living tissues as a result of the action of lecithinases. One type of lecithinase (type B) splits off both fatty acid residues. Another (called type A) removes only one, leaving a product, lysolecithin, which is poisonous because of its power to hemolyze

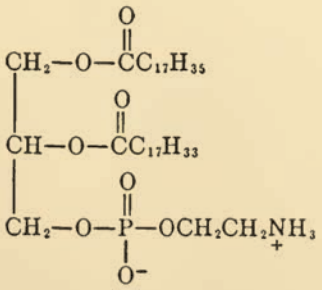
(dissolve) red blood cells. Cobra snake venom contains lecithinase A, and presumably owes its deadly effect, at least in part, to the production of lysolecithin. Still another hydrolysis product of lecithins (or cephalins) are the phosphatidic acids, which are found in the form of metallic salts in various plant and animal tissues.



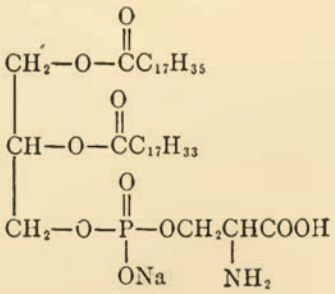
A phosphatidic acid
(calcium salt)

Cephalins

The cephalins are found closely associated with lecithins in many tissues, but particularly in brain tissue. A phospholipide preparation obtained from ether-soluble brain lipides, by precipitation with alcohol, has been called "brain cephalin." It was thought to differ from lecithin only in containing ethanolamine, or cholamine ($HOCH_2CH_2NH_2$), as the basic constituent in place of choline. However, it has been demonstrated that "brain cephalin" is actually a mixture of several phospholipides, only one of which corresponds to the above structure. This component of "brain cephalin" is now designated by the more specific name, *phosphatidyl ethanolamine*. A second component is very similar in structure but contains the amino acid serine, $HOCH_2CHCOOH$, in place of ethanolamine. It is named *phosphatidyl serine*. When obtained from brain tissue, it contains oleic and stearic acids, as the only fatty acids. The chemical formulas of these substances are as follows:



Phosphatidyl ethanolamine
(inner salt formula)



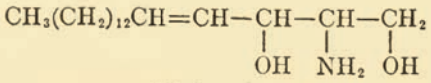
Phosphatidyl serine
(sodium salt)

Phosphatidyl serine is a strong acid, existing at physiological pH values combined as a salt with sodium or potassium. It is alcohol-insoluble, whereas phosphatidyl ethanolamine is easily soluble in alcohol. The "cephalin" of beef brain, and of pig heart and liver, is made up mostly of phosphatidyl ethanolamine (50–75 per cent), mixed with smaller amounts of phosphatidyl serine (13–35 per cent). According to Chargaff, egg yolk cephalin is entirely of the ethanolamine type. Cephalins are thought to play a part in the process of blood clotting.

Sphingomyelins

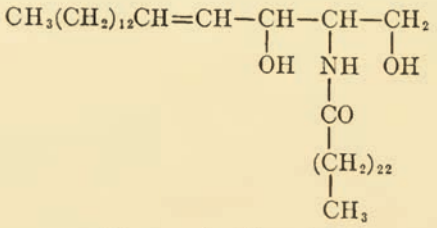
This type of phospholipide differs from most others in being insoluble in ether. It is also insoluble in acetone, but forms a colloidal solution in water. It can be extracted from tissues with hot alcohol, but on cooling crystallizes out as a white solid. First discovered in brain tissue, it is now known to occur in many organs, as well as in blood and milk. Enormously increased amounts are present in all the organs and tissues (except brain) of patients suffering from Niemann-Pick's disease.

Sphingomyelin on hydrolysis gives rise to an amino alcohol (sphingosine), plus a fatty acid (usually lignoceric), phosphoric acid, and choline. The fatty acid is attached to the amino group to form the amide, ligno-



Sphingosine

ceryl sphingosine. This substance is a normal component of pig spleen and liver and can also be obtained by partial hydrolysis of sphingomyelin.



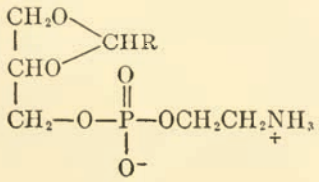
Lignoceryl sphingosine

The phosphoric acid and choline are presumably present in sphingomyelin as a phosphorylcholine, inner-salt group (see formula of lecithin, p. 97), attached to one of the hydroxyl groups of sphingosine.

Other phospholipides

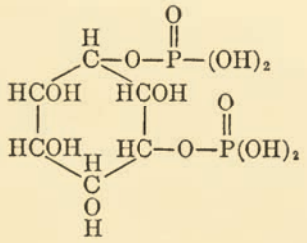
Almost nothing is known of the biological significance of the phospholipides other than lecithins and cephalins, but the fact that they occur

in large amounts in brain and nerve tissue undoubtedly means that they have an important function there. An unusual lipide type is represented by the *plasmalogens*, substances first found in beef muscle, but also present in brain. These lipides make up about 10 per cent of the total phosphatides of these tissues. Plasmalogen on hydrolysis gives rise to cholamine, α-glycerophosphoric acid, and a higher aldehyde corresponding to a common fatty acid, *e.g.*, palmitaldehyde ($C_{15}H_{31}CHO$) or stearaldehyde ($C_{17}H_{35}CHO$). Therefore, plasmalogens most probably have the following formula, in which R represents the alkyl radical of a higher fatty acid.



Plasmalogen

Phospholipides containing inositol have been found in soybeans, as well as in nerve tissue. One of the components of "brain cephalin" has been found to be an alcohol-insoluble phospholipide which on hydrolysis gives rise to equimolar portions of glycerol, a fatty acid, and inositol *meta*-diphosphate:



Inositol *meta*-diphosphate

CEREBROSIDES (GLYCOLIPIDES)

Compound lipides containing a sugar residue but no phosphorus are present in relatively large amounts (7 per cent of the solid matter) in brain tissue and probably occur in all organs of the body. These substances, called cerebrosides, are insoluble in ether and water, but dissolve in pyridine or in hot alcohol. They are white solids which on hydrolysis yield one molar equivalent each of a higher fatty acid (*e.g.*, lignoceric), sphingosine or dihydrosphingosine, and galactose. Such a combination of a carbohydrate, fatty acid, and base shows what unexpected compounds exist in the body and demonstrates that the line between carbohydrates, fats, and nitrogenous compounds is not nearly as sharp as classifications might indicate. All the fatty acids which have been found

to occur in cerebrosides and sphingomyelin are saturated, or have only one double bond, and contain a large number of carbon atoms (mostly C_{24} or C_{26}).

REVIEW QUESTIONS ON LIPIDES

1. Explain the terms: (1) fat, (2) phospholipide, (3) wax, (4) sterol, (5) soap, (6) saponification, (7) hydrogenation of oils, (8) detergent, (9) antioxidant.

2. Explain chemically why tallow is a solid and olive oil is a liquid. Name the chief fatty acids obtained from each.

3. Give (1) the elements in (a) carotene, (b) cholesterol, (2) the equation for the saponification of triolein with potassium hydroxide, (3) the chemical groups contained in lecithin, (4) the graphic formula of glycerol, (5) the formula for one constituent in a wax.

4. Discuss the factors that operate to make a fat rancid.

5. Explain the differences in the chemical composition of the ether extracts (fat) of various food materials. Compare, for example, the ether extracts of oatmeal and spinach.

6. Explain the terms (1) simple glyceride, (2) mixed glyceride, and discuss their occurrence in a natural fat.

7. Explain how an oil is made commercially. What is the source of (1) linseed oil, (2) tallow, (3) "Crisco"?

8. Explain the significance and the limitations of the term "fat" as used in tables of food analyses.

9. Why is January milk so pale in color as compared with June milk? Can you think of any way by which winter milk could be improved in color?

10. Generally compare the biological roles of true fats, waxes, and complex lipides.

11. Correct the following statements if incorrect: (1) "Dreft" is the trade name of a cleaning agent that does not form an insoluble precipitate with hard water. Chemically, it is the sodium salt of sulfated aliphatic alcohols, chiefly lauryl. (2) Hydrogenation of coconut oil will produce a solid fat. (3) Glycerine and glycerol are different names for the same compound. (4) Stearin and sterol also mean the same thing. (5) Linseed oil is obtained from the linen seed.

REFERENCES AND SUGGESTED READINGS

Bloor, W. R., *Biochemistry of the Fatty Acids,* Reinhold Publishing Corporation, New York, 1943.

Bull, H. B., *Biochemistry of the Lipides,* John Wiley and Sons, Inc., New York, 1947.

Chargaff, E., Ziff, M., and Rittenberg, D., "A Study of the Nitrogenous Constituents of Tissue Phosphatides," *J. Biol. Chem.,* **144**, 343 (1942).

Hilditch, T. P., *Chemical Constitution of Natural Fats,* John Wiley and Sons, Inc., New York, 1940.

Jamieson, G. S., *Vegetable Fats and Oils,* Reinhold Publishing Corporation, New York, 1932.

Markley, K. S., *Fatty Acids,* Interscience Publishers, Inc., New York, 1947.

Thannhauser, S. J., and Schmidt, G., "Lipins and Lipidoses," *Physiol. Rev.,* **26**, 275 (1946).

Wittcoff, H., *The Phosphatides,* Reinhold Publishing Corp., New York, 1951.

Chapter 5

PROTEINS

Proteins are organic compounds containing nitrogen, which, together with the carbohydrates and the fats, form the principal part of the solids of living matter. The name comes from a Greek word, *proteios,* meaning first. It was originated in 1839 by a Dutch chemist, Mulder, because of his belief in the widespread occurrence and great importance of the proteins. In nutrition, the proteins are used principally for body-building and maintenance rather than for providing energy. Protein materials are derived from both plant and animal sources, although in the average American diet animal proteins appear to predominate.

Like the polysaccharides and fats, proteins can be broken down into their unit structures. These units are amino acids, the "building stones"

Table 5–1

Economic importance of some industries based on proteins *

Industry	*Wage earners*	*Value of products shipped*
1. Leather and its products	383,175	$3,673,849,000
2. Meat products, including poultry	274,441	8,766,322,000
3. Wool, felt, and hair products	230,524	2,432,355,000
4. Furs and their products	37,561	564,660,000
5. Sea foods (canned)	20,153	226,519,000
6. Glue and gelatin	5,372	99,260,000
	951,226	$15,762,965,000

* Compiled from the 1947 *Census of Manufactures,* Bureau of the Census, 1950, and from the *Statistical Abstract of the United States,* 1951, published by the Department of Commerce.

from which the proteins are constructed. In the common proteins there are about twenty different amino acids; many hundred molecules of these amino acids are combined to form the larger aggregates called proteins. Smaller aggregates of the order of 100 units or less are usually classed as peptides.

Although proteins do not play as large a role in our economic life as carbohydrates, their importance is very great. They are the distinguishing constituents of many essential foods (*e.g.,* meats, fish, poultry, eggs,

beans, peas, etc.) as well as important clothing and furnishing materials (*e.g.*, woolens, felts, furs, silks, leather, etc.). Table 5–1 gives data for some industries utilizing protein materials.

Comparison with Table 3–1 (p. 103) shows that the protein industries approach the machinery industry in monetary value of products, though they do not give employment to as many workers. Since Table 5–1 is not exhaustive, it does not include any nonindustrial business based on protein materials; for example, eggs, which had a value in 1950 of $1,811,387,667 at the farms where they were produced.

Several pure proteins are prepared and sold—gelatin for food and film industries; insulin for the treatment of diabetes; pepsin, trypsin, and other enzymes; and the peptide antibiotics, bacitracin and tyrothricin. Other purified protein materials are marketed as vaccines, toxins, and antitoxins.

Occurrence and preparation

Every kind of cell contains its own special proteins, and, therefore, the number of individual proteins must be enormous. About 700 have been isolated and examined. Perhaps 200, contained in the most important foodstuffs and biological materials, have been studied in some detail. Since there are about 40 known amino acids, each one of which may be used many times, it is evident that an enormous number of proteins is possible. Perhaps comparison with the number of words in the English language will make the possibilities more evident. We have 26 letters, each of which may be used several times in a given word. They comprise about 600,000 words listed in an unabridged dictionary. Many new words are added yearly, just as many new proteins are discovered each year. There are probably thousands of new proteins in the many species of plants, animals, and microorganisms that have not been investigated.

The general method of preparing a protein is to dissolve it in its particular solvent, water, salt solution, or alcohol, and then alternately precipitate impurities or protein by changing one or more factors such as pH, salt concentration, or temperature. Solution and reprecipitation are repeated many times until the protein is obtained as pure as possible, and preferably in crystalline form. Some proteins, for example, egg albumin, can be obtained crystalline after only one or two operations, but others, such as muscle phosphorylase, require about ten different treatments before they will crystallize. A twentyfold purification is usual, but in the case of botulinum toxin A, purification increases the potency more than 200 times that of the crude material.

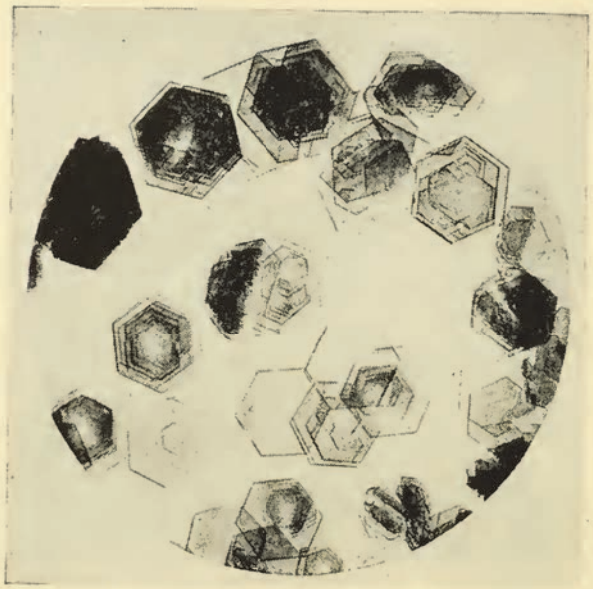
Isolation of proteins challenges the skill and resourcefulness of the most experienced investigators. Great progress, however, has been made in the last twenty years. To date, more than 150 proteins have been

From Hawk and Bergeim, *Practical Physiological Chemistry.*
Courtesy of P. Blakiston's Son & Co., Inc.

Fig. 5–1. Crystalline egg albumin.

From Hawk and Bergeim, *Practical Physiological Chemistry.*
Courtesy of P. Blakiston's Son & Co., Inc.

Fig. 5–2. Oxyhemoglobin of the horse.



From Reichert and Brown, *The Crystallography of Hemoglobins.* Courtesy of the Carnegie Institution of Washington.

Fig. 5–3. Oxyhemoglobin of the squirrel.

From Reichert and Brown, *The Crystallography of Hemoglobins.* Courtesy of the Carnegie Institution of Washington.

Fig. 5–4. Oxyhemoglobin of the guinea pig.

obtained in a crystalline state. Crystallinity, however, is no sure sign of homogeneity (purity). The best tests of homogeneity are electrophoresis, ultracentrifuge sedimentation, solubility measurements, and maximum biological activity (if the protein has such a property).

If a solution of protein is placed in an electrophoresis cell, and an electric current is passed through the solution, the protein moves toward the cathode or anode, depending upon the polarity of the electric charge carried by the protein. The position of the protein in the solution is marked by a change in the refractive index of the solution and shows up as a boundary which can be determined by suitable optical means. The migration of the protein, during a given period of time, can be followed and its electrophoretic mobility calculated. If the protein consists of only one component, there will be only a single boundary in the solution. The protein is then said to be pure or homogeneous. If the solution contains a second component, this will usually migrate at a different rate of speed than the first so that it will be revealed in the optical pattern. Electrophoresis can be used to determine not only the number of components but also the amount of each component in the solution. An example of its usefulness is found in the extensive study being made of the proteins of blood in health and disease.

In the ultracentrifuge, the protein solution is subjected to a centrifugal force about 250,000 times as great as the force of gravity. Heavier particles settle faster than lighter ones. If the solution contains two or more proteins, there will be either two or more boundaries, or else a diffuse boundary in the optical pattern. A single protein in solution shows only one boundary when it is sedimented.

The solubility test of homogeneity depends upon the fact that the quantity of dissolved protein, until saturation is reached, is directly proportional to the weight of sample taken. After saturation there is no increase in dissolved protein, irrespective of the weight of sample. However, if the sample contains more than one protein, the solubility curve does not show a sharp break but continues to rise until the saturation point is reached for each component.

Few of the proteins isolated have been subjected to all of these tests. Most of the proteins, *e.g.*, casein, gliadin, that have been prepared from our common foodstuffs appear to be mixtures of several components.

The names and amounts of the principal proteins in some common foodstuffs are given in Table 5–2. The percentages given are, in most cases, the amounts actually isolated. The totals, in general, agree well with the figures for crude protein ($N \times 6.25$), and hence, it may be assumed that the kinds and amounts of protein in many of our staple foods are well established.

From Table 5–3 it is apparent that proteins have great importance other than as food constituents. Some of our most important articles

Table 5–2

Principal proteins of some common foodstuffs

(Undried basis)

Foodstuff	*Proteins*	*Classification* *	*Percentage in foodstuff*
Barley	Hordein	Alcohol-soluble	4.0
	Hordenin	Glutelin	4.5
	Leucosin	Albumin	0.3
	Edestin	Globulin	1.95
	Proteose	Proteose	
		Total	**10.75**
Beans	Phaseolin	Globulin	21.5
	Phaselin	Globulin	2.0
		Total	**23.5**
Blood	Serum albumins	Albumin	4.0
	Serum globulins	Globulin	3.3
	Fibrinogen	Globulin	0.3
	Hemoglobin	Chromoprotein	14.0
		Total	**21.6**
Corn	Zein	Alcohol-soluble	5.0
	Zeanin	Glutelin	3.15
	Maysin	Globulin	0.25
	Globulin	Globulin	0.14
	Proteose	Proteose	0.06
		Total	**8.60**
Eggs	Egg albumin	Albumin	8.2
	Vitellin	Phosphoprotein	5.2
		Total	**13.4**
Lean meat	Myogen	Albumin	2.0
	Myoalbumin	Albumin	0.2
	Myosin	Globulin	13.6
	Globulin	Globulin	4.2
		Total	**20.0**
Milk	Casein	Phosphoprotein	2.74
	Lactalbumin	Albumin	0.60
	Globulin	Globulin	Trace
		Total	**3.34**
Oats	Avenalin	Globulin	1.5
	Glutelin	Glutelin	11.2
	Prolamin	Alcohol-soluble	1.3
		Total	**14.0**

* See p. 110 for basis of classification.

Table 5–2—*(Continued)*

Foodstuff	*Proteins*	*Classification**	*Percentage in foodstuff*
Peas	Legumilin	Albumin	2.0
	Legumin	Globulin	10.0
	Vicillin	Globulin	
	Proteose	Proteose	1.0
	Undetermined	...	11.5
		Total	**24.5**
Rice	Oryzenin	Glutelin	4.0
	Globulin	Globulin	5.9
	Alc. sol.	Alc. sol.	
		Total	**9.9**
Rye	Gliadin	Alcohol-soluble	4.0
	Secalenin	Glutelin	2.44
	Edestin	Globulin	1.76
	Proteose	Proteose	
	Leucosin	Albumin	0.43
		Total	**8.63**
Wheat	Gliadin	Alcohol-soluble	3.91
	Glutenin	Glutelin	4.17
	Globulin	Globulin	0.63
	Leucosin	Albumin	0.36
	Proteose	Proteose	0.43
	Miscellaneous	...	1.10
		Total	**10.60**

of clothing are protein in character. Probably all enzymes and some hormones are proteins. The discovery of the existence of proteins as viruses, toxins, and poisons, deadly to man, has emphasized their vital importance and greatly increased the impetus to research.

Elementary composition

All proteins contain carbon, hydrogen, oxygen, and nitrogen. Most proteins contain sulfur, some contain phosphorus, and a few contain iron, copper, or manganese. Phosphorus is not really a constituent of the protein part of the molecule. Only the conjugated proteins—proteins united with a prosthetic (additional) group—contain phosphorus, the phosphorus being in the form of a phosphoric acid ester. Iron, likewise, is not found in any amino acid. Hence it is not in the protein molecule proper but is contained in the prosthetic group, *e.g.*, hematin in hemoglobin.

Table 5–3

Examples of enzyme proteins and nonfood proteins

Protein	*Classification*	*Source*
Keratin	Albuminoid	Hair, wool, feathers, hide, nails, etc.
Fibroin	Albuminoid	Silk
Sericin	Albuminoid	Silk
Spongin	Albuminoid	Sponge
Amylopsin	Albumin	Pancreas
Pepsin	Albumin	Stomach
Trypsin	Albumin	Pancreas
Papain	Albumin	Latex of papaya tree
Urease	Globulin?	Jack bean
Catalase	Chromoprotein	Liver, etc.
Flavoprotein	Chromoprotein	Yeast, heart, etc.
Insulin	Globulin?	Pancreas
Mosaic virus	Nucleoprotein	Diseased tobacco plants
Tuberculin	Unclassified	Tubercle bacillus
Avidin	Albumin?	Egg white
Crotoxin	Albumin	Rattlesnake venom
Bacitracin	Polypeptide	Antibiotic from *Bacillus licheniformis*
Botulinum toxin A	Globulin	Toxin from *Cl. botulinum*

The elementary composition of proteins varies within wide limits, but the average figures show C, 53 per cent; H, 7 per cent; O, 23 per cent; N, 16 per cent and S, 1 per cent. Nitrogen shows the greatest variation, ranging from about 10 per cent for the glycoproteins to 30 per cent for the protamines. In the common food proteins it varies within much narrower limits, 15.5–18.7 per cent, and, hence, an average figure of 16 per cent is taken. On the other hand, proteins from entirely different sources, and different in character, may have the same elementary composition. Except in a very general way elementary composition is of but little value in the differentiation of proteins.

CLASSIFICATION

Because of their number and complexity, proteins are difficult to classify. The basis for classification is mainly (1) products on hydrolysis, (2) solubility, (3) coagulability, and (4) precipitability. It is probable that many compounds quite unlike in structure fall into the same group, and it is certain that many so-called proteins are not chemical entities. Even when crystallized, the proteins are not always homogeneous. However, even though imperfect, a classification is indispensable for study and discussion. The official classification, with slight modifications, follows. For examples of many of the classes see Tables 5–2 and 5–3.

Simple proteins

These are naturally occurring proteins that on hydrolysis yield only α-amino acids. This definition is not strictly correct because many of the albumins contain small quantities (1–2 per cent) of carbohydrate, *e.g.*, mannose, galactose. Simple proteins are subdivided as follows:

Albumins. Soluble in pure water and dilute salt solutions, coagulable by heat, precipitated by saturation with ammonium sulfate.

Globulins. Insoluble in pure water but soluble in neutral salt solutions (*e.g.*, 5 per cent NaCl), coagulable by heat, precipitated by half-saturation with ammonium sulfate.

Glutelins. Insoluble in water or salt solution, but soluble in dilute acids or alkalies (*e.g.*, 0.1 per cent).

Prolamins (Alcohol-soluble Proteins). Insoluble in water, dilute salt solutions, or absolute alcohol, but soluble in 70–80 per cent alcohol.

Albuminoids. Insoluble in the reagents given for the preceding proteins. A heterogeneous group of simple proteins found usually in the skeletal structures and protective coatings of animals; examples are keratin from horn, hide, hoof, hair, feathers, and wool, elastin from ligaments, collagen from hide and tendons, and fibroin and sericin from silk. Gelatin, although it does not fit into this group, is classed as an albuminoid because it is obtained from collagen by boiling with water. It is more properly a derived protein.

Histones. Proteins (having basic properties) coagulable by heat, soluble in water, dilute acids, or alkalis, but insoluble in dilute ammonia. They form precipitates with other proteins and yield on hydrolysis large quantities of the basic amino acids. Typical examples are globin from hemoglobin and histones from the thymus gland and leucocytes.

Protamines. These are the simplest natural proteins and contain only a small number of amino acids, among which arginine predominates—in some cases comprising 85 per cent or more of the protein. The protamines are strongly basic, soluble in water, and not coagulable by heat. They form crystalline salts with mineral acids and precipitates with other proteins, *e.g.*, insulin (protamine-insulin). They are found in ripe sperm cells. The most studied compounds have been obtained from fish sperm, *e.g.*, salmine from salmon, sturine from sturgeon, and clupeine from herring.

Conjugated proteins

These proteins are combinations in which a simple protein is united with a characteristic nonprotein group. The nonprotein group is called a prosthetic group. The subdivisions are as follows:

Nucleoproteins. These are basic proteins, *e.g.*, histones, in combination with nucleic acids. They are obtained most readily from thymus and other glands, yeast, and wheat germ.

Glycoproteins or Glucoproteins. The prosthetic group is carbohydrate in character, and makes up a large proportion (25–35 per cent) of the glycoprotein. As a consequence, the nitrogen content of glycoproteins is low, 9–13 per cent. Examples are: proteins from saliva (mucin), vitreous humor, gastric mucosa, and jellyfish.

Phosphoproteins. The prosthetic group is phosphoric acid, linked as an ester to the protein through a hydroxyamino acid, *e.g.*, serine. Two of the best known proteins, casein and vitellin, belong to this class.

Chromoproteins. The prosthetic group is colored. For example: hematin of hemoglobin is red, cyanin of hemocyanin (the respiratory pigment in the blood of the lobster and other molluscs) is blue, melanin of hair proteins is black, riboflavin phosphate of flavoproteins (respiratory enzymes) is red, and retinene (aldehyde of vitamin A) of rhodopsin (the chromoprotein involved in vision) is yellow.

Lipoproteins. The prosthetic group is fatty acid, lecithin, or a phospholipide other than lecithin. Lipoproteins are a poorly defined group occurring in egg yolk, brain tissue, lungs, etc.

Derived proteins

This division includes denatured proteins and cleavage products formed by partial hydrolysis of naturally occurring proteins with acids or enzymes. The hydrolysis products are polypetides of varying size, and the subdivisions have little chemical justification. The principal classes still in use are:

Proteoses. Hydrolytic products of proteins that are soluble in water, not coagulable by heat, and precipitated by saturating the solution containing them with ammonium sulfate.

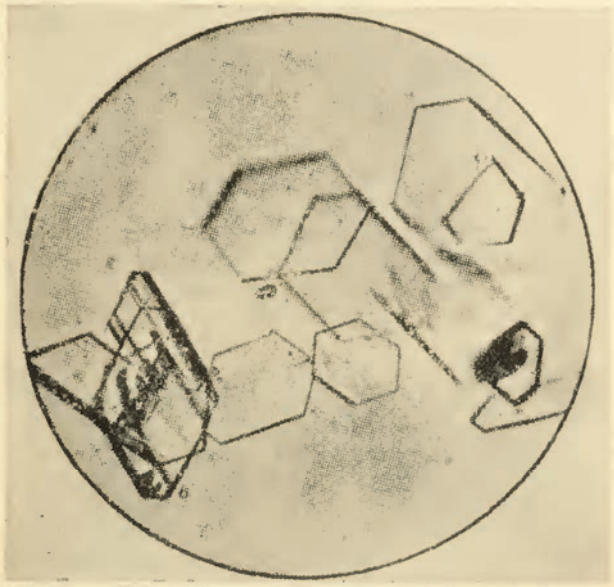
Peptones. These polypeptides are probably of smaller molecular weight than the proteoses since they are found in the filtrate of the ammonium sulfate precipitation of proteoses. So-called "peptones," used in the preparation of bacteriological media, are mixtures of polypeptides, mainly proteoses.

Peptides. Combinations of two or more amino acids. They are called di-, tri-, tetra-, etc., peptides, according to whether they contain two, three, four, or more amino acid residues in the molecule. Some peptides occur naturally, *e.g.*, glutathione, pteroylglutamic acid (a vitamin), penicillins, etc. Large numbers of peptides have been synthesized.

PRODUCTS ON HYDROLYSIS—AMINO ACIDS

Number and kind of amino acids

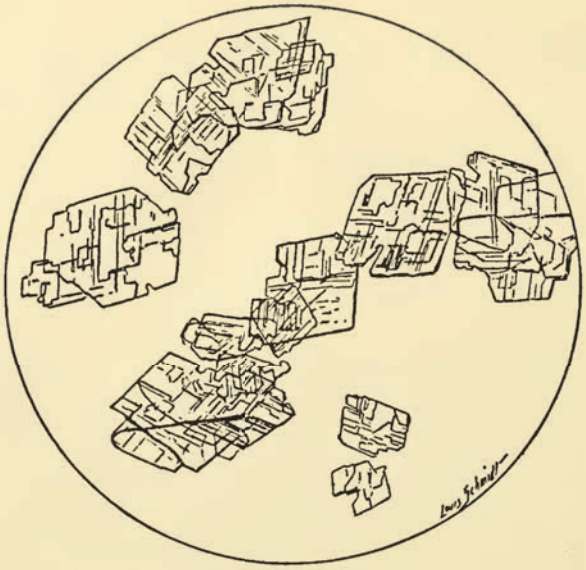
When a protein is hydrolyzed by means of acid, alkali, or enzymes, alpha amino acids, usually called amino acids, are obtained as products. If the protein is a conjugated protein, such as casein, nucleoprotein, etc., other products are obtained in addition. Only the amino acids, however, come from the protein molecule proper; the additional com-



Courtesy of W. C. Rose and *Journal of Biological Chemistry.*

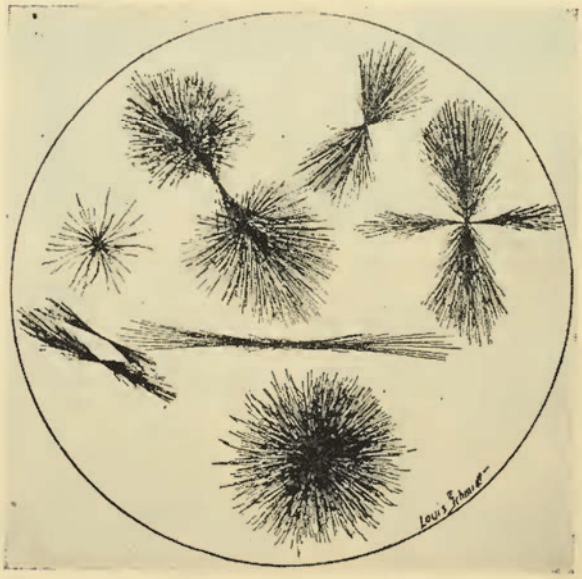
Fig. 5-5. Threonine.

pounds originate in the prosthetic group. The number of amino acids that have been definitely obtained as a result of hydrolysis is about twenty-five; about fifteen more have been reported, but their presence in protein is either not well established or they occur only in unusual protein materials, *e.g.*, djencolic acid of the djencol bean and α,γ-diaminobutyric acid in the polymyxin antibiotics (pp. 118, 119). Several others, citrulline, homocysteine, and cystathionine, do occur free in the body tissues and play an important role in intermediary metabolism. Amino acids are classified according to the number of amino groups, carboxyl groups, and other characteristics as shown on pp. 116–120.



From Hawk and Bergeim, *Practical Physiological Chemistry.*
Courtesy of P. Blakiston's Son & Co., Inc.

Fig. 5–6. Phenylalanine.

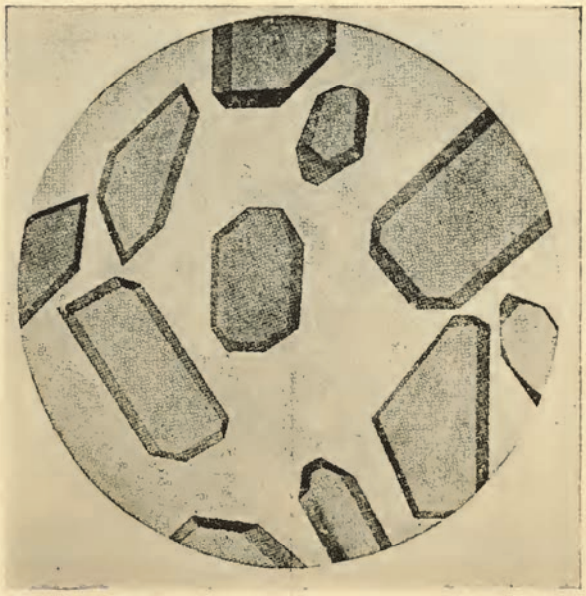


From Hawk and Bergeim, *Practical Physiological Chemistry.*
Courtesy of P. Blakiston's Son & Co., Inc.

Fig. 5–7. Tyrosine.

From Schmidt's *Chemistry of the Amino Acids and Proteins.* Courtesy of Charles C. Thomas, publisher, Springfield, Illinois.

Fig. 5–8. Cystine.



From Hawk and Bergeim, *Practical Physiological Chemistry.* Courtesy of P. Blakiston's Son & Co., Inc.

Fig. 5–9. Histidine.

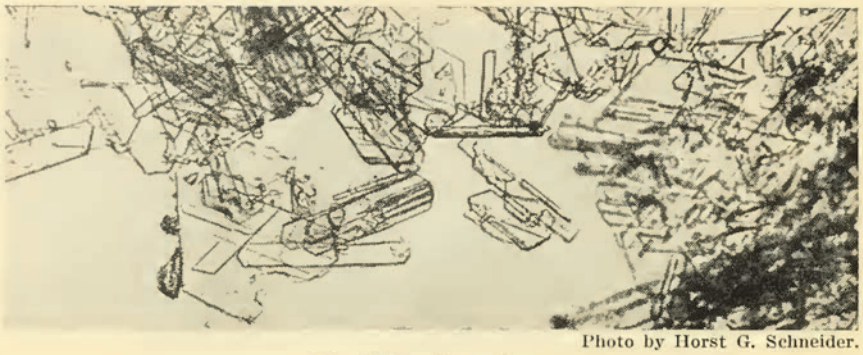


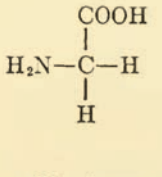
Photo by Horst G. Schneider.

Fig. 5–10. Trytophan.

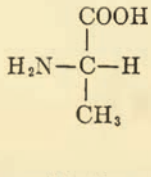
Classification and formulas of amino acids

I. MONOAMINO-MONOCARBOXYLIC ACIDS

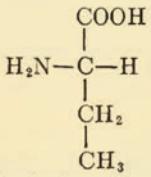
A. *Aliphatic acids:*



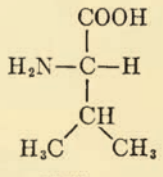
L-Glycine or aminoacetic acid



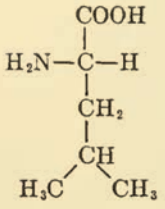
L-Alanine or α-aminopropionic acid



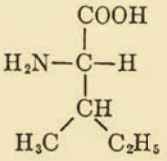
L-Aminobutyric acid



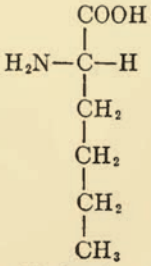
L-Valine or α-aminoisovaleric acid



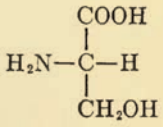
L-Leucine or α-aminoisocaproic acid



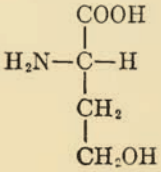
L-Isoleucine or β-methyl-α-amino-valeric acid



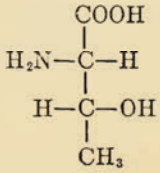
L-Norleucine or α-aminocaproic acid



L-Serine or α-amino-β-hydroxy-propionic acid

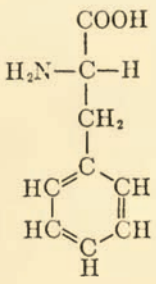


L-Homoserine or α-amino-γ-hydroxy-butyric acid

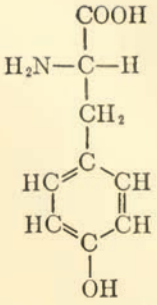


L-Threonine or α-amino-β-hydroxy-butyric acid

B. *Acids with aromatic nuclei:*

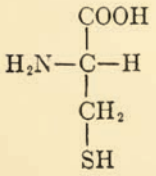


L-Phenylalanine

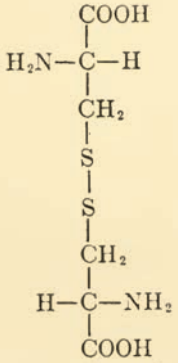


L-Tyrosine or
p-hydroxyphenylalanine

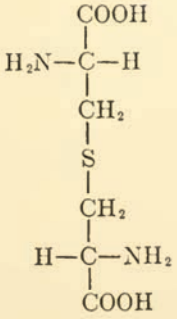
C. *Acids containing sulfur:*



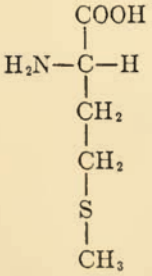
L-Cysteine or
β-thiolalanine



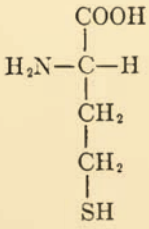
L-Cystine (dicysteine) or
di-β-thiolalanine



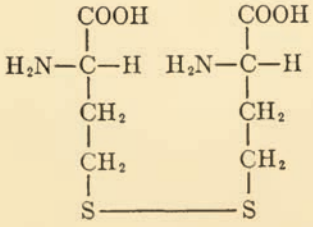
L-Lanthionine



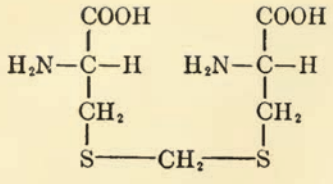
L-Methionine or
α-amino-γ-methylthiol-
butyric acid



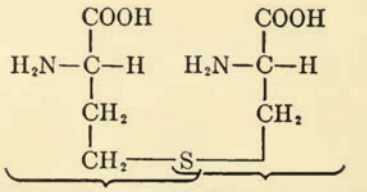
L-Homocysteine



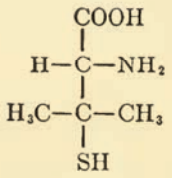
L-Homocystine
(dihomocysteine)



L-Djencolic acid

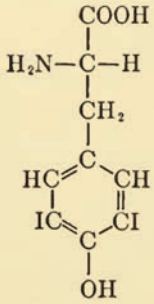


(Homocysteine part) (Cysteine part)
L-Cystathionine

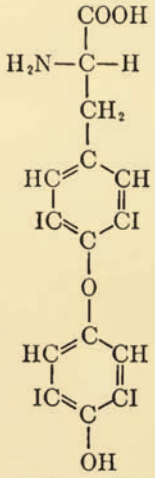


D-Penicillamine
(β-thiolvaline)

D. *Acids containing iodine:*

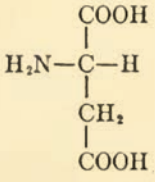


L-Diiodotyrosine

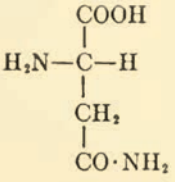


L-Thyroxine

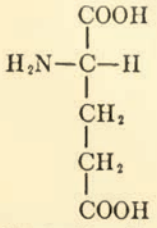
II. Monoamino-dicarboxylic acids and related amides



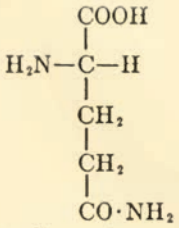
L-Aspartic acid or α-aminosuccinic acid



L-Asparagine (amide of aspartic acid)

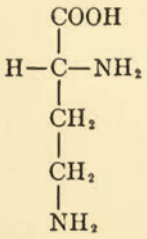


L-Glutamic acid or α-aminoglutaric acid

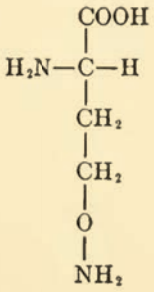


L-Glutamine (amide of glutamic acid)

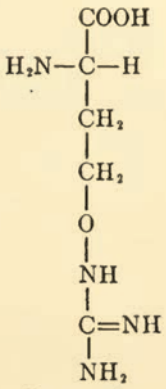
III. Basic amino acids



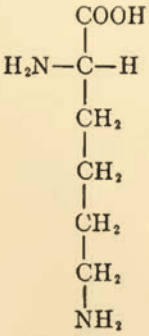
D-α, γ-Diaminobutyric acid



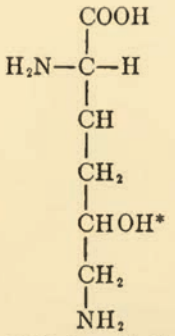
L-Canaline



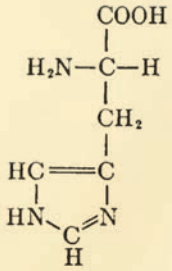
L-Canavanine



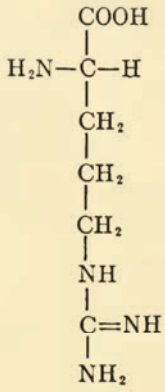
L-Lysine or α, ε-diaminocaproic acid



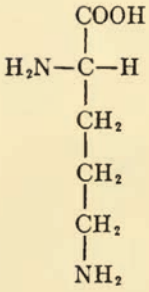
L-Hydroxylysine
*(position of OH, whether D or L, is not known)



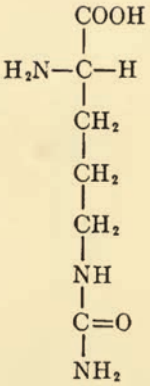
L-Histidine or
β-imidazolealanine



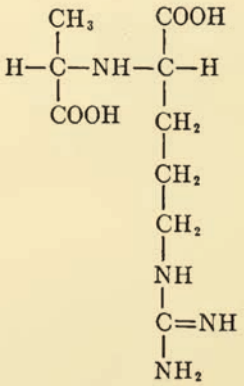
L-Arginine or
∂-guanidino-α-aminovaleric acid



L-Ornithine

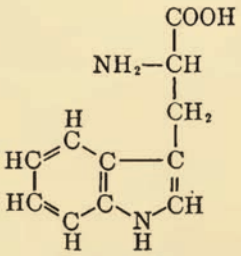


L-Citrulline

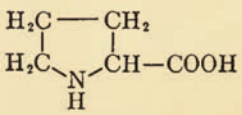


L-Octapine

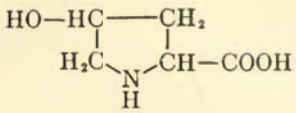
IV. Heterocylic amino acids



L-Tryptophan or
indolealanine



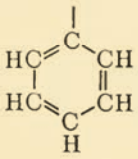
L-Proline or
α-pyrrolidine-carboxylic acid



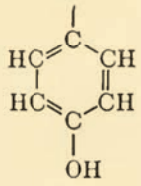
L-Hydroxyproline

Every amino acid contains at least one amino (NH_2) group and one carboxyl (COOH) group. (Proline and hydroxyproline may be regarded as modified amino acids, in which the amino group has been linked to a second carbon, thus forming a ring compound.) One of the amino groups in the above list is always attached to the alpha carbon, hence the name alpha amino acids. Acids with the amino group attached to carbons other than the alpha carbon occur in nature (*e.g.*, beta-alanine in pantothenic acid, beta-lysine in certain antibiotics, and gamma-aminobutyric acid in biological fluids). Other types of amino acids will probably be found as more plants and microorganisms are investigated.

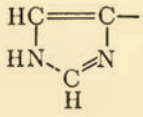
The part of the formula other than $NH_2CH{\cdot}COOH$ is called the side chain and is represented by the letter R. Amino acids differ with respect to their side chains, and, hence, it is this part of the molecule that imparts distinctive features to the compound. Since the most important chemical and physiological properties of amino acids are attributable to the side chains, the student should note these carefully and become familiar with the groups contained therein. Some of these distinguishing groups are:



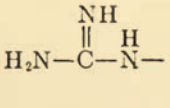
Phenyl



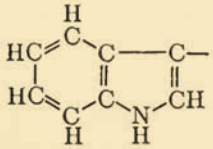
Phenol



Imidazole

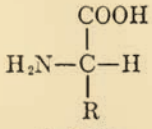


Guanidino

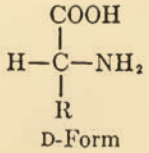


Indole

It is evident that the alpha carbon is asymmetric in all of the amino acids except glycine; hence there are two structural forms of the acids. The two forms may be represented by the general formulas where R



L-Form



D-Form

denotes the remainder of the formula, *e.g.*, CH_3 for alanine, etc. Formerly the small letters *d* and *l* were used to denote the two forms, but since these letters are also used to indicate optical rotation, the American and British chemical societies have adopted SMALL CAPITAL LETTERS to show configuration. The reference compounds, D and L serine, correspond to those used in sugar chemistry, *viz.*, D and L glyceraldehyde. Optical rotation is indicated by the words *dextro* and *levo*, or by plus (+) and minus (—) signs.

Most of the amino acids are of the L type, but small amounts of the D acids have been reported in several common proteins. Cancerous tissue has been reported to be much higher in D-glutamic acid than normal tissue, but the weight of evidence seems to be against this conclusion. Several antibiotics, *e.g.*, gramicidin, tyrocidine and actinomycin, contain large amounts of D-phenylalanine, D-leucine, and D-valine. Other antibiotics, *e.g.*, penicillin and polymyxin, contain the previously unknown amino acids, D-penicillamine and D-α,γ-diaminobutyric acid, respectively. The presence of unusual structures of amino acids, sugars, etc., seems to be a general characteristic of antibiotics.

Some of the amino acids listed have not yet been found in proteins. For a long time a number of these were assumed not to occur in proteins, but recently several have been found in polypeptides, *e.g.*, ornithine in gramicidin and lanthionine in subtilin (an antibiotic). Homocysteine, as yet unreported in any protein, is an intermediary product in the conversion of methionine to cysteine. Cystathionine is also an intermediate in this conversion. The steps are

$$\text{methionine} \longrightarrow \text{homocysteine} \xrightarrow{+\ \text{serine}} \text{cystathionine}$$
$$\longrightarrow \text{cysteine and homoserine (in neurospora) or}$$
$$\alpha\text{-ketobutyric acid (in animals)}$$

Amino acid composition of proteins

To determine the amino acid content of a protein, it must first be hydrolyzed. This is usually done with about twenty per cent hydrochloric acid at 100°C. for 10 to 20 hours. The acid is then removed, and the amount of each amino acid is determined quantitatively. This analysis is one of the most difficult tasks in analytical chemistry, having defied the efforts of some of the world's ablest chemists for the past seventy-five years. Only within the last decade have methods been perfected to such an extent that all the nitrogen or sulfur in a protein can be accounted for in the amino acid figures. The most promising methods at present seem to be microbiological and chromatographic. Bacteria are most widely used in microbiological assays, and the procedures are the same as those used in analyzing for vitamins (p. 234). Chromatographic pro-

cedures are based on differences in the degree of adsorption of amino acids by solids, in the reaction of amino acids with ion exchange materials, and in the solubility of amino acids in organic solvents. For details regarding the operation of these methods, the original papers of Moore and Stein and the book by Block and Bolling should be consulted.

Table 5–4 gives the percentages of amino acids found in some typical proteins. The table includes proteins representative of foods, enzymes, hormones, viruses, antibiotics, and fibers. In a given protein the figures vary greatly. In egg albumin, for example, there is about 14 times as much glutamic acid as there is tryptophan. In another common food protein, gliadin, the glutamic acid exceeds the tryptophan more than one hundredfold.

If all the proteins are considered, glutamic and aspartic acids and leucine are seen to be the most abundant amino acids, while tryptophan, histidine, and methionine are least abundant. The amino acids present in largest amounts are those most closely related to the intermediary compounds of carbohydrate metabolism, *e.g.*, glutamic acid and α-ketoglutaric acid form a pair, and aspartic acid and oxalacetic acid make up a second pair (p. 331). The least abundant amino acids such as tryptophan and methionine are the most complex in structure, probably involving also the largest number of steps in synthesis.

Proteins that contain considerable amounts of all of the amino acids, *e.g.*, egg albumin and casein, are called "complete"; those that are lacking or very high in certain amino acids, *e.g.*, gelatin and zein, are said to be "incomplete." A better term to denote the uneven composition of proteins is disproportionate.

Certain highly specialized proteins such as fibroin and salmine are conspicuously disproportionate in make-up. However, other proteins such as pepsin, insulin, and botulinum toxin A having marked biological properties show no unusual features in amino acid content. Their biological properties must be related to the structure of the molecule as a whole and not to the kind or amount of amino acids that are found in the molecule.

In most cases the sum of the figures for the amino acids amounts to more than 100 per cent. Because of the water taken up in hydrolysis, the total should be about 115 per cent of the starting material. With the improved methods now available, the total nitrogen of the protein can usually be accounted for in the individual amino acids. Likewise, the extent of carbon and sulfur recoveries are useful criteria in judging the validity of the analytical data.

From a practical viewpoint, data on the amino acid composition of foods are more useful than those on the amino acid content of individual proteins. Such data are gradually becoming available, and figures for some of our staple foods are given in Table 5–5. However, we need many

Table 5–4

Amino acid content of some typical proteins *

(Grams of amino acid per 100 g. of protein)

Amino acids	*Albumin (egg)* [1]	*Albumin (serum of human)* [1]	*Bacitracin* [2]	*Botulinum toxin A* [1]	*Casein (cow)* [3]	*Casein (human)* [3]	*Chymotrypsinogen* [1]	*Fibrinogen* [1]	*Fibroin (silk)* [1]	*Gelatin* [1]
Alanine	6.7			3.9	3.5	2.0		3.7	29.7	9.3
Arginine	5.7	6.2		4.6	4.0	3.6	2.8	7.8	1.1	8.6
Aspartic acid	9.3	10.4	17	20.1	7.2	4.6	11.3	13.1	2.8	5.6
Cystine	1.9	6.3	14	0.8	0.3	0.6	4.6	2.7		
Glutamic acid	16.5	17.4	10	15.6	22.0	20.9	9.0	14.5	2.2	11.2
Glycine	3.1	1.6		1.4	1.9	0.0	5.3	5.6	43.6	26.9
Histidine	2.4	3.5	10	1.0	3.2	2.0	1.2	2.6	0.4	0.7
Hydroxyproline		0.0			2.3					14.5
Isoleucine	7.0	1.7	22	11.9	7.6	6.3	5.7	4.8	1.1	5.2
Leucine	9.2	11.2	9	10.3	10.3	12.2	10.4	7.1	0.9	
Lysine	6.3	12.3	9	7.7	8.2	5.6	8.0	9.2	0.7	4.6
Methionine	5.2	1.3		1.1	3.1	2.7	1.2	2.6		0.9
Phenylalanine	7.7	7.8	11	1.2	5.5	5.8	3.6	4.6	3.4	2.6
Proline	3.6	5.1		2.6	11.6	8.9	5.9	5.7	0.7	14.8
Serine	8.2	3.7		4.4	5.9	5.4	11.4	7.0	16.2	3.2
Threonine	4.0	5.0		8.5	4.5	4.5	11.4	6.1	1.6	2.2
Tryptophan	1.2	0.2		1.9	1.2	1.0	5.6	3.3		
Tyrosine	3.7	4.7		13.5	6.1	6.1	3.0	5.5	12.8	1.0
Valine	7.1	7.7		5.3	7.2	5.0	10.1	4.1	3.6	3.3
Total:	108.8	106.1	102	115.8	115.6	97.2	110.5	110.0	120.8	114.6

Table 5–4 (continued)

Amino acids	*Gliadin (wheat)*[1]	*Gliadin (rye)*[4]	*γ-Globulin*[1]	*Glutenin*[4]	*Hemoglobin (horse)*[1]	*Hordein*[4]	*Insulin*[1]	*Keratin (wool)*[1]	*Lactogenic hormone*[5]	*β-Lactoglobulin*[1]	*Legumin*[4]	*Myogen A (Aldolase)*[1]
Alanine	2.1	1.4		6.2	7.4	0.5	4.5	4.1		6.2	2.1	8.6
Arginine	2.7	2.2	4.8	4.7	3.7	2.9	3.1	10.4	8.6	2.9	11.7	6.3
Aspartic acid	1.3	0.3	8.8	2.0	10.6		6.8	6.6	11.6	11.4	5.3	9.7
Cystine	2.6		3.1	1.8	1.0	1.6	12.5	11.9	3.1	3.4	0.9	1.1
Glutamic acid	45.7	38.0	11.8	26.5	8.5	43.2	18.6	14.1	14.1	19.5	17.0	11.4
Glycine		0.2	4.2	0.9	5.6	0.0	4.3	6.5	4.0	1.4	0.4	5.6
Histidine	1.8	0.4	2.5	1.8	8.7	2.1	4.9	1.1	4.5	1.6	1.7	4.2
Hydroxyproline					0.0			6.5		0.0		
Isoleucine	11.9	6.3	2.7	6.3	0.0	5.7	2.8	11.3	7.2	8.4	8.0	7.9
Leucine			9.3		15.4		13.2		12.5	15.6		11.5
Lysine	0.7		8.1	1.9	8.5	1.0	2.5	2.8	5.3	11.4	5.0	9.5
Methionine	1.7		1.1		1.0			0.7	4.3	3.2		1.2
Phenylalanine	6.4	2.7	4.6	2.8	7.7	5.0	8.1	3.7	4.1	3.5	3.8	3.1
Proline	13.4	9.8	8.1	6.2	3.9	13.7	2.5	9.5	6.2	4.1	3.2	5.7
Serine	4.9	1.2	11.4	0.7	5.8	1.7	5.2	10.0	6.5	5.0	0.5	7.3
Threonine	2.1		8.4		4.4		2.1	6.4	4.8	5.9		7.5
Tryptophan	0.6		2.9	2.1	1.7	1.1		1.8	1.2	1.9	1.8	2.3
Tyrosine	3.2		6.8	5.4	3.0		13.0	4.7	4.7	3.8	1.6	5.3
Valine	2.7		9.7	1.0	9.1	0.2	7.8	4.6	5.9	5.8		7.4
Total:	103.8	62.5	108.3	70.3	106.0	78.7	111.9	116.7	108.6	115.0	63.0	115.6

Table 5–4 (continued)

Amino acids	*Myosin*[1]	*Ovovitellin*[6]	*Pepsin*[1]	*Phascolin*[4]	*Ribonuclease*[1]	*Salmine*[1]	*Tobacco mosaic virus*[7]	*Zein*[1]
Alanine	6.5	4.0		1.8		1.1	5.1	10.5
Arginine	7.4	8.4	1.0	4.9	5.2	85.2	9.8	1.7
Aspartic acid	8.9	8.1	16.0	5.2	14.2		13.5	4.6
Cystine	1.4	1.5	2.1	0.2	7.1		0.7	0.8
Glutamic acid	22.1	11.0	11.9	14.5	13.0		11.3	26.9
Glycine	1.9	2.8	6.4	0.6	1.3	2.9	1.9	0.0
Histidine	2.4	3.0	0.9	2.6	4.2		0.0	1.3
Isoleucine	15.6	5.3	10.8		3.1		6.6	22.5
Leucine		8.6	10.4	9.6	0.0	1.6	9.3	
Lysine	11.9	6.9	0.9	4.6	10.4		1.5	0.0
Methionine	3.4	2.8	1.7		4.4		0.0	2.4
Phenylalanine	4.3	4.0	6.4	3.2	3.6		8.4	5.9
Proline	1.9	4.4	5.0	2.8	3.6	5.8	5.8	10.5
Serine	4.3	11.2	12.2	0.4	12.0	9.1	7.2	7.1
Threonine	5.1	4.7	9.6		9.0		9.9	3.5
Tryptophan	0.8	1.1	2.4	0.9	0.0		2.1	0.1
Tyrosine	3.4	3.8	8.5	2.8	7.9		3.8	5.3
Valine	2.6	6.2	7.1	1.0	7.3	3.1	9.2	3.5
Total:	103.9	97.8	113.3	55.1	106.3	108.8	106.1	106.6

* The source of the figures is indicated by the numbers 1 to 7 in the block at the head of each column. In a few cases, *e.g.*, hydroxyproline, the figures are from various other sources. The numbers refer to the following sources:

[1] Tristram, G. R., *Advances in Protein Chemistry*, **5**, 83 (1949) Academic Press, Inc., New York.

[2] Barry, G. T., Gregory, J. C., and Craig, L. C., *J. Biol. Chem.*, **175**, 485 (1948).

[3] McMeekin, T. L., and Polis, B. D., *Advances in Protein Chemistry*, **5**, 201 (1949) Academic Press, Inc., New York.

[4] Sherman, H. C., *Chemistry of Food and Nutrition*, 7th ed. The Macmillan Company, New York, 1946.

[5] Li, C. H., *J. Biol. Chem.*, **178**, 459 (1949).

[6] Lewis, J. C., Snell, N. S., Hirschman, D. J., and Frankel-Conrat, H., *J. Biol. Chem.*, **186**, 23 (1950).

[7] Knight, C. A., *J. Biol. Chem.* **171**, 301 (1947).

Table 5–5

Amino acid content of some foods

(Per cent on undried basis) *

Amino Acids	*Cheese*	*Corn*	*Liver*	*Milk (entire)*	*Muscle (animal)*	*Muscle (fish)*	*Oats*	*Peas and beans*
Alanine			0.94		1.34	1.18		0.18
Arginine	0.86	0.48	1.32	0.15	1.40	1.25	0.98	1.58
Aspartic acid			1.38		1.41			
Cystine	0.09	0.15	0.28	0.04	0.22	0.20	0.26	0.29
Glutamic acid			2.12	0.75	2.80			
Glycine			1.6	0.08	0.91			1.78
Histidine	0.78	0.25	0.5	0.09	0.60	0.44	0.33	0.50
Isoleucine	1.73	0.64	0.96	0.26	1.09	1.10	0.71	1.24
Leucine	2.1	1.50	1.68	0.39	1.45	1.60	1.15	1.58
Lysine	2.0	0.25	1.40	0.30	1.81	1.52	0.52	1.46
Methionine	0.83	0.31	0.64	0.11	0.60	0.54	0.29	0.45
Phenylalanine	1.52	0.50	1.22	0.19	0.91	0.74	0.79	1.12
Proline					1.09	0.51		1.55
Serine		0.85	1.48	0.15	1.09	0.68		0.95
Threonine	0.88	0.37	1.06	0.16	0.91	0.79	0.52	0.88
Tryptophan	0.38	0.06	0.30	0.05	0.25	0.20	0.19	0.18
Tyrosine	1.62	0.60	0.78	0.14	0.73	0.64	0.65	0.63
Valine	1.85	0.53	1.20	0.19	1.00	1.01	0.78	1.24

(Per cent on undried basis) *

Amino Acids	*Rice*	*Soy bean meal*	*Sweet potatoes*	*White potatoes*	*Wheat flour*	*Whole wheat*	*Whole eggs*	*Yeasts*
Alanine		1.16						
Arginine	0.54	2.56	0.06	0.10	0.41	0.56	0.87	0.59
Aspartic acid								
Cystine	0.10	0.67			0.20	0.24	0.32	0.14
Glutamic acid		6.44				3.81		1.93
Glycine					0.74		0.35	
Histidine	0.13	1.02	0.03	0.04	0.23	0.28	0.32	0.39
Isoleucine	0.39	2.10	0.07	0.07	0.45	0.53	1.01	0.79
Leucine	0.62	2.80	0.10	0.19	0.74	0.92	1.21	0.99
Lysine	0.24	2.38	0.09	0.17	0.20	0.35	0.92	0.99
Methionine	0.23	0.60	0.03	0.05	0.21	0.33	0.53	0.26
Phenylalanine	0.38	1.86	0.09	0.12	0.58	0.67	0.83	0.59
Proline		1.75						
Serine		1.47			0.46	0.56		
Threonine	0.29	1.37	0.08	0.14	0.29	0.43	0.57	0.72
Tryptophan	0.10	0.49	0.04	0.04	0.09	0.16	0.20	0.17
Tyrosine	0.43	1.40			0.40	0.53	0.59	0.47
Valine	0.46	1.86	0.11	0.11	0.44	0.56	0.95	0.76

* Calculated from data in Block and Bolling, 2nd ed.

more such data before the amino acid content of the diet can be calculated with any assurance of being correct. It is still not certain that the ratio of an individual amino acid to the total nitrogen in a food remains constant from sample to sample. In a standard product such as milk the variation is probably small, but in a food such as white potatoes large fluctuations are to be expected.

In recent work Brand has obtained a yield of 116.3 per cent for β-lactoglobulin and a recovery of 100 per cent of the nitrogen. Since all the nitrogen, and also all the surfur, in the protein was accounted for in the percentages of amino acids, it was possible to calculate the number of molecules (actually residues equivalent to the molecular weights minus the molecular weight of water) of each amino acid in the protein. For example, the number of alanine residues is calculated as follows:

The alanine content of β-lactoglobulin is 6.2 per cent (Table 5–4) and the molecular weight is 41,500.

6.2% alanine (m.w. 89) = 4.95% alanine residue (m.w. 71).

.0495 × 41,500 = 2054 g. alanine residues per gram molecular weight of β-lactoglobulin.

2054 ÷ 71 = 28.9, or in round numbers 29 residues.

The calculations gave a total for all the amino acids of 370 residues. Adding together the molecular weights of all the residues, a total of 42,020 was obtained, which is in good agreement with that obtained by ultracentrifuge measurements (41,500). From these data Brand proposed an amino acid formula for β-lactoglobulin as follows:

$$Gly_8Ala_{29}Val_{21}Leu_{50}Ileu_{27}Pro_{15}Phe_9CySH_1(CyS{-}{-})_8Met_9Try_4Arg_7His_4Lys_{33}$$
$$Asp_{36}Glu_{24}(GluNH_2)_{32}Ser_{20}Thr_{21}Tyr_9(H_2O)_4$$

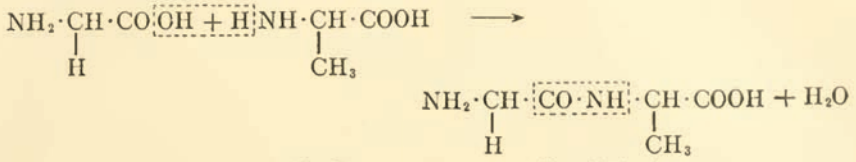
In this formula the amino acids are shown by the first three letters of the name, and the number of residues by the figure following the abbreviation. $(CyS{-})_8$ means that 8 half molecules or 4 whole molecules of cystine are present. Glutamic acid residues with the second carboxyl group neutralized as an amide, *i.e.*, glutamine, are represented as $(GluNH_2)_{32}$. The four molecules of water come from the H and OH of terminal NH_2— and COOH groups, respectively, in polypeptide chains. β-Lactoglobulin is believed to contain 4 polypeptide chains per molecule.

In a similar manner the number of amino acid residues present can be calculated for the proteins listed in Table 5–4, for which the data are sufficiently complete.

Linkage of amino acids to form protein

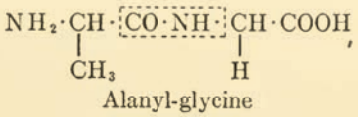
As indicated below in the formulas, the amino acids contain at least one amino group and one carboxyl group. The amino group gives basic

character to the compound, while the carboxyl group gives it acidic properties. When two amino acids are joined together, water is eliminated, and the acids are linked through the carboxyl group of one and the amino group of the other. *This is a fundamental point and should be clearly noted.* For example, the combination of glycine and alanine may be represented thus:



The linkage between the carboxyl group of the glycine and the amino group of the alanine is a peptide linkage. In the formula it is enclosed by dotted lines. A combination of two amino acids is known as a dipeptide. If three amino acids are joined together, a tripeptide is obtained; four amino acids give a tetrapeptide.

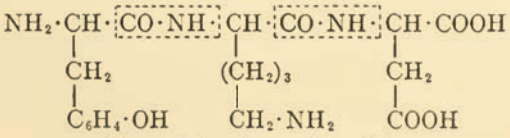
There is obviously another dipeptide of these two amino acids in which the carboxyl group of the alanine is linked to the amino group of the glycine:



Alanyl-glycine

If two molecules of glycine or two of alanine are combined, two more peptides, glycyl-glycine and alanyl-alanine, respectively, are obtained. Thus with two amino acids there are four possible peptides.

A tripeptide consists of three amino acids joined together by two peptide linkages, as is illustrated by the tripeptide, tyrosyl-lysyl-aspartic acid:



Tyrosyl-lysyl-aspartic acid

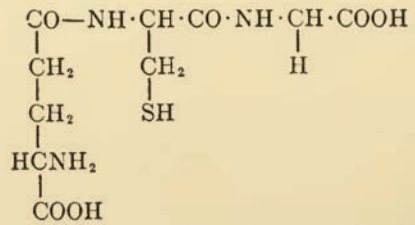
There are 27 possible tripeptides with three amino acids, provided each acid is used once, twice, or three times in a given combination. The number of combinations (polypeptides) that can be obtained by use of 25 different amino acids is almost unlimited. Sherman stated, "If a protein be imagined made up of 30 molecules of 18 different amino acids, one taken twice, one three times, another 3, one 4, one 5 times and 13

taken once each, there would be 10^{27} isomers." There is an adequate mathematical basis for the existence of literally billions of proteins—many more than probably occur in nature. The surprising thing is that with so many possibilities, cells through countless generations produce proteins of identical chemical and physical characteristics.

Attention is called to the occurrence of a free amino and a free carboxyl group at the left and right ends, respectively, in the formulas of the above peptides. Because of these groups, peptides possess both basic and acidic properties. The side chains may also contribute to the basicity and acidity of the peptide because of free amino and carboxyl groups contained therein. Lysine and aspartic acid are examples of amino acids containing basic and acidic groups in the side chains. There are several more such dicarboxylic and basic amino acids. (See formulas of the amino acids.) In long peptides such as proteins it is the side chains, and not the end groups, that contribute most to the basicity or acidity of the molecule. Note also that the phenol group of the tyrosine is free in the above peptide, and, hence, it will have properties characteristic of this group, *e.g.*, positive Millon and xanthoproteic tests. In proteins there will be many such distinctive groups free, *e.g.*, phenol, indole, imidazole, to impart their characteristic properties to the molecule.

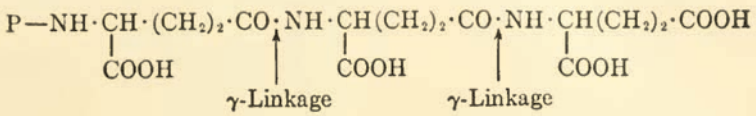
Hundreds of peptides have been made in the laboratory. Fischer, one of the most famous investigators of the composition of the proteins, made a large number of polypeptides. One of these contained 3 leucine and 15 glycine radicals, which gives a molecular weight of 1213—one of the largest molecules that has ever been produced synthetically. These synthetic polypeptides possessed many of the properties of native proteins such as solubility, color tests, and hydrolysis by enzymes. Although they are far from being as complex as the native proteins, their synthesis is a considerable step toward an understanding of the way in which a protein molecule is put together.

In the preceding discussion of peptides no consideration has been given to the possibility of the second carboxyl group of aspartic acid and glutamic acid being involved in the linkages. Glutathione, an important constituent of all cells, is a tripeptide in which glutamic acid is linked to the next amino acid through the γ-carboxyl instead of the α-carboxyl. Thus,



Glutathione (γ,L-glutamyl-L-cysteyl-glycine)

Glutamic acid is also linked through its γ-carboxyl in several peptides of bacteriological origin. In the vitamin, pteroyltriglutamic acid (one of the several forms of folic acid), the second and third glutamic acid residues are linked to the preceding by a γ-linkage. Thus,



P stands for the pteroyl group. For the complete formula see p. 247. The corresponding triglutamate with α-glutamic acid linkages has also been synthesized, but it has no vitamin potency.

A still more unusual polypeptide is the capsular substance produced by *Bacillus anthracis*. This substance is a long chain polypeptide containing only D-glutamic acid residues, 50 to 100 in number, joined together mainly through γ-linkages. There is some indication that small numbers of γ-glutamic acid linkages occur in such common proteins as casein, edestin, hemoglobin, etc. To date, the authors have found no reports regarding the occurrence of second or β-carboxyl linkages for aspartic acid in naturally occurring polypeptides.

Sequence of the amino acids in the protein molecule

Data are now available for the number of amino acid residues present in many proteins, and information concerning the order in which the amino acids occur in the chain is gradually accumulating. Data have been obtained by partially hydrolyzing the protein and isolating some of the peptides that have been formed, then determining the order of the amino acids in the peptides. Examples of such peptides are as follows: glycyl-alanine, alanyl-glycine, and glycyl-tyrosine from silk fibroin; prolyl-phenylalanine and leucyl-glutamic acid from gliadin; phosphoseryl-glutamic acid from casein.

Applying this technique to the antibiotic gramicidin S, Consden *et al.* isolated four dipeptides and one tripeptide from the partial hydrolysis products. These pieces contained all of the five amino acids found in gramicidin S, and, hence, the order of the amino acids was worked out to be:

(-α-(L-valyl)-L-ornithyl-L-leucyl-D-phenylalanyl-L-proline-)$_{1 \text{ or } 2}$

Gramicidin S then appears to be a cyclic penta- or decapeptide. It is obviously a very simple compound in comparison to the common proteins. An open chain peptide having the same sequence of amino acids has been synthesized by Harris and Work, but its antibiotic activity was small as compared to that of the natural compound. These results

suggest that the antibiotic activity of gramicidin S is related largely to its cyclic structure.

A very extensive piece of work has been done recently by Sanger and associates on insulin. This protein consists of two types of polypeptide chains. One type has glycine as the N-terminal (free NH_2 group) residue, and the other chain is headed by phenylalanine. Each molecule of insulin contains two glycine chains and two phenylalanine chains held together by —S—S— bridges of cystine residues (p. 117). The bridges can be broken by oxidation with performic acid to $—SO_3H$ groups, thus setting free the chains. The two types of chains can then be separated and the sequence of the amino acids in them determined. This was done by partial hydrolysis with hydrochloric acid and enzymes and isolation of the peptides split off from the chain. More than 60 peptides ranging from dipeptides to hexapeptides were isolated from the hydrolyzate of the phenylalanine chain, and their structures determined by chromatography. From all these fractions the sequence of the amino acids in the whole chain was deduced to be:

Phe.Val.Asp.Glu.His.Leu.$CySO_3H$.Gly.Ser.His.Leu.Val.Glu.Ala.Leu.Tyr.Leu.Val.

$CySO_3H$.Gly.Glu.Arg.Gly.Phe.Phe.Tyr.Thr.Pro.Lys.Ala.

The total number of amino acid residues in the chain is 30.

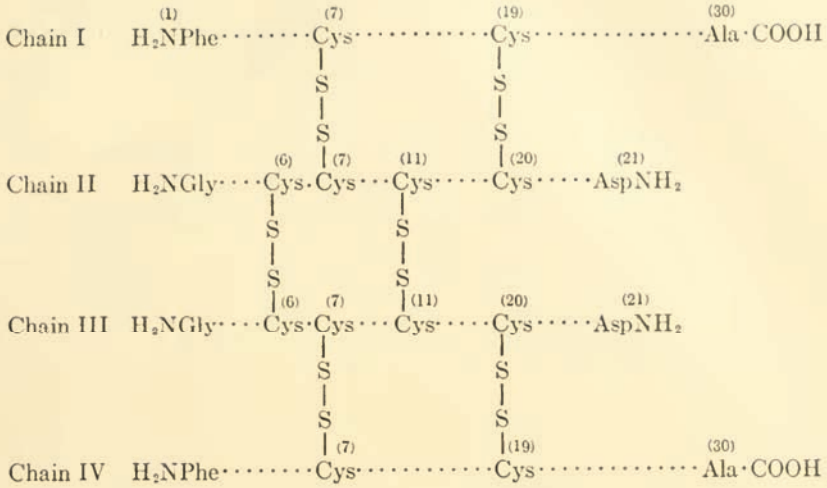
The sequence of the amino acids in the glycine chain was worked out in the same way. The chain consists of 21 amino acid residues arranged as follows:

Gly.Ileu.Val.Glu.$CySO_3H$.$CySO_3H$.Ala.Ser.Val.$CySO_3H$.Ser.Leu.Tyr.Glu.Leu.

Glu.Asp.Tyr.$CySO_3H$.Asp.

In the phenylalanine chain the last amino acid residue in the chain, that is, the residue with a free carboxyl group, is alanine. In the glycine chain the carboxyl terminal residue is designated as Asp., *i.e.*, an aspartic acid residue. In the intact insulin it is an asparagine residue. During acid hydrolysis the amide group is split to give aspartic acid. It is believed that two other aspartic acid residues in the chain are in reality asparagine residues, and six of the glutamic acid residues are actually glutamine residues. The exact location of these amide groups is not known.

The relation of the four chains to one another is still to be determined. Several arrangements are possible, but the presence of four cysteic acid residues in the glycine chain and two in the phenylalanine chain suggests that the two glycine chains lie between the phenylalanine chains. A diagram of such an arrangement for the intact insulin follows:



This diagram shows the four chains headed by phenylalanine and glycine and ending in alanine and asparagine, $AspNH_2$. The cystine residues with one-half in one chain and the other half in the adjacent chain form the —S—S— bridges that hold the chains together. The numbers above the chains indicate the order of the amino acid residues in the chains. The intervening amino acid residues given on p. 132 have been omitted because of limitations of space. The chains are probably not strung out in a long line, as shown in the diagram, but may be coiled to form layers, as will be apparent from the discussion in the next section.

Even though the sequence of the amino acids in the chains is known and the relation of the chains to one another may be as assumed in the diagram, the question as to what there is about this arrangement that gives insulin its hormone property still remains unanswered. In time there may be an answer even to that question.

Structure of the protein molecule

A protein of moderate size such as egg albumin, having a molecular weight of 40,000 and an average amino acid residue weight of 120, contains about 350 amino acid residues. The residues are joined together to form a number of chains of varying lengths. For example, egg albumin and β-lactoglobulin contain four chains, edestin—six, lactalbumin —nine, and insulin—four. The average number of amino acid residues in the chains of these proteins is calculated to be 89 for egg albumin, 92 for β-lactoglobulin, 16 for lactalbumin, and 25 for insulin.

Since a protein has three-dimensional form, the chains must be arranged so as to provide such a structure. Many theories, based mainly

on X-ray data, have been advanced regarding the arrangement of the chains in the protein molecule. Since the carbon and nitrogen atoms in the backbone of the chain are at angles to one another, the chain has a zig-zag appearance with hydrogens, oxygens, and "tails" (side chains of the amino acid residues) sticking out at various angles with the backbone carbons and nitrogens. (See hydrogen bonding, p. 135.)

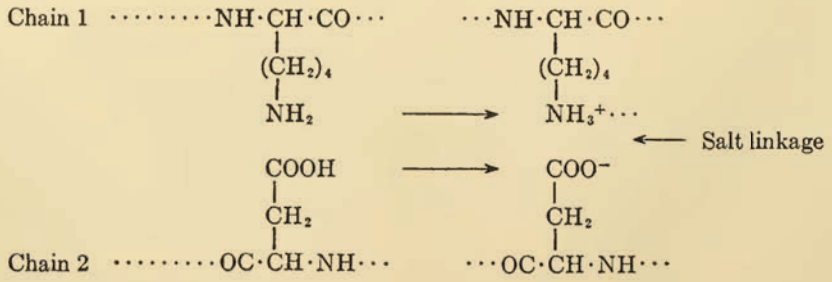
If such a chain is folded back on itself, or if two chains are arranged in the right order, the result is a puckered two-dimensional structure or layer. Such a layer may be compared to a piece of lace crocheted from a single thread. Several such layers may be superimposed on one another to build a three-dimensional structure, as in a layer cake. Layers may also be folded back and forth in some such fashion as a road map is folded. When a protein is spread on a water surface, the layers unfold and form a film only one layer thick. In such cases the polar groups (*e.g.*, amide, carboxyl, hydroxyl, phenol, etc.) are drawn into the water surface while the nonpolar groups (*e.g.*, hydrogen, paraffin, benzene, etc.) are repelled and extend upward from the water surface.

A somewhat different arrangement of the chains has been proposed by Pauling for the structure of fibrous proteins, for example, α-keratin of wool. According to this view, the chains are arranged in the form of a helix (a spiral spring is an example of a helix) to give a hollow cylinder-like structure. For collagen, Pauling concludes that the molecule consists of three chains twisted about one another to give a rope-like effect.

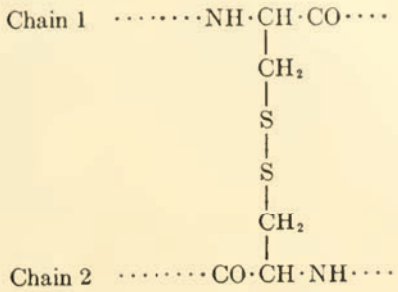
Theories on the structure of proteins are constantly changing as more X-ray, infra-red, and other physical measurements are made. To quote Bernal, "The problem of the protein structure is now a definite and not unattainable goal."

One of the difficult problems in protein structure is to find a satisfactory explanation for the manner in which the folds in a molecular chain, or the layers in a molecule, are held together. Several types of forces have been postulated, three of which will be mentioned here.

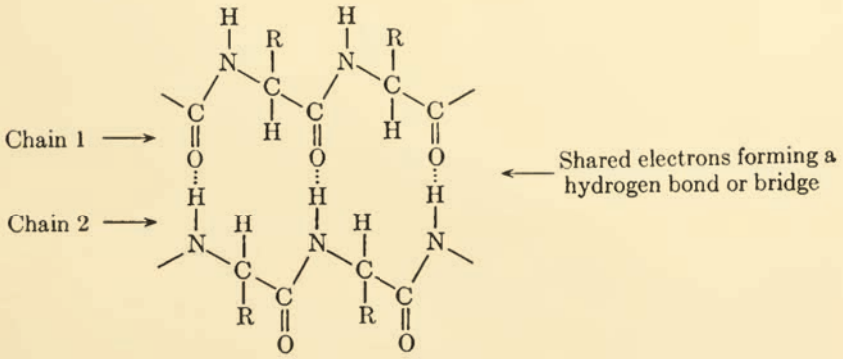
Salt Linkage. Basic groups of one side chain (the second amino group of lysine, the imadazole group of histidine, and the guanidino group of arginine) may be united to acidic groups (the second carboxyl group of aspartic and glutamic acids) of another chain to form an electrostatic bond, that is, a salt.



Sulfide Linkage. One-half of a cystine molecule may be part of one chain; the other half may be located in a second chain to form an —S—S— linkage.



Hydrogen Bonding. A third, and more important, type of binding is the hydrogen bond, in which electrons are shared between the hydrogen of an imino group (—NH—) located in one chain and the oxygen of a carboxyl group (—CO—) in another chain.



There are approximately as many —NH— and —CO— groups in a protein molecule as there are amino acid residues. Hence the number of hydrogen bonds set up would be in the hundreds. Although a single hydrogen bond constitutes only a weak chemical linkage, a dozen together provide about as much strength as a covalent (*e.g.,* —C—C—) bond.

Proteins are classified on the basis of their shape as fibrous or corpuscular. ("Globular" was formerly used as the descriptive term, but "corpuscular" seems now to be the preferred designation.) As the term implies, the fibrous proteins are long and slender, or unsymmetrical. In some cases the length is 30 times the cross section. The corpuscular proteins are more nearly symmetrical. Many of these have a long axis (length) only twice as great as the short axis (thickness). Methemoglobin is reported to be 64 Å long, 48 Å wide, and 36 Å thick.[1]

[1] One millimeter equals 10,000,000 Å (angstrom units).

CONJUGATED PROTEINS AND THEIR PROSTHETIC GROUPS

The conjugated proteins (protein plus a prosthetic group) will, of course, give other products than amino acids on hydrolysis. Some important compounds that either constitute the prosthetic group or arise from it on hydrolysis follow.

Nucleoproteins

Nucleic acids occur either in the free state or in combination with proteins to form nucleoproteins. Because of their number, complexity, and importance, the nucleoproteins and the nucleic acids will be discussed in a separate chapter.

Phosphoproteins

The phosphoproteins also give phosphoric acid on hydrolysis. Phosphoric acid is linked to serine through the hydroxy group of the amino acid to form a phosphoric acid ester that can unite with bases, *e.g.*, calcium, to form salts. The serine phosphoric acid esters seem to account for most of the phosphorus in the phosphoproteins.

Glycoproteins

These are ill-defined proteins containing carbohydrate complexes (chondroitin sulfuric acid or mucoitin sulfuric acid) as prosthetic groups. Typical examples of these proteins occur in salivary mucin, gastrointestinal mucus, and the vitreous humor of the eye. They are slippery materials and, hence, serve as useful lubricants. They facilitate the movement of the food through the intestinal tract and, since they are not digestible, protect the tract against the proteolytic enzymes.

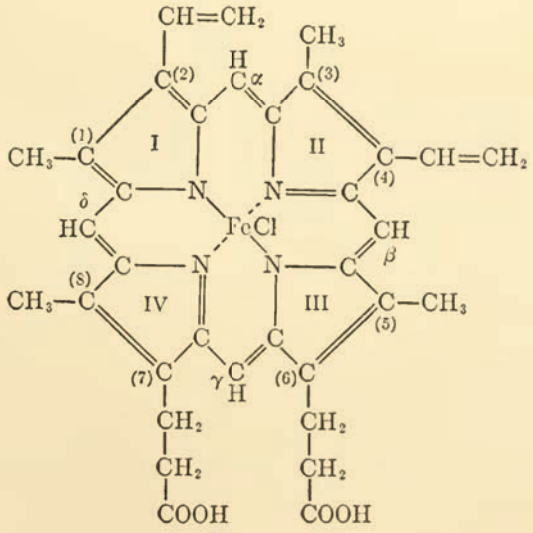
On hydrolysis, chondroitin sulfuric acid gives one mole each of glucuronic acid, galactosamine, acetic acid, and sulfuric acid. Mucoitin sulfuric acid gives glucosamine instead of galactosamine, but the other hydrolysis products are the same as from chondroitin sulfuric acid. There is still some uncertainty as to how the several products are bound together. These carbohydrate complexes account for 25–35 per cent of the glycoprotein. In consequence the nitrogen content is low, 9–13 per cent.

Chromoproteins

This group of proteins probably includes a larger number of important proteins than any other subdivision. They are functional rather than structural, playing a role of the first order as carriers of gases, mediators in the oxidation process, and intermediates in the phenomenon of vision.

Hemoglobin. The chromoproteins contain colored prosthetic groups. Hemoglobin, the red pigment of the blood cells, is the best known protein of this type. Heme is the prosthetic group and globin, a histone, is the protein part of the combination. Heme accounts for only 3.8 per cent of the hemoglobin; globin comprises the other 96.2 per cent. Heme contains four atoms of iron, which amounts to 0.33 per cent of the hemoglobin. There are four heme units attached to one globin part to make up hemoglobin.

The exact nature of the attachment is not known, but it is generally considered to be either an ionic or covalent bond between the iron of the heme and the histidine of the globin. The union is weak and can be easily broken by warming with acetic acid and sodium chloride. A salt, hemin ($C_{34}H_{32}N_4O_4FeCl$), is formed and this crystallizes readily. The structural formula of hemin is



Heme

It consists of four pyrrole groups (I–IV) joined together through four methene groups ($\alpha,\beta,\gamma,\delta$). To the pyrrole groups are also attached four methyl groups (1,3,5,8), two vinyl groups (2 and 4), and two propionic acid residues (6 and 7). In the center of the formula is a trivalent atom

of iron Fe^{+++} (the iron is ferrous Fe^{++} in the original hemoglobin and is oxidized to ferric Fe^{+++} in the process of isolating hemin) which is joined to the four nitrogens of the pyrrole rings by partial valences and to a chlorine atom to form the chloride salt. The four pyrrole groups, without the side chains and iron, form a unit known as porphin, and derivatives thereof are called porphyrins. Many porphyrins occur in nature, *e.g.*, coproporphyrin of feces, uroporphyrin of urine, and chlorophyll (p. 389). Chlorophyll, the green pigment of plants, contains an additional ring, differs from heme in its side chains, and has Mg at the center instead of Fe. One of the propionic side chains in chlorophyll is linked as an ester to the unsaturated alcohol phytol. That the most important animal pigment and the most important plant pigment should have related structures is something to be noted carefully.

The outstanding chemical feature of heme (hemoglobin) is its ability to combine with oxygen and thus serve as a transport agent in the blood. Each heme combines with one molecule of oxygen. Since there are four heme units in each molecule of hemoglobin, the reaction between hemoglobin and oxygen may be represented thus:

$$Hb + 4O_2 \rightleftarrows Hb(O_2)_4$$

Hemoglobin Oxyhemoglobin

more conveniently, but less exactly,

$$Hb + O_2 \rightleftarrows HbO_2$$

This reaction takes place in the lungs; in the tissues it is reversed.

Hemoglobin also reacts with carbon monoxide to form a combination that is several hundred times stronger than that with oxygen. It therefore takes much oxygen to displace the carbon monoxide from the hemoglobin complex and makes breathing of carbon monoxide a very dangerous matter. Hemoglobin also combines with carbon dioxide, and a considerable part of the carbon dioxide contained in the blood is combined with it. Hemoglobin is thus a carrier of gases both in going from the lungs and in returning to them.

If hemoglobin is exposed to mild oxidizing agents, for example, potassium ferricyanide, the iron is converted from the ferrous state Fe^{++} to the ferric form Fe^{+++}. The resulting hemoglobin is called methemoglobin. It can carry only one-half as much oxygen as hemoglobin and does not readily release oxygen.

The hemoglobins of different species differ in the globin part of the molecule, but the heme part is the same in all. Note the difference in crystalline structure of hemoglobins from different species (Figs. 5–1 to 5–4).

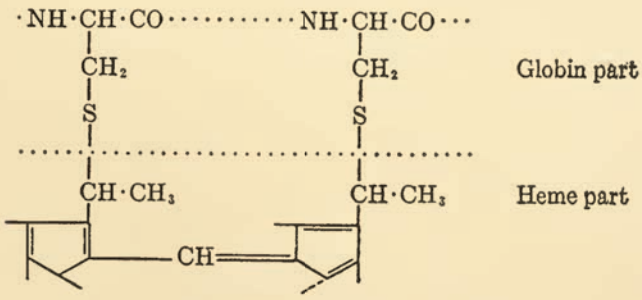
A form of hemoglobin has recently been reported to occur in the

nodules of leguminous plants and is believed to play a role in the fixation of atmospheric nitrogen by these nodules.

Erythrocruorin and Chlorocruorin. Many proteins other than hemoglobin contain iron porphyrins as prosthetic groups. The basic structure of these iron porphyrins is the same as that of the heme in hemoglobin, and in many forms of life they also serve as the oxygen-carrying agent. An example of this type of compound is erythrocruorin, the respiratory pigment of the common earthworm and other worms. Chlorocruorin, a green pigment, serves the same purpose for marine worms, *e.g., Spirographis.* These pigments are dissolved in the blood, not contained in blood cells as are the hemoglobins. They are usually of large molecular weight, several million, and contain many heme groups, *e.g.,* 190 in the chlorocruorin of *Spirographis.* The side chains of the hemes differ from those in hemoglobin, but the iron is in the ferrous state as in hemoglobin.

Cytochromes. There are at least three cytochromes, *a, b,* and *c,* that differ from one another in solubility, reaction to cyanide, and other properties. The cytochromes occur in all oxygen-using cells and, hence, are the most widely distributed heme proteins in nature. They form an oxidation-reduction system and serve as carriers of hydrogen in the oxidation scheme of cells. (See pp. 283 and 333.)

Cytochrome *c* is the best known cytochrome and has been obtained in a crystalline and homogeneous state. It is a small protein, molecular weight 13,000, and contains only one heme per molecule. The iron content is 0.43 per cent and is present either in the ferrous or ferric state, according to whether the cytochrome is in the reduced or oxidized state. The heme part of the molecule is bound not only by an iron-to-histidine bonding, as in hemoglobin, but, in addition, is joined by two covalent linkages between vinyl groups of the heme and cysteine residues of the globin. Thus,



These bonds are very stable, and on hydrolysis the cysteine residues go with the heme. In other words, the —S—C— linkage is more stable than the peptide linkage —CO—N—.

Heme-containing Enzymes. Catalases occur widely in plant and animal tissues. Beef liver catalase has been crystallized and contains

approximately 0.1 per cent of iron. On the basis of 225,000 for the molecular weight, the iron content corresponds to four atoms per molecule. The iron appears to be in the ferric state, and the heme unit is the same as that in hemoglobin.

Two peroxidases, horse radish peroxidase II and cytochrome *c* peroxidase from yeast, have been obtained in crystalline, or highly purified, form and found to contain the same heme as hemoglobin. Two other peroxidases, myeloperoxidase from leucocytes and lactoperoxidase from milk, contain green-colored hemes and, hence, are also called verdoperoxidases. Both have hemes containing iron, but the structures of these hemes have not yet been determined.

Other Metal-containing Proteins. Ferritin is a brown-colored protein containing up to 23 per cent of iron. The iron is present as colloidal iron oxide or phosphate and is very loosely bound to the protein. Ferritin occurs in the liver, spleen, and bone marrow and is believed to serve as a storage form of iron.

Hemocyanins serve as respiratory proteins for the lobster, octopus, and other marine animals. The blood of the lobster becomes blue when aerated, hence the term hemocyanin, literally blue blood (a term that man has applied to himself as a mark of distinction, without considering its connotations). Hemocyanin contains copper (about 0.35 per cent), but the nature of the prosthetic group, if any, carrying the copper is still an unsettled question. The molecular weights ascribed to the hemocyanins are enormous, 2 to 5 million.

Additional copper-containing proteins are: hemocuprein from red blood corpuscles, hepatocuprein from the liver and several oxidizing enzymes, *e.g.*, tyrosinase, ascorbic acid oxidase, etc. Other metals forming complexes with proteins, but not having color, are magnesium in carboxylase, zinc in insulin and carbonic anhydrase, and manganese in arginase.

Other Colored Proteins. Flavoproteins or "yellow enzymes" have riboflavin phosphate, or a dinucleotide of riboflavin and adenine, as the prosthetic group and are yellow in color, hence the name "yellow enzyme" given to them. About a dozen such enzyme proteins have been reported. They play an important role as hydrogen transport agents and can exist in either a reduced or oxidized state (see Fig. 10–4).

Rhodopsin is a red, light-sensitive pigment found in the retina of land and marine animals, *e.g.*, man, cattle, squid, and plays an important role in vision. It is a chromoprotein that consists of the prosthetic group, *cis*-retinene 1, and a protein called opsin. Retinene 1 is the aldehyde ($C_{19}H_{27}CHO$) corresponding to vitamin A_1 ($C_{19}H_{27}CH_2OH$). In fresh water vertebrates, such as fish, a somewhat different chromoprotein called porphyropsin takes the place of rhodopsin. Porphyropsin contains retinene 2, which corresponds to vitamin A_2. Opsins from different sources

combine with either retinene, but it is not certain that they are identical proteins (p. 204).

Lipoproteins

As a class, these proteins are not well-defined and are regarded by some investigators as mixtures of lipides and proteins rather than as chemical entities. They occur in cell nuclei, blood, milk, bacteria, etc. Lipovitellin of egg yolk, thromboplastic protein (functioning in blood clotting) of lung tissue, and the polymyxin antibiotics are examples of such proteins. The prosthetic groups that have been obtained from lipoproteins include lecithin, cephalin, and fatty acids.

GENERAL PROPERTIES OF THE PROTEINS

Form

As usually obtained, proteins are amorphous, but, under carefully controlled conditions, they can be made to crystallize. Great progress has been made in the preparation of crystalline proteins in recent years, especially of those that are enzymes. About 150 proteins have been crystallized and, of these, approximately 40 are enzymes, respiratory pigments, toxins, viruses, or hormones. The most commonly prepared crystalline proteins are the animal albumins and the plant globulins.

Size of the protein molecule

From several lines of evidence it is known that the protein molecule is large. A minimum molecular weight can be determined in various ways: from the percentage of some element, such as sulfur, phosphorus, or iron, which occurs in small quantities; from the percentage of some amino acid found in the protein. The true molecular weight is obtained by physical methods such as ultracentrifuge measurements and osmotic pressure determinations.

The following example illustrates the method of calculating the minimum molecular weight of hemoglobin from the iron content, 0.33 per cent. Assuming one atom of iron, atomic weight 56, a proportion is set up as follows:

$$0.33 : 56 = 100 : x; \quad x = 17{,}000 \text{ minimum m.w.}$$

Since there are four atoms of iron in hemoglobin, the true molecular weight becomes 68,000, which is in excellent agreement with ultracentrifuge data.

If all the nitrogen of a protein is accounted for in the amino acid analysis, the number of amino acid residues can be calculated, and the sum of the molecular weights of the residues probably equals the molecular weight of the protein. The molecular weight of β-lactoglobulin obtained in this way is 42,020, and that given by the ultracentrifuge method is 41,500.

Protamines are small proteins, m.w. 3000 to 5000, and hemocyanins are very large, m.w. several million. Many viruses have apparent molecular weights up to several hundred million. Home authorities prefer to call these values particle weights rather than molecular weights, giving proteins probable molecular weights approximating 17,600, or a multiple thereof; that is, 2, 4, 6, 8, 16, 24 . . . , 384 times 17,600. For example, zein and hemoglobin are reported to have molecular weights of about 35,000 (2 units) and 68,000 (4 units), respectively. The molecular weight of the proteoses is supposed to range from 4000 to 5000 and that of the peptones from 800 to 1000.

Color reactions

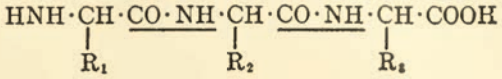
Many color tests have been proposed by different investigators to detect the presence of proteins. The most common are given below.

The Xanthoproteic Test. With concentrated nitric acid, most proteins give a yellow color that becomes more pronounced if the solution is made alkaline. The familiar stain that is formed if nitric acid comes in contact with the skin is due to the action of this acid upon the proteins of the skin. The cause of the test is the formation of a nitro-phenyl-derivative somewhat similar, perhaps, to picric acid. A modified phenyl grouping such as is contained in tyrosine and tryptophan seems to be necessary to the test. Tryptophan gives a better test than tyrosine, whereas phenylalanine, although it contains the phenyl group (C_6H_5), does not give the test at all. All the common proteins give the test, but considerable quantities of protein are required. Although the test is general, it is not very sensitive.

The Millon Test. A brick-red color is developed when some proteins are heated with the Millon reagent (mercury dissolved in nitric acid). The reaction is due to the presence of the phenol group (C_6H_4OH), which is contained in the amino acid, tyrosine. Proteins that contain no tyrosine will therefore fail to give a Millon test. Gelatin gives a faint Millon test either because it contains a minute quantity of tyrosine or because it has not yet been freed of other tyrosine-containing proteins. Carbolic acid and salicylic acid, which are not amino acids, likewise give this test because they contain the phenol group.

The Biuret Test. A pink to purple color is obtained when proteins are treated with alkali and minute quantities of copper sulfate. The

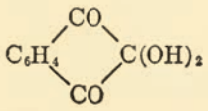
color is due to the presence of two peptide groups, —CO·NH—. When three amino acids are joined, two such groups are formed, for example,



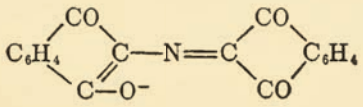
Tripeptides, with the exception of glycyl-glycyl-glycine, give the test, as do also peptones, proteoses, and all native proteins. It is one of the most general of the protein color tests.

The Hopkins-Cole Test. A purplish color is developed when a protein containing tryptophan radicals is treated with the Hopkins-Cole reagent (magnesium glyoxylate). The cause of the color development is the indole group, which exists in the amino acid, tryptophan. It has been assumed that the color is due to the formation of indigo by the action of the reagents on indole groups.

The Ninhydrin Test. All amino acids (except proline and hydroxyproline) and, hence, all proteins give a blue to purple color with ninhydrin.



Only a free amino and a free carboxyl group are required for the test. All amino acids except the two mentioned possess such groups. The blue color results from a condensation of two molecules of the reagent with ammonia, which splits off from the amino group of the amino acid. The color compound is the anion of a salt and has the following formula:



Precipitation

The precipitatin of proteins varies considerably with the reagent that is used. Some reagents precipitate only the globulins, whereas others precipitate a larger number of proteins but do not precipitate the peptones. Some reagents precipitate not only the proteins and peptones but also carry down certain amino acids. Among the most effective precipitants are ammonium sulfate, magnesium sulfate, mercury salts, trichloracetic acid, phosphotungstic acid, tungstic acid, tannic acid, colloidal iron, and strong solutions of alcohol. Phosphotungstic acid appears to be the reagent which precipitates the largest percentage of nitrogen, whether this is in the form of proteins, peptones, amino acids, or other nitrogenous compounds.

The theory for the precipitation of proteins is based on the amphoteric character of the proteins, *i.e.*, the presence of both basic and acidic groups in the molecule. If the protein is more basic than the reagent, the precipitate is a protein salt of that reagent. If it is more acidic than the reagent, the precipitate comes down as proteinate; that is, the protein behaves as an acid and the precipitant acts as a base. For each protein there is a pH value called the iso-electric point, on one side of which the protein acts as a base and on the other side as an acid; for example, the iso-electric point of gelatin is at the pH value of 4.7. Below 4.7, gelatin acts as a base and is precipitated by acids as a gelatin salt such as gelatin hydrochloride. Above 4.7, gelatin behaves as an acid and is precipitated as a gelatinate such as sodium gelatinate. The two types of precipitation may be represented in simplified form by the following equations.

By acids: $RNH_2 + HCl \rightarrow RNH_2 \cdot HCl$

By bases: $R'COOH + NaOH \rightarrow R'COONa$

In place of hydrochloric acid in the acid precipitation, we may have acids such as picric, tannic, tungstic, phosphotungstic, etc., which form more insoluble compounds with proteins. Instead of sodium hydroxide, calcium or barium hydroxide may be used to give the corresponding calcium or barium salt. The metallic salt of the protein may also be formed by adding a soluble salt of the metal to the sodium hydroxide-protein solution. For example, if lead acetate, one of the best protein precipitants, is added to a solution, lead proteinate is formed and, since this is insoluble, a precipitate is produced.

The above theory of precipitation and salt formation assumes the formation of an excess of positive charges on the protein molecule

$$RNH_2 + H_2O \rightarrow RNH_3^+ + OH^-$$

below the iso-electric point and an excess of negative charges on the protein molecule

$$R'COOH \rightarrow R'COO^- + H^+$$

above the iso-electric point. Under the first condition, the protein reacts with negative ions, *e.g.*, Cl^-, to form a protein salt; under the second, it reacts with positive ions, *e.g.*, Na^+, to form a proteinate. At the iso-electric point the positive charges equal the negative charges; hence the protein combines with neither acid nor base. Above or below the iso-electric point the protein is unbalanced and therefore combines with oppositely charged ions.

Denaturation

Neurath and associates define protein denaturation as "any nonproteolytic modification of the unique structure of a native protein, giving

rise to definite changes in chemical, physical, or biological properties." Modifications of structure by addition of a group such as acetylation, which obviously changes the properties of the protein, are not included in the term.

Denaturation, or more properly coagulation, of protein has been observed by man ages before he knew about proteins. It is a phenomenon that must have come to his attention soon after he began to prepare his food by heating it. Despite its apparent nature, a clear understanding of coagulation is still lacking.

A most striking example of coagulation is the conversion of egg white from a liquid to a solid when an egg is boiled. This change appears first to involve denaturation of the egg white followed by aggregation of the denatured protein into flocs and then into a solid coagulum. The evidence for these stages is the fact that if a solution of egg white is heated in a salt-free medium below or above the isoelectric point of the protein, about pH 4.7, the solution remains clear. However, the increases in viscosity and SH groups, *e.g.*, in cysteine, show that the egg white has been denatured. If the pH of the clear solution is adjusted to 4.7, coagulation takes place without any further heating.

A decrease in solubility is only one of the changes that occur when a protein is denatured. A greater susceptibility to the action of enzymes is another effect. The digestibility of egg white, for example, is much increased by heating. A third effect is a partial or complete loss of biological activity, if the native protein possesses such a property. Enzymes, antibodies, and viruses lose their potency when completely denatured. Loss of crystallizability, changes in viscosity, and an increase in the number of reactive groups (*e.g.*, sulfhydryl (—SH), disulfide, phenol, and indole) are other changes brought about by denaturation. Not all of these changes are equally apparent and, if measured quantitatively, do not run parallel to one another. These differences are probably an indication that different structural arrangements in the molecule are affected to a varying degree by the denaturing agent. It is apparent that no single criterion is adequate as an index of denaturation. If the denaturation has not gone too far, the process may be reversed under suitable conditions, and much of the protein recovered in the original state.

Denaturation may be brought about not only by heat but also by freezing, irradiation with ultraviolet light, acid, alkali, alcohol, urea, guanidine salts, and some of the new type of detergents (p. 87). The last three compounds are the reagents ordinarily used in experimental work to produce denaturation.

The measurement most commonly used to detect denaturation is titration of the SH and S—S groups with a mild oxidizing agent. The S—S groups are first reduced with cyanide to SH and then titrated. The

increase in SH groups after reduction is a measure of the S—S groups present. The reagents commonly used in titrating are potassium ferricyanide and a dye, porphydrindin blue. The end point of the former is determined with sodium nitroprusside as indicator. It gives a red color with SH groups. On reduction the dye becomes a colorless compound.

Phenol and indole groups have weak reducing properties in alkaline solution toward ferricyanide and can be determined after previous oxidation of the SH groups at a different pH.

The question naturally arises as to what occurs in the structure of the molecule when a protein is denatured. Denaturation is undoubtedly a disruption of the highly complex and precisely organized folding of the peptide chains and the interrelation of these chains to one another. In the disruption of this organization, groups (*e.g.*, SH) that were previously buried deep in the molecule become exposed and reactive. Biological activity, which depends upon very specific arrangements in the chain structures, is lost with the disappearance of these arrangements. The regeneration of the crystalline form and partial recovery of biological activity that has been observed in some instances may be ascribed to an only partial disorganization of the chains. If the pattern still exists, the chains may refold themselves into the original structure.

QUANTITATIVE DETERMINATION OF PROTEIN (KJELDAHL METHOD)

Crude protein

The protein content of any food material is obtained by determining the total nitrogen and multiplying the result by 6.25. This value is called the crude protein content of the food material. Two assumptions are made in determining protein in this way:

1. All nitrogen is assumed to be present in the substance as protein. This is not necessarily the case, as food materials frequently contain a large proportion of the nitrogen in other than protein form. Examples of compounds containing nitrogen that are not protein are amino acids, amides, alkaloids, cyanates, purines, pyrimidines, creatine, and creatinine. From this list it is seen that in foodstuffs there are many compounds that contain nitrogen but are not proteins. In finished products such as seeds the major portion of the nitrogen is in the form of protein, while in actively metabolizing tissue such as partly formed seeds and vegetables, particularly string beans and green peas, the larger part of the nitrogen may be in other than protein form. In muscle and other animal tissue there is a considerable proportion of so-called meat extractives that contain nitrogen but are not protein.

2. The second assumption is that all proteins contain 16 per cent of nitrogen, in which case the ratio of the protein to nitrogen is 100:16 or 6.25. Since the nitrogen content varies from 15.5 to 18.7 per cent, the use of the factor 6.25 may give too high or too low results, depending upon the protein under consideration. In the case of the cereals, the nitrogen content is higher than 16 per cent, and the factor 5.7 more nearly represents the ratio between the protein and the nitrogen. With milk the factor 6.37 is used. In general, however, the factor 6.25 is used and is the one that is employed in the calculation of general tables of analysis for the composition of food materials.

It should be observed that the term "crude protein" has nothing to do with the purity of the protein, but merely represents results obtained by laboratory procedure.

Purpose of reagents used in determining total nitrogen

In the Kjeldahl determination of total nitrogen, carbon and hydrogen must be oxidized and the nitrogen converted into ammonium sulfate. Sulfuric acid is the oxidizing reagent that is used. The material first becomes black and charred, but after prolonged heating it clears up and becomes free from carbon. The action of the sulfuric acid may be represented by the following equation:

$$H_2SO_4 \rightarrow H_2O + SO_2 + O$$

The sodium or potassium sulfate added raises the boiling point of the solution and thus aids in the oxidation of the material. A small amount of copper selenite is also added and acts as a catalyst to promote oxidation. The oxidation of protein may be represented by the oxidation of a fragment of a protein, for example, glycine.

$$CH_2(NH_2)\cdot COOH + 3O \rightarrow 2CO_2 + H_2O + NH_3$$

$$2NH_3 + H_2SO_4 \rightarrow (NH_4)_2SO_4$$

Through the addition of concentrated sodium hydroxide, ammonia is liberated and collected by distilling it into a known amount of standard acid. From the amount of acid neutralized by the ammonia, the nitrogen and, hence, the crude protein ($N \times 6.25$) are obtained by a simple calculation.

REVIEW QUESTIONS ON PROTEINS

1. Name two of the chief proteins in (1) milk, (2) eggs, (3) blood, (4) wheat; one protein in (5) corn and (6) peas. Which of these proteins probably contain (a) phosphorus? (b) sulfur?

2. Give the graphic formula for (1) cystine, (2) tyrosine, (3) tryptophan, (4) lysine, (5) uric acid. Write out the graphic formula for (a) a dipeptide, (b) a tripeptide. Give the elements contained in hematin.

3. Name eight chemical groups contained in amino acid molecules. Which of the above are responsible for color tests, and which reagent is used to show the presence of the particular group?

4. What is meant by the term "incomplete protein"? Name two proteins that are incomplete, and show in what respect they are incomplete.

5. Explain the use of the following reagents in the Kjeldahl determination of crude protein, (1) concentrated H_2SO_4, (2) Na_2SO_4, (3) concentrated NaOH. What is meant by the term "crude protein"? As usually determined, what assumptions are involved?

6. Which three proteins are the most abundant in the food that you have eaten in the last three meals? Give data on which you base your conclusions.

7. Approximately how many grams of tryptophan are there in a glass of milk?

8. Which two amino acids are found most abundantly in proteins?

9. Name two kinds of reagents that may be used to precipitate proteins. In each case explain how the reagent brings about the precipitation, and name the product formed.

10. Give the name and source of (1) three proteins that are enzymes, (2) three proteins contained in industrial nonfood products, (3) one protein that is a hormone, (4) two proteins that are respiratory pigments, (5) two proteins other than those named in (1) to (4) that have been crystallized.

11. Which protein color tests will be positive with the tripeptide, tyrosyltryptophylcystine? Give reasons for conclusion in case of each test.

12. Correct the following statements if incorrect:

(1) Milk and eggs are proteins that belong in every diet.

(2) In calculating crude protein it is assumed that all the protein of the sample is in the form of nitrogen.

(3) In practical nutrition the deficiencies of incomplete proteins are remedied by the addition of individual amino acids.

13. If the tyrosine content of a protein is 3.78 per cent and the molecular weight is 42,000, calculate the number of moles of tyrosine per mole of protein.

14. From the data given in the text calculate the number of residues of glycine, alanine, and valine in β-lactoglobulin. Do your figures in round numbers check with those of the text?

15. In which two ways does cystathionine appear to be split in S-transfer?

16. What is the least number of dipeptides that would have to be isolated and identified to prove the sequence of amino acids in gramicidin S?

17. For outside reading: How would you determine whether a dipeptide containing alanine and leucine is alanyl-leucine or leucyl-alanine?

18. For outside reading: How many isomers of pteroyltriglutamic acid are due to the arrangement of the glutamic acid residues.

REFERENCES AND SUGGESTED READINGS

Anson, M. L. and Edsall, J. T., *Advances in Protein Chemistry,* Vols. I–VII. Academic Press Inc., New York, 1945–51.

Baldwin, E., *Dynamic Aspects of Biochemistry,* 2nd ed., The University Press, Cambridge, England, 1952.

Bernal, J. D., "Structure of Proteins," *Proc. Roy. Inst. Gt. Brit.,* **30,** 541 (1939).

Block, R. J. and Bolling, D., *The Amino Acid Composition of Proteins and Foods,* 2nd ed., Charles C. Thomas, Springfield, Illinois, 1951.

Brand, E., Saidel, L. J., Goldwater, W. H., Kassell, B., and Ryan, F. J., "The Empirical Formula of β-Lactoglobulin," *J. Am. Chem. Soc.,* **67**, 1524 (1945).

Cohn, E. J. and Edsall, J. T., *Proteins, Amino Acids and Peptides,* Reinhold Publishing Corp., New York, 1943.

Consden, R., Gordon, A. H., Martin, A. J. P., and Synge, R. L. M., "Gramicidin S: the Sequence of the Amino Acid Residues," *Biochem. J.,* **41**, 596 (1947).

Gortner, R. A. and Gortner, W. A., *Outlines of Biochemistry,* 3rd ed., John Wiley and Sons, Inc., New York, 1949.

Greenberg, D. M., *Amino Acids and Proteins,* Charles C. Thomas, Springfield, Illinois, 1951.

Harris, J. I. and Work, T. S., "The Synthesis of Peptides Related to Gramicidin S and the Significance of Optical Configuration in Antibiotic Peptides," *Biochem. J.,* **46**, 582 (1950).

Haurowitz, F., *Chemistry and Biology of Proteins,* Academic Press Inc., New York, 1950.

Hawk, P. B., Oser, B. L., and Summerson, W. H., *Practical Physiological Chemistry,* 12th ed., The Blakiston Company, Philadelphia, 1947.

Hubbard, R. and Wald, G., "Cis-trans Isomers of Vitamin A and Retinene in Vision," *Science,* **115**, 60 (1952).

Luck, J. M., Loring, H. S., and Mackinney, G., *Annual Review of Biochemistry,* Vols. I–XXI, Annual Reviews Inc., Stanford, California, 1932–52.

Moore, S. and Stein, W. H., "Chromatography of Amino Acids on Starch Columns," *J. Biol. Chem.,* **176**, 337 (1948), **178**, 53 (1949).

Neurath, H., Greenstein, J. P., Putnam, F. W., and Erickson, J. O., "The Chemistry of Protein Denaturation," *Chem. Rev.,* **34**, 157 (1944).

Pauling, L., "The Configuration of Polypeptide Chains in Proteins," *Record of Chemical Progress,* **12**, 155 (1951).

Roughton, F. J. W. and Kendrew, J. C., *Haemoglobin,* Interscience Publishers Inc., New York, 1949.

Sanger, F. and Thompson, E. O. P., "The Amino-acid Sequence in the Glycyl Chain of Insulin," *Biochem. J.,* **52**, iii (1952).

Sanger, F. and Tuppy, H., "The Amino Acid Sequences in the Phenylalanyl Chain of Insulin," *Biochem. J.,* **49**, 463, 481 (1951).

Schmidt, C. L. A., *The Chemistry of the Amino Acids and Proteins With Addendum,* 2nd ed., Charles C. Thomas, Springfield, Illinois and Baltimore, 1945.

Sherman, H. C., *Chemistry of Food and Nutrition,* 7th ed., The Macmillan Company, New York, 1946.

Wald, G., "The Chemistry of Rod Vision," *Science,* **113**, 287 (1951); **115**, 60 (1952).

West, E. S. and Todd, W. R., *Text Book of Biochemistry,* The Macmillan Company, New York, 1951.

Chapter 6

NUCLEOPROTEINS, NUCLEIC ACIDS AND RELATED SUBSTANCES

Introduction

As already stated in the previous chapter, nucleoproteins are conjugated proteins having nucleic acids as prosthetic groups. The literature on nucleoproteins and nucleic acids is extensive and increasing at a rapid rate. Hundreds of papers and more than a dozen reviews dealing with various phases of the subject have appeared in the last five years. The great activity in this field can be attributed to the widespread occurrence, intriguing chemical nature, and metabolic importance of these substances.

One compelling reason for the attention being given to nucleoproteins and nucleic acids is a growing belief that these compounds are closely associated with the reproductive processes and may furnish the physical basis of heredity. Chromosomes, the constituents of cells carrying the hereditary characters, or genes, are largely, if not wholly, nucleoproteins. Whatever chemical compounds make up the genes, it is obvious that such compounds must be sufficiently diverse in character to permit the almost infinite number of combinations that occur in nature. In the nucleic acids there is adequate diversification to meet this requirement.

NUCLEOPROTEINS

The term nucleoprotein arose because nucleic acids and the associated protein, protamine, were first obtained from the highly nucleated material of pus cells and fish sperm. Other nuclear cells such as thymus, liver, spleen, and yeast are rich in nucleoproteins, but some nonnuclear cells, for example, red blood corpuscles, also are a good source of nucleoprotein. In fish sperm cells nucleoprotein makes up 50–80 per cent of the solid material and over 90 per cent of the defatted nucleus. In the cell sap of tobacco plants infected with virus, the nucleoprotein which makes up the virus may amount to 2 g. per liter of sap. In yeast cells the nucleoprotein amounts to only about 0.15 per cent of the dry matter. Bacteria are much higher than yeast in nucleoproteins, *e.g.*, 2–3 per cent of the dry matter in *Escherichia coli* cells.

Preparation

Nucleoproteins are labile substances; hence to obtain them from cells only mild reagents and low temperatures can be used.

An example of present day methods is the procedure Mirsky and Ris used for the preparation of chromosomes from calf spleen. The tissue was broken up in a Waring blender and soluble matter removed with 0.14*M* sodium chloride, in which the chromosomes were insoluble. Unbroken cells and other coarse materials were separated by filtering through finely woven cloth and sedimenting the chromosomes in a centrifuge at 3500 rpm. The suspension, filtration through cloth, and centrifugation were repeated several times until the phosphorus figure, which was taken as a measure of the nucleic acid content, became constant. Under the microscope the material showed the characteristic coiled pattern of chromosomes. Differential centrifugation methods are rapid, and 10 to 15 g. of purified material can be obtained in the course of a morning's work.

Tobacco bushy stunt virus, which is destroyed by almost any chemical, was separated by Stanley from the other constituents of ground tobacco leaves by means of differential centrifugation and purified still further by crystallization.

Separation of components

The nucleic acid-protein complex is usually a very loose one, and the combination can be broken up and the two components separated in various ways. Some of these procedures are as follows:

The protein may be denatured either by heating or by treating with urea to give an insoluble compound, this then being removed by filtration. Alternatively, the protein may be obtained by extraction with chloroform and octyl alcohol, leaving the nucleic acid in solution. A third method is to destroy the nucleic acid with the enzyme, ribonuclease, and then separate the unchanged protein from the digested material.

If the nucleic acid is the component wanted, it may be obtained from the solution after the protein has been removed by one of the methods described above. The protein may also be destroyed by trypsin and the nucleic acid recovered. If the protein is a histone, separation may be accomplished by dialysis against 1*M* sodium chloride. Histone passes through the membrane and leaves the nucleic acid behind.

Linkage between protein and nucleic acid

The bond between protein and nucleic acid is in some cases electrostatic, or salt-like, since the two parts may be separated by the passage of

an electric current through the solution. The positively charged protein moves to the cathode and the negatively charged nucleic acid goes to the anode. Thymus nucleoprotein is an example of this type. The protein is histone, and, since this is a strongly basic substance, it forms a salt with nucleic acid. The two components are probably joined together through the basic groups of arginine, histidine and lysine, and the phosphoric acid groups of the nucleic acid. Histones and protamines are particularly high in arginine. If a sample of histone or protamine dissolved in 0.14*M* sodium chloride is added to a solution of nucleic acid having the same strength of sodium chloride, the two components react and form a precipitate. Such precipitates are probably the result of interaction of the molecules as a whole, and not arginine alone. The same quantity of arginine, histidine, and lysine as is contained in the protamine forms no precipitate at the same pH.

There may be a second type of nucleoprotein in which the protein and nucleic acid are bound together by nonpolar linkages. Some nucleoproteins migrate as single entities, and the protein cannot be separated from the nucleic acid until it has first been denatured. However, some investigators do not regard this as conclusive evidence of a nonsalt type of bonding because the structure of the native protein is quite different from that of the denatured protein, and the configuration may modify the strength of the bonding groups.

Quantitative data on components

In Table 6–1 are given examples of nucleoproteins, the kinds of proteins contained therein, and the proportion of protein to nucleic acid. Histones and protamines are the common type, but lipoproteins occur frequently in the nucleoproteins of animal tissues. Chromosomes seem to contain two kinds of nucleoprotein, histone and nonhistone types. Mirsky, from whose papers these data are taken, differentiates the two types on the basis of the insolubility of the nonhistone protein in $HgSO_4$-H_2SO_4 solution (histone is soluble) and its greater tryptophan content (nonhistone contains 1.36 per cent and histone only 0.14 per cent).

Nucleolipoprotein, containing lipoprotein in combination with nucleic acid, indicates a protein having two prosthetic groups, lipide and nucleic acid. Several viruses, *e.g.*, vaccinia virus, have been found to be nucleolipoprotein complexes.

The nucleic acid portion of the nucleoproteins may range from a small fraction, 5 per cent, to more than half of the total. The desoxyribonucleic acid (DNA) seems to make up a larger percentage of the nucleoprotein than the ribose form (RNA). Both types of nucleic acid often occur in the same cell. In chromosomes the DNA type predominates, but in yeast cells the RNA form is in excess.

Table 6–1

Types of protein and nucleic acid found in some typical nucleoproteins

Source of nucleoprotein	Composition of nucleoprotein *	
	Protein, %	*Nucleic acid, %*
Calf thymus	Histone, 40	DNA, 60
Sperm heads of fish	Protamine, 40	DNA, 60
Liver	Lipoprotein, 95	DNA, 5
Chromosomes of calf thymus:		
Soluble fraction, 90%	Histone, 47	DNA, 45
Residual fraction, 10%	Nonhistone, ?	RNA, 11 DNA, 2
Tobacco mosaic virus	Not classified, 94	RNA, 6
Tobacco ring-spot virus	Not classified, 60	RNA, 40
Tuberculin from tubercle bacillus	Not classified, 60	DNA, 40
Yeast	Not classified, 90–105	RNA, 5–10 DNA, ?
Bacteria	Not classified, 80–85	RNA, 15–20

* DNA denotes desoxyribonucleic acid; RNA means ribonucleic acid. The nature of the different types of nucleic acid will be discussed later. Where no figures regarding the amounts are available, this is indicated by a question mark (?) in the second and third columns.

The viruses of the tobacco plant are well-characterized. They have been obtained in crystalline form and their properties carefully determined. The protein part varies from 60 to 94 per cent of the nucleoprotein in the two viruses listed in Table 6–1. The amino acids of tobacco mosaic virus account for 106 per cent of the virus. (See Table 5–4). The nucleic acid is of the ribose type, which is the most abundant type found in plant material.

The molecular weights reported for nucleoproteins are large, 2 million for calf thymus nucleohistone and 40 million for tobacco mosaic virus.

NUCLEIC ACIDS

Component units

The nucleic acids are themselves complex structures with molecular weights ranging from 17,000 for yeast nucleic acid to more than a million for the acid from the thymus gland. The particle size varies with the method of preparation, hence, the smaller weights may represent split products of the larger units. The molecules appear to be rod-like in shape, with the length of the particles 40 to 400 times that of their diameter.

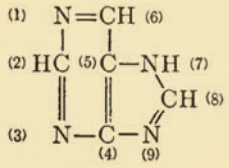
Nucleic acids are divided into two classes depending upon the kind of hydrolysis products. This will be evident from an inspection of the following tabulation.

Products from ribonucleic acid	*Products from desoxyribonucleic acid*	*Classification of products*
Adenine, guanine	Adenine, guanine	Purine
Cytosine, *uracil*	Cytosine, *5-methylcystosine, thymine*	Pyrimidine
D-*Ribose*	D-*Desoxyribose*	Pentose
Phosphoric acid	Phosphoric acid	Acid

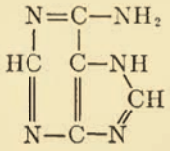
Four of the products are found in both kinds of nucleic acid. The distinguishing products are uracil and D-ribose for ribonucleic acid and thymine, 5-methylcytosine and D-desoxyribose for the other type. The sugars are the products from which the terms ribonucleic acid (RNA) and desoxyribonucleic acid (DNA) are derived.[1] It was at one time believed that ribonucleic acid was found only in plants and desoxyribonucleic acid only in animal cells. This view is incorrect, and it now appears probable that all cells contain both types. The desoxyribonucleic acid seems to be most abundant in the nucleus of the cell, and the ribonucleic type to be preponderant in the cytoplasm surrounding the nucleus.

Purines and pyrimidines

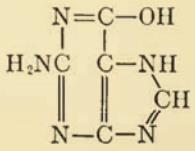
The structural formulas of the purines are given below:



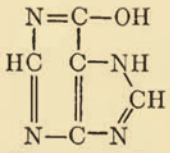
Purine
(synthetic base)



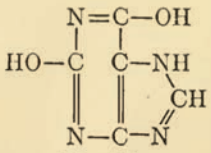
Adenine



Guanine

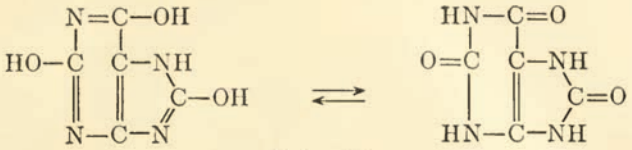


Hypoxanthine



Xanthine

[1] Research workers use the terms pentosenucleic acid (PNA) and desoxypentosenucleic acid (DNA) as general terms and limit the more specific names to nucleic acids where the sugars have been definitely established as ribose (in yeast, liver, and tobacco mosaic virus) and desoxyribose (in calf thymus). This cautious attitude is probably desirable for the research worker, but since ribose and desoxyribose are the usual sugars found in nucleic acids, it is less complicated for the beginning student to start with the particular name and proceed to the general term when it becomes necessary.



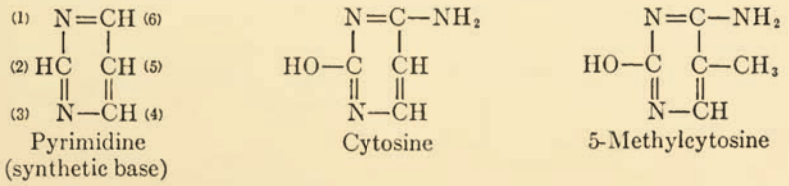
(hydroxy form) Uric acid (carbonyl form)

The naturally occurring purines may be referred to the synthetic base, purine. The various atoms in the rings are numbered to denote the position in the structure. Thus adenine is designated 6-aminopurine, and guanine is 2-amino-6-hydroxypurine. In solution the hydroxy compounds exist in two tautomeric forms, as is shown in the formulas for uric acid.

Adenine and guanine are constituents of native nucleic acids, and the other three compounds are products derived from the first two as a result of metabolism. Hypoxanthine is formed in the body by deamination of adenine, and on oxidation this product forms xanthine, which may also originate from deamination of guanine. Oxidation of xanthine gives uric acid, which is the end product of purine metabolism in man.

Three other purines occur in our common beverages. Caffeine (1,3,7-trimethylxanthine) is found in the coffee bean to the extent of about 1 per cent and in tea leaves to about 2 per cent. Theobromine (3,7-dimethylxanthine) occurs in the cocoa bean (about 2 per cent) and theophylline (1,3-dimethylxanthine) is found is small quantities in tea leaves. Caffeine, removed from the coffee bean in the making of decaffeinized coffee, is used in the manufacture of cola drinks. However, this supply is not sufficient for the purpose, and much of the caffeine used in soft drinks is made synthetically.

The pyrimidines have the following structural formulas:



Pyrimidine (synthetic base) Cytosine 5-Methylcytosine

Uracil Thymine

The pyrimidine ring forms a part of the purine structure and is numbered in the same way. Thus cytosine is 2-hydroxy-6-aminopyrimidine, uracil is 2,6-hydroxydipyrimidine, and thymine is 2,6-dihydroxy-5-methylpyrimidine. 5-Methylcytosine, reported many years ago as occurring

in the tubercle bacillus, has now been found in small quantities in desoxyribonucleic acids from cattle spleen, fish sperm, and wheat germ, but not in the desoxyribonucleic acids from bacteria and viruses. To date, none has been found in ribonucleic acids. The vitamin, thiamine, is a pyrimidinethiazole combination. Its pyrimidine can be designated as 2,5-dimethyl-6-aminopyrimidine.

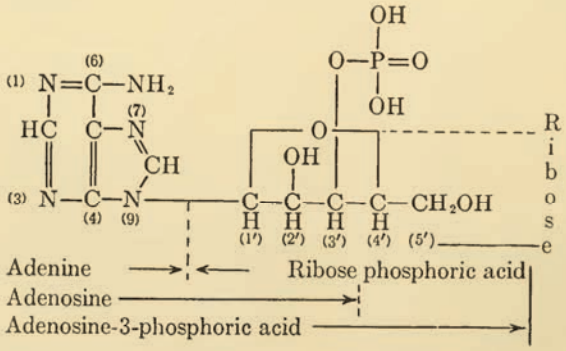
Nucleosides and nucleotides

If the hydrolysis of a nucleic acid is done under suitable conditions, the breakdown may be stopped before it is complete and nucleosides and nucleotides obtained. A nucleoside is a purine or pyrimidine-pentose combination, and a nucleotide is a nucleoside-phosphoric acid complex. The various bases, and the corresponding well-known nucleosides and nucleotides, are listed in the following tabulation:

Base	*Nucleoside*	*Mononucleotide*
Adenine	Adenosine	Adenylic acid
Guanine	Guanosine	Guanylic acid
Cytosine	Cytidine	Cytidylic acid
Uracil	Uridine	Uridylic acid
Thymine	Thymidine	Thymidylic acid

The nucleosides are designated, according to the sugar contained in them, as ribosides (adenosine, guanosine, cytidine and uridine) or as desoxyriboside (thymidine). The corresponding mononucleotides contain the same sugars as the nucleosides. There are obviously other nucleosides and nucleotides of desoxyribose, but, to date, these have not been given specific names. They are often designated by prefixing the term desoxy to the names of the ribose-containing compounds: desoxyadenosine, desoxyadenylic acid, etc.

The structural formula of one of the mononucleotides, adenylic acid, will be given to show the order of the components and the linkages that join the parts together.



Adenylic acid (adenosine monophosphate)

Adenine and ribose are joined by a β-glycosidic linkage from the nitrogen of position-9 of the adenine to carbon-1 of the ribose. The adenosine and phosphoric acid are united by an ester linkage. Adenosine and adenylic acid have now been synthesized by Todd and co-workers, so that there is no doubt remaining as to their structure. These workers have also synthesized a number of other nucleosides and nucleotides.

A second type of adenylic acid has been obtained from yeast nucleic acid. In this type the phosphoric acid is thought to be linked to carbon-2′ instead of 3′. If this proves to be correct, it provides strong support for the view that nucleotides are linked together through phosphoric acid, which is joined to one nucleotide at carbon-2′ and to the other nucleotide at carbon-3′. These structures are complicated and difficult to determine, but distinct progress is being made toward their final solution.

In the pyrimidines the β-glycosidic linkage is between the nitrogen at number 3 position and carbon-1′ of the ribose. The phosphoric acid is located at carbon-3′, as in the purine nucleotides. These structures have been established beyond doubt by synthesis of cytidine and uridylic acid.

The nucleosides and nucleotides of desoxyribose are believed to have the same linkages between base, sugar, and phosphoric acid as those of ribose, but the data are not so conclusive as for the ribose compounds.

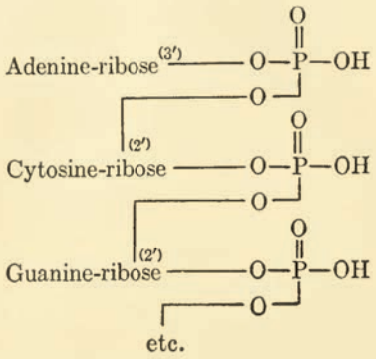
Polynucleotides

Nucleic acids found in nature are usually polynucleotides, consisting of many purine and pyrimidine nucleotides joined together to form a single structure. Estimates ranging from 60 nucleotides for yeast nucleic acid to 4000 for thymus nucleic acid have been given. Such estimates are in accord with the large molecular weights obtained for these nucleic acids. Formerly, it was believed that these large molecules were made up of many tetranucleotide units, but this view is now generally abandoned.

The molar ratios of the different purines and pyrimidines to one another do not bear out the idea of a regularly occurring tetranucleotide unit. For example, Chargaff and co-workers found that the desoxynucleic acid of salmon sperm gave molar ratios of the constituents as follows: Adenine to guanine, 1.43; thymine to cytosine, 1.43; adenine to thymine, 1.02; guanine to cytosine, 1.02; purines to pyrimidines, 1.02. Adenine occurred in excess of guanine, and thymine was more abundant than cytosine. Oddly enough the ratios are the same in both cases, and the total purine is equal to the total pyrimidine content. In other desoxynucleic acids Chargaff found adenine exceeded guanine, and thymine outweighed cytosine, but the ratios were different than in the salmon nucleic acid.

The nucleotides are joined together through phosphoric acid groups, but just how is not known. One possibility is a linkage from carbon-2′

of one ribose to phosphoric acid and a second linkage of this to carbon-3′ of the next nucleotide Such an arrangement for three ribonucleotides can be represented as follows:



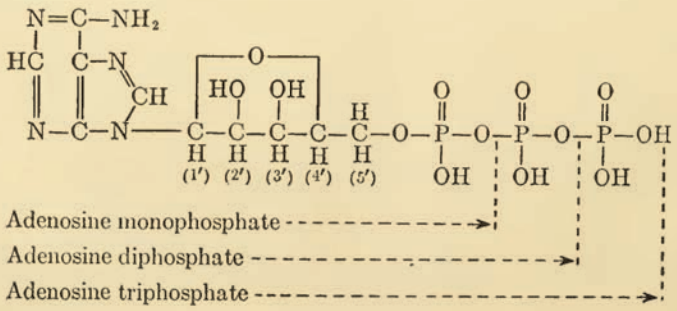
The numbers 2′ and 3′ denote the carbon atoms in the ribose to which the phosphoric acid group is linked.

Obviously, there are other ways of joining the guanine nucleotide to the other two ribonucleotides. For example, the ribose part of the guanine nucleotide could be linked to the adenine nucleotide, instead of to the phosphoric acid in the cytosine nucleotide. The result would be a triester structure, instead of the diester form given by the first type of combination. The number of possibilities would increase as the number of nucleotides joined together became larger. A more complicated branching structure would be the result. There is no information as to the sequence of the nucleotides in the nucleic acid structure.

In the desoxynucleotides, carbon-2′ of the sugar can not serve as a point of linkage because it has no hydroxyl group. It is generally assumed that the desoxynucleotides are joined together by way of carbons-3′ and 5′ of their respective sugars.

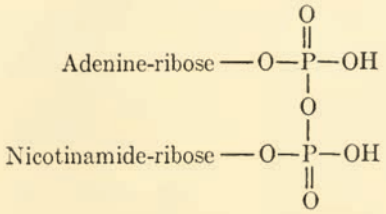
Substances related to nucleosides and nucleotides

Adenosine Phosphates. A nucleotide with the phosphoric acid at carbon-5′ of the ribose, instead of at carbon-3′, is the well-known muscle-adenylic acid. It is also called adenosine monophosphate (AMP).



Besides the monophosphate, adenosine forms a diphosphate, ADP, and a triphosphate, ATP. The three derivatives of adenosine play an outstanding role in enzyme chemistry and intermediary metabolism.

Coenzymes I *and* II. These compounds are dinucleotides of adenine and the base, nicotinamide. Coenzyme I is also known as diphosphopyridine nucleotide (DPN), and coenzyme II as triphosphopyridine nucleotide (TPN). The make-up of DPN can be seen from the following designation:

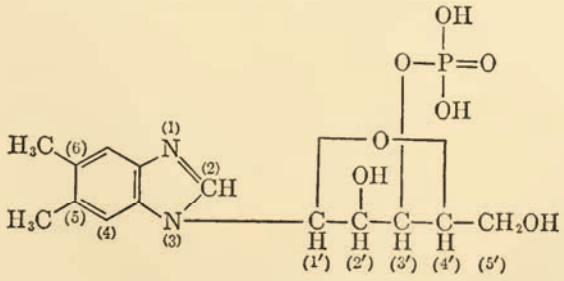


Note that the two nucleotides are joined through the phosphoric acid molecules, instead of the pentose-phosphoric acid-pentose structure found in nucleic acids. The phosphoric acid groups are linked to carbon-5′ of the ribose units.

TPN is like DPN except that it has a third phosphoric group attached to carbon-2′ of the ribose found in the adenine nucleotide. For the structural formulas of DPN and TPN see p. 276.

Flavin Nucleotides. There is a so-called mononucleotide, riboflavin phosphate, and a dinucleotide of riboflavin phosphate and adenylic acid. The riboflavin phosphate is not a true nucleotide because the ribose part is replaced by the sugar alcohol corresponding to ribose, *viz.*, ribitol. The difference in structure is evident from the formula on p. 278. The flavin nucleotides are coenzymes, and a discussion of their function will be given in the chapter on enzymes.

5,6-*Dimethylbenzimidazole Riboside.* This nucleotide and the corresponding nucleoside have been obtained as degradation products of vitamin B_{12}. The structural formula is



Note that the ribose has a furanose, instead of a pyranose, structure. In the formula, the phosphoric acid is attached to carbon 3′ of the ribose, but the point of attachment is uncertain. It may be at carbon 2′.

REVIEW QUESTIONS ON NUCLEOPROTEINS

1. Which types of protein have been found in nucleoproteins? How much is protein and how much is nucleic acid in a few typical nucleoproteins?

2. Name eight compounds that may be obtained from nucleic acids on hydrolysis? To what class of compounds does each belong? Name some other examples of each class.

3. Define nucleoside and nucleotide, giving an example of each. Which nucleotides are associated with (1) vitamins, (2) enzymes?

4. Compare the probable structure of the dinucleotide, adenylic acid—cytidylic acid with the pyridine and flavin dinucleotides. Point out similarities and differences.

5. Catalog the linkages that join together the parts of a nucleic acid.

REFERENCES AND SUGGESTED READINGS

Chargaff, E., Lipshitz, R., Green, C., and Hodes, M. F., "The Composition of Desoxyribonucleic Acid of Salmon Sperm," *J. Biol. Chem.*, **192**, 223 (1951).

Cold Spring Harbor Symposia on Quantitative Biology, *Nucleic Acids and Nucleoproteins*, Vol. XII, The Biological Laboratory, Cold Spring Harbor, N. Y., 1947.

Davidson, J. N., *The Biochemistry of the Nucleic Acids*, John Wiley and Sons, Inc., New York, 1950.

Green, H. N. and Stoner, H. B., *Biological Actions of the Adenine Nucleotides*, H. K. Lewis and Co., Ltd., 1950.

Mirsky, A. E. and Ris, H., "Isolated Chromosomes," *J. Gen. Physiol.*, **31**, 1 (1948).

Schlenk, F., "Chemistry and Enzymology of Nucleic Acids," *Advances in Enzymology*, **9**, 455 (1949).

Stanley, W. M., "Purification of Tomato Bushy Stunt Virus by Differential Centrifugation," *J. Biol. Chem.*, **135**, 437 (1940).

Symposia of the Society for Experimental Biology, No. 1, *Nucleic Acid*, The University Press, Cambridge, England, 1951.

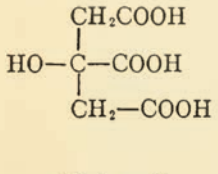
Symposium on Biochemistry of Nucleic Acids, *J. Cellular Comp. Physiol.*, **38**, Supplement, 1951.

Tamm, C., Hodes, M. E., and Chargaff, E., "The Formation of Apurinic Acid from the Desoxyribonucleic Acid of Calf Thymus," *J. Biol. Chem.*, **195**, 49 (1952).

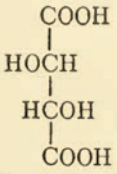
Chapter 7

ACIDITY

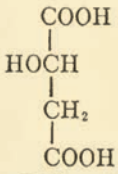
So many acids and bases, both organic and inorganic, occur in living organisms that only a few of them can be considered here. Some of the more common organic acids are citric in citrus fruits, tartaric in grapes, malic in apples, and oxalic in rhubarb and spinach. In some cases these acids are present in the plant tissues in the free condition,



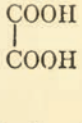
Citric acid



L-Tartaric acid



L-Malic acid

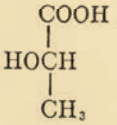


Oxalic acid

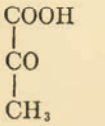
but often they occur as salts. Thus lemon juice contains about 5 per cent of free citric acid, but much of the tartaric acid in grape juice has had its acidity partially neutralized by potassium. Crystals of potassium acid tartrate, "cream of tartar," often are deposited from grape juice or wine on long standing. Similarly, the oxalic acid in some plants is free, and in others exists as calcium oxalate. This is a matter of considerable consequence because free oxalic acid, as well as its water-soluble salts, is a strong poison. The halogeton weed, which grows in Nevada and several other Western states, contains up to 18 per cent of soluble oxalates on the dry weight basis. Livestock, particularly sheep, have been killed by the thousands by eating this weed (see Fig. 7–1). Fortunately, the oxalic acid in rhubarb and spinach is present largely in the form of the very insoluble calcium oxalate which is nontoxic. Rhubarb leaves, however, are said to contain harmful concentrations of soluble oxalates.

A whole series of organic acids is involved in the normal metabolism of carbohydrates and fats in the animal body (Chap. 13). Among the

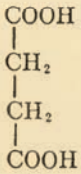
Acetic acid



L-Lactic acid



Pyruvic acid



Succinic acid

simpler members are acetic, lactic, pyruvic, and succinic acids. In addition, living cells contain many organic phosphates (Chap. 13) which are strongly acidic because they contain the phosphate group.

By *Life* photographer Carl Iwasaki (c) *Time* Inc.

Fig. 7–1. Sheep eating the halogeton weed, which is poisonous because of its high content of soluble oxalates. One animal has already died from eating the weed.

Alkaline substances encountered in biological materials are somewhat less numerous than the acids. Methylamine and trimethylamine are weak bases, which are present in decayed fish and contribute to the unpleasant "dead-fish" odor. The strong base, choline, is found in many tissues, usually as a salt such as the chloride, or combined in more complex forms (phospholipides, p. 97). Other amines found in decaying animal matter are formed through decomposition of amino acids (p. 321).

CH_3NH_2	$(CH_3)_3N$	$(CH_3)_3NCH_2CH_2OH$ $\vert$ Cl
Methylamine	Trimethylamine	Choline chloride

All of the bases and acids contained in a tissue contribute to the acidity or alkalinity of that tissue. Acidity (or alkalinity) in biological materials is of two kinds, "total" and "active."

Active acidity refers only to the concentration of hydrogen ions present in the material. Hydrogen ions, it will be remembered, are produced by ionization of an acid, for example,

$$CH_3COOH \rightleftarrows CH_3COO^- + H^+$$

The hydrogen-ion concentration is expressed in terms of pH, which will be considered later.

Total acidity, on the other hand, refers to the total amount of acid present, both ionized and nonionized. It is usually expressed as per cent by weight; for example, vinegar contains 4 to 5 per cent acetic acid.

TOTAL ACIDITY

Determination of total acidity or alkalinity is important when one needs to know how much of some material is required to react with the acid or alkali in another material. The determination usually is accomplished by titration with standard solutions, that is, by measuring the volume of base or acid that is required to react with a given amount of the sample. The first requisite for such an analysis, therefore, is a standard solution, which is simply a solution of known concentration. The concentration of standard solutions is usually expressed in terms of molarity or normality.

Molar solutions

By definition a molar solution is of such concentration that one liter of the solution contains exactly one gram molecular weight, or one *mole*, of the solute. Any given fraction of a liter of such a solution, therefore, would contain an equivalent portion of a gram molecular weight. Hence, withdrawal of aliquots from such a standard solution offers a speedier method of obtaining a known weight of reagent than can be effected by the process of weighing. Moreover, the accuracy is much greater than that which can be attained by weighing minute quantities of material.

Normal solutions

Since a mole of one compound may react with one, two, or more moles of another compound, or with only a fraction—one-half, one-third, and so on—of a mole of the second compound, molar solutions are seldom convenient to use. This is not true of normal solutions, which are so prepared that a given volume of a solution of one compound is equivalent to exactly the same volume of solution of any other compound. This is very clearly evident in the case of acids and bases. In displaying