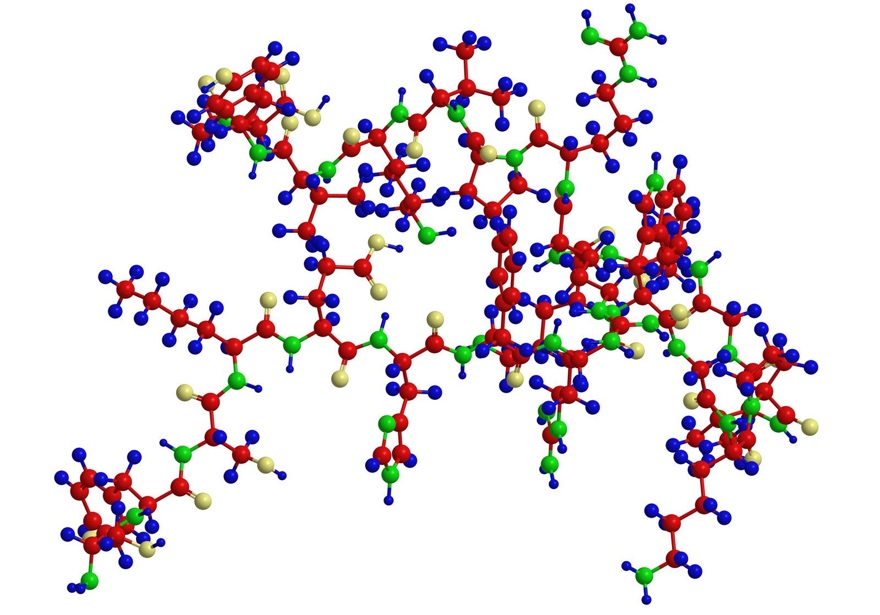
**Protein**

Highly complex substance that is present in all living organisms. Proteins are of great nutritional value and are directly involved in the chemical processes essential for life. The importance of proteins was recognized by chemists in the early 19th century, including Swedish chemist Jöns Jacob Berzelius, who in 1838 coined the term protein, a word derived from the Greek prōteios, meaning “holding first place.” Proteins are species-specific; that is, the proteins of one species differ from those of another species. They are also organ-specific; for instance, within a single organism, muscleproteins differ from those of the brain and liver. 

**peptide**

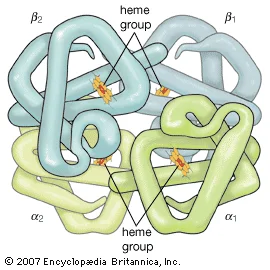
A protein molecule is very large compared with molecules of sugar or salt and consists of many amino acids joined together to form long chains, much as beads are arranged on a string. There are about 20 different amino acids that occur naturally in proteins. Proteins of similar function have similar amino acid compositionand sequence. Although it is not yet possible to explain all of the functions of a protein from its amino acid sequence, established correlations between structure and function can be attributed to the properties of the amino acids that compose proteins.

[](https://cdn.britannica.com/86/161386-050-A91146EF/Legumes-bean-salad-amino-acids-source.jpg)

**legume; amino acid**

Plants can synthesize all of the amino acids; animals cannot, even though all of them are essential for life. Plants can grow in a medium containing inorganic nutrients that provide nitrogen, potassium, and other substances essential for growth. They utilize the carbon dioxide in the air during the process of photosynthesis to form organic compoundssuch as carbohydrates. Animals, however, must obtain organic nutrients from outside sources. Because the protein content of most plants is low, very large amounts of plant material are required by animals, such as ruminants (e.g., cows), that eat only plant material to meet their amino acid requirements. Nonruminant animals, including humans, obtain proteins principally from animals and their products—e.g., meat, milk, and eggs. The seeds of legumes are increasingly being used to prepare inexpensive protein-rich food (see human nutrition).

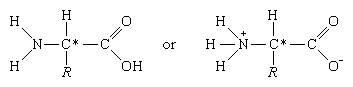
The protein content of animal organs is usually much higher than that of the blood plasma. Muscles, for example, contain about 30 percent protein, the liver 20 to 30 percent, and red blood cells 30 percent. Higher percentages of protein are found in hair, bones, and other organs and tissues with a low water content. The quantity of free amino acids and peptidesin animals is much smaller than the amount of protein; protein molecules are produced in cellsby the stepwise alignment of amino acids and are released into the body fluids only after synthesis is complete.



The high protein content of some organs does not mean that the importance of proteins is related to their amount in an organism or tissue; on the contrary, some of the most important proteins, such as enzymes and hormones, occur in extremely small amounts. The importance of proteins is related principally to their function. All enzymes identified thus far are proteins. Enzymes, which are the catalysts of all metabolic reactions, enable an organism to build up the chemical substances necessary for life—proteins, nucleic acids, carbohydrates, and lipids—to convert them into other substances, and to degrade them. Life without enzymes is not possible. There are several protein hormones with important regulatory functions. In all vertebrates, the respiratory protein hemoglobinacts as oxygen carrier in the blood, transporting oxygen from the lung to body organs and tissues. A large group of structural proteins maintains and protects the structure of the animal body.

**General structure and properties of proteins**

**The amino acid composition of proteins**

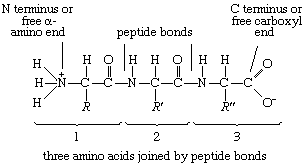
The common property of all proteins is that they consist of long chains of α-amino (alpha amino) acids. The general structure of α-amino acids is shown in . The α-amino acids are so called because the α-carbon atom in the molecule carries an amino group (―NH2); the α-carbon atom also carries a carboxyl group (―COOH).

In acidic solutions, when the pH is less than 4, the ―COO groups combine with hydrogen ions (H+) and are thus converted into the uncharged form (―COOH). In alkaline solutions, at pH above 9, the ammonium groups (―NH+3) lose a hydrogen ion and are converted into amino groups (―NH2). In the pH range between 4 and 8, amino acids carry both a positive and a negative charge and therefore do not migrate in an electrical field. Such structures have been designated as dipolar ions, or zwitterions (i.e., hybrid ions).

Although more than 100 amino acids occur in nature, particularly in plants, only 20 types are commonly found in most proteins. In protein molecules the α-amino acids are linked to each other by peptide bonds between the amino group of one amino acid and the carboxyl group of its neighbour.

Proteins. Formula 2: The peptide bond.

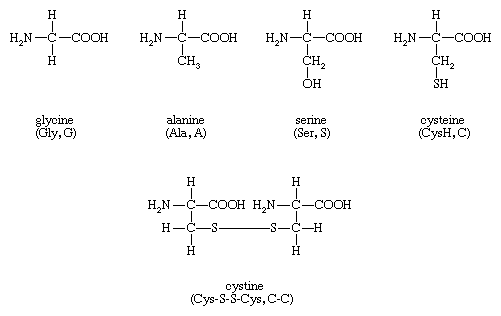
The condensation (joining) of three amino acids yields the tripeptide.



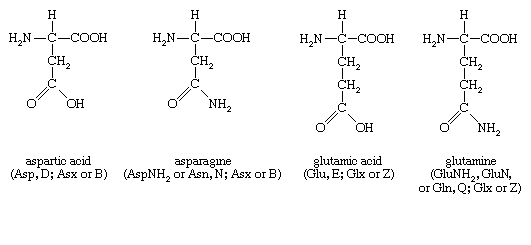
It is customary to write the structure of peptides in such a way that the free α-amino group (also called the N terminus of the peptide) is at the left side and the free carboxyl group (the C terminus) at the right side. Proteins are macromolecular polypeptides—i.e., very large molecules (macromolecules) composed of many peptide-bonded amino acids. Most of the common ones contain more than 100 amino acids linked to each other in a long peptide chain. The average molecular weight (based on the weight of a hydrogenatom as 1) of each amino acid is approximately 100 to 125; thus, the molecular weights of proteins are usually in the range of 10,000 to 100,000 daltons (one dalton is the weight of one hydrogen atom). The species-specificity and organ-specificity of proteins result from differences in the number and sequences of amino acids. Twenty different amino acids in a chain 100 amino acids long can be arranged in far more than 10100 ways (10100 is the number one followed by 100 zeroes).

**Structures of common amino acids**

The amino acids present in proteins differ from each other in the structure of their side (R) chains. The simplest amino acid is glycine, in which R is a hydrogen atom. In a number of amino acids, R represents straight or branched carbon chains. One of these amino acids is alanine, in which R is the methyl group(―CH3). Valine, leucine, and isoleucine, with longer R groups, complete the alkyl side-chain series. The alkyl side chains (R groups) of these amino acids are nonpolar; this means that they have no affinity for water but some affinity for each other. Although plants can form all of the alkyl amino acids, animals can synthesize only alanine and glycine; thus valine, leucine, and isoleucine must be supplied in the diet.

Two amino acids, each containing three carbon atoms, are derived from alanine; they are serineand cysteine. Serine contains an alcohol group (―CH2OH) instead of the methyl group of alanine, and cysteine contains a mercapto group (―CH2SH). Animals can synthesize serine but not cysteine or cystine. Cysteine occurs in proteins predominantly in its oxidized form (oxidation in this sense meaning the removal of hydrogen atoms), called cystine. Cystine consists of two cysteine molecules linked by the disulfide bond (―S―S―) that results when a hydrogen atom is removed from the mercapto group of each of the cysteines. Disulfide bonds are important in protein structure because they allow the linkage of two different parts of a protein molecule to—and thus the formation of loops in—the otherwise straight chains. Some proteins contain small amounts of cysteine with free sulfhydryl (―SH) groups.

Four amino acids, each consisting of four carbon atoms, occur in proteins; they are aspartic acid, asparagine, threonine, and methionine. Aspartic acid and asparagine, which occur in large amounts, can be synthesized by animals. Threonine and methionine cannot be synthesized and thus are essential amino acids; i.e., they must be supplied in the diet. Most proteins contain only small amounts of methionine.

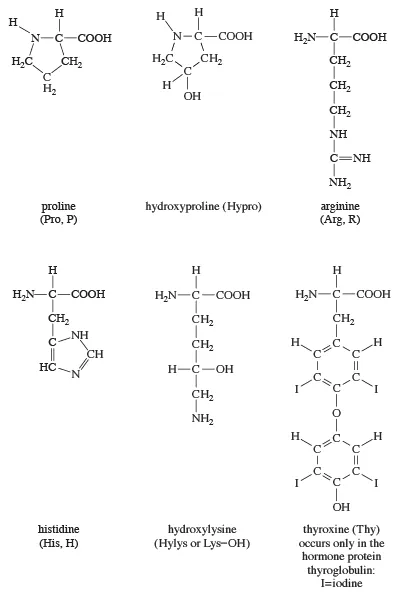
Proteins also contain an amino acid with five carbon atoms (glutamic acid) and a secondary amine (in proline), which is a structure with the amino group (―NH2) bonded to the alkyl side chain, forming a ring. Glutamic acid and aspartic acid are dicarboxylic acids; that is, they have two carboxyl groups (―COOH). 

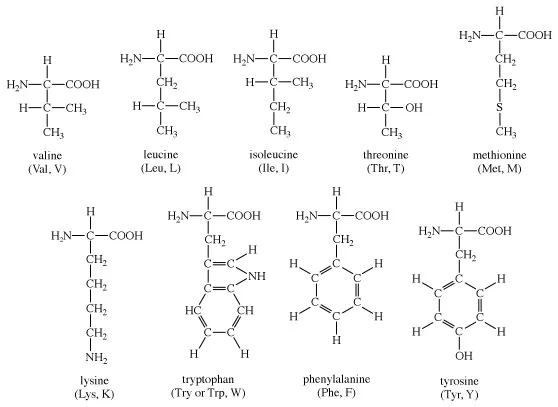
Glutamine is similar to asparagine in that both are the amides of their corresponding dicarboxylic acid forms; i.e., they have an amide group (―CONH2) in place of the carboxyl (―COOH) of the side chain. Glutamic acid and glutamine are abundant in most proteins; e.g., in plant proteins they sometimes comprise more than one-third of the amino acids present. Both glutamic acid and glutamine can be synthesized by animals.

| **Amino acid content of some proteins\*** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **amino acid** | **protein** | | | | | |
| **alpha-casein** | **gliadin** | **edestin** | **collagen (ox hide)** | **keratin (wool)** | **myosin** |
| \*Number of gram molecules of amino acid per 100,000 grams of protein. | | | | | | |
| \*\*The values for aspartic acid and glutamic acid include asparagine and glutamine, respectively. | | | | | | |
| \*\*\*Isoleucine plus leucine. | | | | | | |
| **lysine** | 60.9 | 4.45 | 19.9 | 27.4 | 6.2 | 85 |
| **histidine** | 18.7 | 11.7 | 18.6 | 4.5 | 19.7 | 15 |
| **arginine** | 24.7 | 15.7 | 99.2 | 47.1 | 56.9 | 41 |
| **aspartic acid\*\*** | 63.1 | 10.1 | 99.4 | 51.9 | 51.5 | 85 |
| **threonine** | 41.2 | 17.6 | 31.2 | 19.3 | 55.9 | 41 |
| **serine** | 63.1 | 46.7 | 55.7 | 41.0 | 79.5 | 41 |
| **glutamic acid\*\*** | 153.1 | 311.0 | 144.9 | 76.2 | 99.0 | 155 |
| **proline** | 71.3 | 117.8 | 32.9 | 125.2 | 58.3 | 22 |
| **glycine** | 37.3 | — | 68.0 | 354.6 | 78.0 | 39 |
| **alanine** | 41.5 | 23.9 | 57.7 | 115.7 | 43.8 | 78 |
| **half-cystine** | 3.6 | 21.3 | 10.9 | 0.0 | 105.0 | 86 |
| **valine** | 53.8 | 22.7 | 54.6 | 21.4 | 46.6 | 42 |
| **methionine** | 16.8 | 11.3 | 16.4 | 6.5 | 4.0 | 22 |
| **isoleucine** | 48.8 | 90.8\*\*\* | 41.9 | 14.5 | 29.0 | 42 |
| **leucine** | 60.3 |  | 60.0 | 28.2 | 59.9 | 79 |
| **tyrosine** | 44.7 | 17.7 | 26.9 | 5.5 | 28.7 | 18 |
| **phenylalanine** | 27.9 | 39.0 | 38.4 | 13.9 | 22.4 | 27 |
| **tryptophan** | 7.8 | 3.2 | 6.6 | 0.0 | 9.6 | — |
| **hydroxyproline** | 0.0 | 0.0 | 0.0 | 97.5 | 12.2 | — |
| **hydroxylysine** | — | — | — | 8.0 | 1.2 | — |
| **total** | 839 | 765 | 883 | 1,058 | 863 | 832 |
| **average residual weight** | 119 | 131 | 113 | 95 | 117 | 120 |

The amino acids proline and hydroxyproline occur in large amounts in collagen, the protein of the connective tissue of animals. Proline and hydroxyproline lack free amino (―NH2) groups because the amino group is enclosed in a ring structure with the side chain; they thus cannot exist in a zwitterion form. Although the nitrogen-containing group (>NH) of these amino acids can form a peptide bond with the carboxyl group of another amino acid, the bond so formed gives rise to a kink in the peptide chain; i.e., the ring structure alters the regular bond angle of normal peptide bonds.

Proteins usually are almost neutral molecules; that is, they have neither acidic nor basic properties. This means that the acidic carboxyl ( ―COO−) groups of aspartic and glutamic acid are about equal in number to the amino acids with basic side chains. Three such basic amino acids, each containing six carbon atoms, occur in proteins. The one with the simplest structure, lysine, is synthesized by plants but not by animals. Even some plants have a low lysine content. Arginine is found in all proteins; it occurs in particularly high amounts in the strongly basic protamines (simple proteins composed of relatively few amino acids) of fish sperm. The third basic amino acid is histidine. Both arginine and histidine can be synthesized by animals. Histidine is a weaker base than either lysine or arginine. The imidazole ring, a five-membered ring structure containing two nitrogen atoms in the side chain of histidine, acts as a buffer (i.e., a stabilizer of hydrogen ion concentration) by binding hydrogen ions (H+) to the nitrogen atoms of the imidazole ring.



The remaining amino acids—phenylalanine, tyrosine, and tryptophan—have in common an aromatic structure; i.e., a benzene ring is present. These three amino acids are essential, and, while animals cannot synthesize the benzene ring itself, they can convert phenylalanine to tyrosine. 

Because these amino acids contain benzene rings, they can absorb ultraviolet light at wavelengths between 270 and 290 nanometres (nm; 1 nanometre = 10−9 metre = 10 angstrom units). Phenylalanine absorbs very little ultraviolet light; tyrosine and tryptophan, however, absorb it strongly and are responsible for the absorption band most proteins exhibit at 280–290 nanometres. This absorption is often used to determine the quantity of protein present in protein samples.

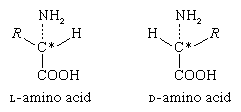
Most proteins contain only the amino acids described above; however, other amino acids occur in proteins in small amounts. For example, the collagen found in connective tissue contains, in addition to hydroxyproline, small amounts of hydroxylysine. Other proteins contain some monomethyl-, dimethyl-, or trimethyllysine—i.e., lysine derivativescontaining one, two, or three methyl groups (―CH3). The amount of these unusual amino acids in proteins, however, rarely exceeds 1 or 2 percent of the total amino acids.

Physicochemical properties of the amino acids

The physicochemical properties of a protein are determined by the analogous properties of the amino acids in it.

The α-carbon atom of all amino acids, with the exception of glycine, is asymmetric; this means that four different chemical entities (atoms or groups of atoms) are attached to it. As a result, each of the amino acids, except glycine, can exist in two different spatial, or geometric, arrangements (i.e., isomers), which are mirror images akin to right and left hands.

These isomers exhibit the property of opticalrotation. Optical rotation is the rotation of the plane of polarized light, which is composed of light waves that vibrate in one plane, or direction, only. Solutions of substances that rotate the plane of polarization are said to be optically active, and the degree of rotation is called the optical rotation of the solution. The direction in which the light is rotated is generally designed as plus, or d, for dextrorotatory (to the right), or as minus, or l, for levorotatory (to the left). Some amino acids are dextrorotatory, others are levorotatory. With the exception of a few small proteins (peptides) that occur in bacteria, the amino acids that occur in proteins are L-amino acids.



In bacteria, D-alanine and some other D-amino acids have been found as components of gramicidin and bacitracin. These peptides are toxic to other bacteria and are used in medicine as antibiotics. The D-alanine has also been found in some peptides of bacterial membranes.

**Britannica Quiz**

**Levels of structural organization in proteins**

**Primary structure**

Analytical and synthetic procedures reveal only the primary structure of the proteins—that is, the amino acid sequence of the peptide chains. They do not reveal information about the conformation (arrangement in space) of the peptide chain—that is, whether the peptide chain is present as a long straight thread or is irregularly coiled and folded into a globule. The configuration, or conformation, of a protein is determined by mutual attraction or repulsion of polar or nonpolar groups in the side chains (Rgroups) of the amino acids. The former have positive or negative charges in their side chains; the latter repel water but attract each other. Some parts of a peptide chain containing 100 to 200 amino acids may form a loop, or helix; others may be straight or form irregular coils.

The terms secondary, tertiary, and quaternary structure are frequently applied to the configuration of the peptide chain of a protein. A nomenclature committee of the International Union of Biochemistry (IUB) has defined these terms as follows: The primary structure of a protein is determined by its amino acid sequence without any regard for the arrangement of the peptide chain in space. The secondary structure is determined by the spatial arrangement of the main peptide chain without any regard for the conformation of side chains or other segments of the main chain. The tertiary structure is determined by both the side chains and other adjacent segments of the main chain, without regard for neighbouring peptide chains. Finally, the term quaternary structure is used for the arrangement of identical or different subunits of a large protein in which each subunit is a separate peptide chain.

**The isolation and determination of proteins**

Animal material usually contains large amounts of protein and lipids and small amounts of carbohydrate; in plants, the bulk of the dry matter is usually carbohydrate. If it is necessary to determine the amount of protein in a mixture of animal foodstuffs, a sample is converted to ammonium salts by boiling with sulfuric acid and a suitable inorganic catalyst, such as copper sulfate (Kjeldahl method). The method is based on the assumption that proteins contain 16 percent nitrogen, and that nonprotein nitrogen is present in very small amounts. The assumption is justified for most tissues from higher animals but not for insectsand crustaceans, in which a considerable portion of the body nitrogen is present in the form of chitin, a carbohydrate. Large amounts of nonprotein nitrogen are also found in the sap of many plants. In such cases, the precise quantitative analyses are made after the proteins have been separated from other biological compounds.

Proteins are sensitive to heat, acids, bases, organic solvents, and radiation exposure; for this reason, the chemical methods employed to purify organic compounds cannot be applied to proteins. Salts and molecules of small size are removed from protein solutions by dialysis—i.e., by placing the solution into a sac of semipermeable material, such as cellulose or acetylcellulose, which will allow small molecules to pass through but not large protein molecules, and immersing the sac in water or a salt solution. Small molecules can also be removed either by passing the protein solution through a column of resin that adsorbs only the protein or by gel filtration. In gel filtration, the large protein molecules pass through the column, and the small molecules are adsorbed to the gel.

Groups of proteins are separated from each other by salting out—i.e., the stepwise addition of sodium sulfate or ammonium sulfate to a protein solution. Some proteins, called globulins, become insoluble and precipitatewhen the solution is half-saturated with ammonium sulfate or when its sodium sulfate content exceeds about 12 percent. Other proteins, the albumins, can be precipitated from the supernatant solution (i.e., the solution remaining after a precipitation has taken place) by saturation with ammonium sulfate. Water-soluble proteins can be obtained in a dry state by freeze-drying (lyophilization), in which the protein solution is deep-frozen by lowering the temperature below −15 °C (5 °F) and removing the water; the protein is obtained as a dry powder.

Most proteins are insoluble in boiling water and are denatured by it—i.e., irreversibly converted into an insoluble material. Heat denaturation cannot be used with connective tissue because the principal structural protein, collagen, is converted by boiling water into water-soluble gelatin.**More From Britannica**

**What Is the Difference Between a Peptide and a Protein?**

Fractionation (separation into components) of a mixture of proteins of different molecular weight can be accomplished by gel filtration. The size of the proteins retained by the gel depends upon the properties of the gel. The proteins retained in the gel are removed from the column by solutions of a suitable concentration of salts and hydrogen ions.

Many proteins were originally obtained in crystalline form, but crystallinity is not proof of purity; many crystalline protein preparations contain other substances. Various tests are used to determine whether a protein preparation contains only one protein. The purity of a protein solution can be determined by such techniques as chromatography and gel filtration. In addition, a solution of pure protein will yield one peak when spun in a centrifuge at very high speeds (ultracentrifugation) and will migrate as a single band in electrophoresis(migration of the protein in an electrical field). After these methods and others (such as amino acid analysis) indicate that the protein solution is pure, it can be considered so. Because chromatography, ultracentrifugation, and electrophoresis cannot be applied to insoluble proteins, little is known about them; they may be mixtures of many similar proteins.

Very small (microheterogeneous) differences in some of the apparently pure proteins are known to occur. They are differences in the amino acid composition of otherwise identical proteins and are transmitted from generation to generation; i.e., they are genetically determined. For example, some humans have two hemoglobins, hemoglobin A and hemoglobin S, which differ in one amino acid at a specific site in the molecule. In hemoglobin A the site is occupied by glutamic acid and in hemoglobin S by valine. Refinement of the techniques of protein analysis has resulted in the discovery of other instances of microheterogeneity.

The quantity of a pure protein can be determined by weighing or by measuring the ultraviolet absorbancy at 280 nanometres. The absorbency at 280 nanometres depends on the content of tyrosine and tryptophan in the protein. Sometimes the slightly less sensitive biuret reaction, a purple colour given by alkaline protein solutions upon the addition of copper sulfate, is used; its intensity depends only on the number of peptide bonds per gram, which is similar in all proteins.

**Classification of proteins**

**Classification by solubility**

After two German chemists, Emil Fischer and Franz Hofmeister, independently stated in 1902 that proteins are essentially polypeptides consisting of many amino acids, an attempt was made to classify proteins according to their chemical and physical properties, because the biological function of proteins had not yet been established. (The protein character of enzymeswas not proved until the 1920s.) Proteins were classified primarily according to their solubility in a number of solvents. This classification is no longer satisfactory, however, because proteins of quite different structure and function sometimes have similar solubilities; conversely, proteins of the same function and similar structure sometimes have different solubilities. The terms associated with the old classification, however, are still widely used. They are defined below.

Albumins are proteins that are soluble in waterand in water half-saturated with ammonium sulfate. On the other hand, globulins are salted out (i.e., precipitated) by half-saturation with ammonium sulfate. Globulins that are soluble in salt-free water are called pseudoglobulins; those insoluble in salt-free water are euglobulins. Both prolamins and glutelins, which are plant proteins, are insoluble in water; the prolamins dissolve in 50 to 80 percent ethanol, the glutelins in acidified or alkaline solution. The term protamine is used for a number of proteins in fish sperm that consist of approximately 80 percent arginine and therefore are strongly alkaline. Histones, which are less alkaline, apparently occur only in cellnuclei, where they are bound to nucleic acids. The term scleroproteins has been used for the insoluble proteins of animal organs. They include keratin, the insoluble protein of certain epithelial tissues such as the skin or hair, and collagen, the protein of the connective tissue. A large group of proteins has been called conjugated proteins, because they are complex molecules of protein consisting of protein and nonprotein moieties. The nonprotein portion is called the prosthetic group. Conjugatedproteins can be subdivided into mucoproteins, which, in addition to protein, contain carbohydrate; lipoproteins, which contain lipids; phosphoproteins, which are rich in phosphate; chromoproteins, which contain pigments such as iron-porphyrins, carotenoids, bile pigments, and melanin; and finally, nucleoproteins, which contain nucleic acid.

The weakness of the above classification lies in the fact that many, if not all, globulins contain small amounts of carbohydrate; thus there is no sharp borderline between globulins and mucoproteins. Moreover, the phosphoproteins do not have a prosthetic group that can be isolated; they are merely proteins in which some of the hydroxyl groups of serine are phosphorylated (i.e., contain phosphate). Finally, the globulins include proteins with quite different roles—enzymes, antibodies, fibrous proteins, and contractile proteins.

**Classification by biological functions**

In view of the unsatisfactory state of the old classification, it is preferable to classify the proteins according to their biological function. Such a classification is far from ideal, however, because one protein can have more than one function. The contractile protein myosin, for example, also acts as an ATPase (adenosine triphosphatase), an enzyme that hydrolyzes adenosine triphosphate (removes a phosphate group from ATP by introducing a water molecule). Another problem with functional classification is that the definite function of a protein frequently is not known. A protein cannot be called an enzyme as long as its substrate (the specific compound upon which it acts) is not known. It cannot even be tested for its enzymatic action when its substrate is not known.

**Milk proteins**

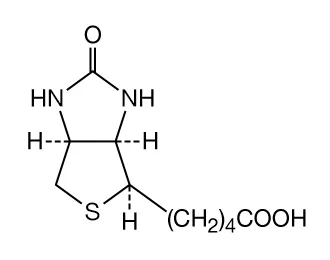
Milk contains the following: an albumin, α-lactalbumin; a globulin, beta-lactoglobulin; and a phosphoprotein, casein. If acid is added to milk, casein precipitates. The remaining watery liquid (the supernatant solution), or whey, contains α-lactalbumin and β-lactoglobulin. Both have been obtained in crystalline form; in bovine milk, their molecular weights are approximately 14,000 and 18,400, respectively. Lactoglobulin also occurs as a dimer of molecular weight 37,000. Genetic variations can produce small variations in the amino acidcomposition of lactoglobulin. The amino acid composition and the tertiary structure of lactalbumin resemble that of lysozyme, an egg protein.

Casein is precipitated not only by the addition of acid but also by the action of the enzymerennin, which is found in gastric juice. Rennin from calf stomachs is used to precipitate casein, from which cheese is made. Milk fat precipitates with casein; milk sugar, however, remains in the supernatant (whey). Casein is a mixture of several similar phosphoproteins, called α-, β-, γ−, and κ-casein, all of which contain some serine side chains combined with phosphoric acid. Approximately 75 percent of casein is α-casein. Cystine has been found only in κ-casein. In milk, casein seems to form polymeric globules (micelles) with radially arranged monomers, each with a molecular weight of 24,000; the acidic side chains occur predominantly on the surface of the micelle, rather than inside.

Egg proteins

About 50 percent of the proteins of egg white are composed of ovalbumin, which is easily obtained in crystals. Its molecular weight is 46,000 and its amino acid composition differs from that of serum albumin. Other proteins of egg white are conalbumin, lysozyme, ovoglobulin, ovomucoid, and avidin. Lysozyme is an enzyme that hydrolyzes the carbohydrates found in the capsules certain bacteria secretearound themselves; it causes lysis (disintegration) of the bacteria. The molecular weight of lysozyme is 14,100. Its three-dimensional structure is similar to that of α-lactalbumin, which stimulates the formation of lactose by the enzyme lactose synthetase. Lysozyme has also been found in the urine of patients suffering from leukemia, meningitis, and renal disease.

Avidin is a glycoprotein that combines specifically with biotin, a vitamin. In animals fed large amounts of raw egg white, the action of avidin results in “egg-white injury.” The molecular weight of avidin, which forms a tetramer, is 16,200. Its amino acid sequence is known.



Egg-yolk proteins contain a mixture of lipoproteins and livetins. The latter are similar to serum albumin, α-globulin, and β-globulin. The yolk also contains a phosphoprotein, phosvitin. Phosvitin, which has also been found in fish sperm, has a molecular weight of 40,000 and an unusual amino acid composition; one third of its amino acids are phosphoserine.

**Protamines and histones**

Protamines are found in the sperm cells of fish. The most thoroughly investigated protamines are salmine from salmon sperm and clupeine from herring sperm. The protamines are bound to deoxyribonucleic acid (DNA), forming nucleoprotamines. The amino acid composition of the protamines is simple; they contain, in addition to large amounts of arginine, small amounts of five or six other amino acids. The composition of the salmine molecule, for example, is: Arg51, Ala4, Val4, Ile1, Pro7, and Ser6, in which the subscript numbers indicate the number of each amino acid in the molecule. Because of the high arginine content, the isoelectric points of the protamines are at pH values of 11 to 12; i.e., the protamines are alkaline. The molecular weights of salmine and clupeine are close to 6,000. All of the protamines investigated thus far are mixtures of several similar proteins.

The histones are less basic than the protamines. They contain high amounts of either lysine or arginine and small amounts of aspartic acidand glutamic acid. Histones occur in combination with DNA as nucleohistones in the nuclei of the body cells of animals and plants, but not in animal sperm. The molecular weights of histones vary from 10,000 to 22,000. In contrast to the protamines, the histones contain most of the 20 amino acids, with the exception of tryptophan and the sulfur-containing ones. Like the protamines, histone preparations are heterogeneousmixtures. The amino acid sequence of some of the histones has been determined.

**Plant proteins**

Plant proteins, mostly globulins, have been obtained chiefly from the protein-rich seeds of cereals and legumes. Small amounts of albumins are found in seeds. The best known globulins, insoluble in water, can be extracted from seeds by treatment with 2 to 10 percent solutions of sodium chloride. Many plant globulins have been obtained in crystalline form; they include edestin from hemp, molecular weight 310,000; amandin from almonds, 330,000; concanavalin A (42,000) and B (96,000); and canavalin (113,000) from jack beans. They are polymers of smaller subunits; edestin, for example, is a hexamer of a subunit with a molecular weight of 50,000, and concanavalin B a trimer of a subunit with a molecular weight of 30,000. After extraction of lipids from cereal seeds by ether and alcohol, further extraction with water containing 50 to 80 percent of alcohol yields proteins that are insoluble in water but soluble in water–ethanol mixtures and have been called prolamins. Their solubility in aqueous ethanol may result from their high proline and glutamine content. Gliadin, the prolamin from wheat, contains 14 grams of proline and 46 grams of glutamic acidin 100 grams of protein; most of the glutamic acid is in the form of glutamine. The total amounts of the basic amino acids (arginine, lysine, and histidine) in gliadin are only 5 percent of the weight of gliadin. Because the glysine content is either low or nonexistent, human populations dependent on grain as a sole protein source suffer from lysine deficiency.

**Conjugated proteins**

**Combination of proteins with prosthetic groups**

The link between a protein molecule and its prosthetic group is a covalent bond (an electron-sharing bond) in the glycoproteins, the biliproteins, and some of the heme proteins. In lipoproteins, nucleoproteins, and some heme proteins, the two components are linked by noncovalent bonds; the bonding results from the same forces that are responsible for the tertiary structure of proteins: hydrogen bonds, salt bridges between positively and negatively charged groups, disulfide bonds, and mutual interaction of hydrophobic groups. In the metalloproteins (proteins with a metal element as a prosthetic group), the metal ion usually forms a centre to which various groups are bound.

Some of the conjugated proteins have been mentioned in preceding sections because they occur in the blood serum, in milk, and in eggs; others are discussed below in sections dealing with respiratory proteins and enzymes.

**Mucoproteins and glycoproteins**

The prosthetic groups in mucoproteins and glycoproteins are oligosaccharides(carbohydrates consisting of a small number of simple sugar molecules) usually containing from four to 12 sugar molecules; the most common sugars are galactose, mannose, glucosamine, and galactosamine. Xylose, fucose, glucuronic acid, sialic acid, and other simple sugars sometimes also occur. Some mucoproteins contain 20 percent or more of carbohydrate, usually in several oligosaccharides attached to different parts of the peptide chain. The designationmucoprotein is used for proteins with more than 3 to 4 percent carbohydrate; if the carbohydrate content is less than 3 percent, the protein is sometimes called a glycoprotein or simply a protein.

Mucoproteins, highly viscous proteins originally called mucins, are found in saliva, in gastric juice, and in other animal secretions. Mucoproteins occur in large amounts in cartilage, synovial fluid (the lubricating fluid of joints and tendons), and egg white. The mucoprotein of cartilage is formed by the combination of collagen with chondroitinsulfuric acid, which is a polymer of either glucuronic or iduronic acid and acetylhexosamine or acetylgalactosamine. It is not yet clear whether or not chondroitinsulfate is bound to collagen by covalent bonds.

**Lipoproteins and proteolipids**

The bond between the protein and the lipidportion of lipoproteins and proteolipids is a noncovalent one. It is thought that some of the lipid is enclosed in a meshlike arrangement of peptide chains and becomes accessible for reaction only after the unfolding of the chains by denaturing agents. Although lipoproteins in the α- and β-globulin fraction of blood serum are soluble in water (but insoluble in organic solvents), some of the brain lipoproteins, because they have a high lipid content, are soluble in organic solvents; they are called proteolipids. The β-lipoprotein of human blood serum is a macroglobulin with a molecular weight of about 1,300,000, 70 percent of which is lipid; of the lipid, about 30 percent is phospholipid and 40 percent cholesterol and compounds derived from it. Because of their lipid content, the lipoproteins have the lowest density (mass per unit volume) of all proteins and are usually classified as low- and high-density lipoproteins (LDL and HDL).

Coloured lipoproteins are formed by the combination of protein with carotenoids. Crustacyanin, the pigment of lobsters, crayfish, and other crustaceans, contains astaxanthin, which is a compound derived from carotene. Among the most interesting of the coloured lipoproteins are the pigments of the retina of the eye. They contain retinal, which is a compound derived from carotene and which is formed by the oxidation of vitamin A. In rhodopsin, the red pigment of the retina, the aldehyde group (―CHO) of retinal forms a covalent bond with an amino (―NH2) group of opsin, the protein carrier. Colour vision is mediated by the presence of several visual pigments in the retina that differ from rhodopsin either in the structure of retinal or in that of the protein carrier.

**Heme proteins and other chromoproteins**

Although the heme proteins contain iron, they are usually not classified as metalloproteins, because their prosthetic group is an iron-porphyrin complex in which the iron is bound very firmly. The intense red or brown colour of the heme proteins is not caused by iron but by porphyrin, a complex cyclic structure. All porphyrin compounds absorb light intensely at or close to 410 nanometres. Porphyrin consists of four pyrrole rings (five-membered closed structures containing one nitrogen and four carbon atoms) linked to each other by methine groups (―CH=). The iron atom is kept in the centre of the porphyrin ring by interaction with the four nitrogen atoms. The iron atom can combine with two other substituents; in oxyhemoglobin, one substituent is a histidine of the protein carrier, the other is an oxygenmolecule. In some heme proteins, the protein is also bound covalently to the side chains of porphyrin. Heme proteins are described below (see Respiratory proteins).

The chromoprotein melanin, a pigment found in dark skin, dark hair, and melanotic tumours, occurs in every major group of living organisms and appears to be remarkably diverse in structure. In humans, melanin produced by melanocytes may be dark brown (eumelanin) or pale red or yellowish (phaeomelanin). The different types are synthesized via different pathways, though they share the same initial step—the oxidation of tyrosine.

Green chromoproteins called biliproteins are found in many insects, such as grasshoppers, and also in the eggshells of many birds. The biliproteins are derived from the bile pigment biliverdin, which in turn is formed from porphyrin; biliverdin contains four pyrrole rings and three of the four methine groups of porphyrin. Large amounts of biliproteins have been found in red algae and blue-green algae; the red protein is called phycoerythrin, the blue one phycocyanobilin.

**Protein hormones**

Some hormones that are products of endocrineglands are proteins or peptides, others are steroids. (The origin of hormones, their physiological role, and their mode of action are dealt with in the article hormone.) None of the hormones has any enzymatic activity. Each has a target organ in which it elicits some biological action—e.g., secretion of gastric or pancreatic juice, production of milk, production of steroidhormones. The mechanism by which the hormones exert their effects is not fully understood. Cyclic adenosine monophosphate is involved in the transmittance of the hormonal stimulus to the cells whose activity is specifically increased by the hormone.