**Classification Of Proteins** <https://www.britannica.com/science/protein/Special-structure-and-function-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 enzymes was 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 water and 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 cell nuclei, 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. Conjugated proteins 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.

Special Structure And Function Of Proteins

Despite its weaknesses, a functional classification is used here in order to demonstrate, whenever possible, the correlation between the structure and function of a protein. The structural, fibrous proteins are presented first, because their structure is simpler than that of the globular proteins and more clearly related to their function, which is the maintenance of either a rigid or a flexible structure.

Collagen is the structural protein of bones, tendons, ligaments, and skin. For many years collagen was considered to be insoluble in water. Part of the collagen of calf skin, however, can be extracted with citrate buffer at pH 3.7. A precursor of collagen called procollagen is converted in the body into collagen. Procollagen has a molecular weight of 120,000. Cleavage of one or a few peptide bonds of procollagen yields collagen, which has three subunits, each with a molecular weight of 95,000; therefore, the molecular weight of collagen is 285,000 (3 × 95,000). The three subunits are wound as spirals around an elongated straight axis. The length of each subunit is 2,900 angstroms, and its diameter is approximately 15 angstroms. The three chains are staggered, so that the trimer has no definite terminal limits.

Collagen differs from all other proteins in its high content of proline and hydroxyproline. Hydroxyproline does not occur in significant amounts in any other protein except elastin. Most of the proline in collagen is present in the sequence glycine–proline-X, in which X is frequently alanine or hydroxyproline. Collagen does not contain cystine or tryptophan and therefore cannot substitute for other proteins in the diet. The presence of proline causes kinks in the peptide chain and thus reduces the length of the amino acid unit from 3.7 angstroms in the extended chain of the β-structure to 2.86 angstroms in the collagen chain. In the intertwined triple helix, the glycines are inside, close to the axis; the prolines are outside.

Native collagen resists the action of trypsin but is hydrolyzed by the bacterial enzyme collagenase. When collagen is boiled with water, the triple helix is destroyed, and the subunits are partially hydrolyzed; the product is gelatin. The unfolded peptide chains of gelatin trap large amounts of water, resulting in a hydrated molecule.

When collagen is treated with tannic acid or with chromium salts, cross links form between the collagen fibres, and it becomes insoluble; the conversion of hide into leather is based on this tanning process. The tanned material is insoluble in hot water and cannot be converted to gelatin. On exposure to water at 62° to 63° C (144° to 145° F), however, the cross links formed by the tanning agents collapse, and the leather contracts irreversibly to about one-third its original volume.

Collagen seems to undergo an aging process in living organisms that may be caused by the formation of cross links between collagen fibres. They are formed by the conversion of some lysine side chains to aldehydes (compounds with the general structure RCHO), and the combination of the aldehydes with the ε-amino groups of intact lysine side chains. The protein elastin, which occurs in the elastic fibres of connective tissue, contains similar cross links and may result from the combination of collagen fibres with other proteins. When cross-linked collagen or elastin is degraded, products of the cross-linked lysine fragments, called desmosins and isodesmosins, are formed.

Keratin

Keratin, the structural protein of epithelial cells in the outermost layers of the skin, has been isolated from hair, nails, hoofs, and feathers. Keratin is completely insoluble in cold or hot water; it is not attacked by proteolytic enzymes (i.e., enzymes that break apart, or lyse, protein molecules), and therefore cannot replace proteins in the diet. The great stability of keratin results from the numerous disulfide bonds of cystine. The amino acid composition of keratin differs from that of collagen. Cystine may account for 24 percent of the total amino acids. The peptide chains of keratin are arranged in approximately equal amounts of antiparallel and parallel pleated sheets, in which the peptide chains are linked to each other by hydrogen bonds between the carbonyl and imino groups.

Reduction of the disulfide bonds to sulfhydryl groups results in dissociation of the peptide chains, the molecular weight of which is 25,000 to 28,000 each. The formation of permanent waves in the beauty treatment of hair is based on partial reduction of the disulfide bonds of hair keratin by thioglycol, or some other mild reducing agent, and subsequent oxidation of the sulfhydryl groups (―SH) in the reoriented hair to disulfide bonds (―S―S―) by exposure to the oxygen of the air.

The length of keratin fibres depends on their water content. They can bind approximately 16 percent of water; this hydration is accompanied by an increase in the length of the fibres of 10 to 12 percent.

The most thoroughly investigated keratin is hair keratin, particularly that of wool. It consists of a mixture of peptides with high and low cystine content. When wool is heated in water to about 90° C (190° F), it shrinks irreversibly. This is attributed to the breakage of hydrogen bonds and other noncovalent bonds; disulfide bonds do not seem to be affected.

Others

The most thoroughly investigated scleroprotein has been fibroin, the insoluble material of silk. The raw silk comprising the cocoon of the silkworm consists of two proteins. One, sericin, is soluble in hot water; the other, fibroin, is not. The amino acid composition of the latter differs from that of all other proteins. It contains large amounts of glycine, alanine, tyrosine, and serine; small amounts of the other amino acids; and no sulfur-containing ones. The peptide chains are arranged in antiparallel β-structures. Fibroin is partly soluble in concentrated solutions of lithium thiocyanate or in mixtures of cupric salts and ethylene diamine. Such solutions contain a protein of molecular weight 170,000, which is a dimer of two subunits.

Little is known about either the scleroproteins of the marine sponges or the insoluble proteins of the cellular membranes of animal cells. Some of the membranes are soluble in detergents; others, however, are detergent-insoluble.

The muscle proteins

The total amount of muscle proteins in mammals, including humans, exceeds that of any other protein. About 40 percent of the body weight of a healthy human adult weighing about 70 kilograms (150 pounds) is muscle, which is composed of about 20 percent muscle protein. Thus, the human body contains about 5 to 6 kilograms (11 to 13 pounds) of muscle protein. An albumin-like fraction of these proteins, originally called myogen, contains various enzymes—phosphorylase, aldolase, glyceraldehyde phosphate dehydrogenase, and others; it does not seem to be involved in contraction. The globulin fraction contains myosin, the contractile protein, which also occurs in blood platelets, small bodies found in blood. Similar contractile substances occur in other contractile structures; for example, in the cilia or flagella (whiplike organs of locomotion) of bacteria and protozoans. In contrast to the scleroproteins, the contractile proteins are soluble in salt solutions and susceptible to enzymatic digestion.

The energy required for muscle contraction is provided by the oxidation of carbohydrates or lipids. The term mechanochemical reaction has been used for this conversion of chemical into mechanical energy. The molecular process underlying the reaction is known to involve the fibrous muscle proteins, the peptide chains of which undergo a change in conformation during contraction.

Myosin, which can be removed from fresh muscle by adding it to a chilled solution of dilute potassium chloride and sodium bicarbonate, is insoluble in water. Myosin, solutions of which are highly viscous, consists of an elongated—probably double-stranded—peptide chain, which is coiled at both ends in such a way that a terminal globule is formed. The length of the molecule is approximately 160 nanometres and its average diameter 2.6 nanometres. The equivalent weight of each of the two terminal globules is approximately 30,000; the molecular weight of myosin is close to 500,000. Trypsin splits myosin into large fragments called meromyosin. Myosin contains many amino acids with positively and negatively charged side chains; they form 18 and 16 percent, respectively, of the total number of amino acids. Myosin catalyzes the hydrolytic cleavage of ATP (adenosine triphosphate). A smaller protein with properties similar to those of myosin is tropomyosin. It has a molecular weight of 70,000 and dimensions of 45 by 2 nanometres. More than 90 percent of its peptide chains are present in the α-helix form.

Myosin combines easily with another muscle protein called actin, the molecular weight of which is about 50,000; it forms 12 to 15 percent of the muscle proteins. Actin can exist in two forms—one, G-actin, is globular; the other, F-actin, is fibrous. Actomyosin is a complex molecule formed by one molecule of myosin and one or two molecules of actin. In muscle, actin and myosin filaments are oriented parallel to each other and to the long axis of the muscle. The actin filaments are linked to each other lengthwise by fine threads called S filaments. During contraction the S filaments shorten, so that the actin filaments slide toward each other, past the myosin filaments, thus causing a shortening of the muscle (for a detailed description of the process, see muscle: Striated muscle).

Fibrinogen and fibrin

Fibrinogen, the protein of the blood plasma, is converted into the insoluble protein fibrin during the clotting process. The fibrinogen-free fluid obtained after removal of the clot, called blood serum, is blood plasma minus fibrinogen. The fibrinogen content of the blood plasma is 0.2 to 0.4 percent.

Fibrinogen can be precipitated from the blood plasma by half-saturation with sodium chloride. Fibrinogen solutions are highly viscous and show strong flow birefringence. In electron micrographs the molecules appear as rods with a length of 47.5 nanometres and a diameter of 1.5 nanometres; in addition, two terminal and a central nodule are visible. The molecular weight is 340,000. An unusually high percentage, about 36 percent, of the amino acid side chains are positively or negatively charged.

The clotting process is initiated by the enzyme thrombin, which catalyzes the breakage of a few peptide bonds of fibrinogen; as a result, two small fibrinopeptides with molecular weights of 1,900 and 2,400 are released. The remainder of the fibrinogen molecule, a monomer, is soluble and stable at pH values less than 6 (i.e., in acid solutions). In neutral solution (pH 7) the monomer is converted into a larger molecule, insoluble fibrin; this results from the formation of new peptide bonds. The newly formed peptide bonds form intermolecular and intramolecular cross links, thus giving rise to a large clot, in which all molecules are linked to each other. Clotting, which takes place only in the presence of calcium ions, can be prevented by compounds such as oxalate or citrate, which have a high affinity for calcium ions.

Albumins, globulins, and other soluble proteins

The blood plasma, the lymph, and other animal fluids usually contain one to seven grams of protein per 100 millilitres of fluid, which includes small amounts of hundreds of enzymes and a large number of protein hormones. The discussion below is limited largely to the proteins that occur in large amounts and can be easily isolated from the body fluids.

Proteins of the blood serum

Human blood serum contains about 7 percent protein, two-thirds of which is in the albumin fraction; the other third is in the globulin fraction. Electrophoresis of serum reveals a large albumin peak and three smaller globulin peaks, the alpha-, beta-, and gamma-globulins. The amounts of alpha-, beta-, and gamma-globulin in normal human serum are approximately 1.5, 1.9, and 1.1 percent, respectively. Each globulin fraction is a mixture of many different proteins, as has been demonstrated by immunoelectrophoresis. In this method, serum from an animal (e.g., a rabbit) injected with human serum is allowed to diffuse into the four protein bands—albumin, alpha-, beta-, and gamma-globulin—obtained from the electrophoresis of human serum. Because the animal has previously been injected with human serum, its blood contains antibodies (substances formed in response to a foreign substance introduced into the body) against each of the human serum proteins; each antibody combines with the serum protein (antigen) that caused its formation in the animal. The result is the formation of about 20 regions of insoluble antigen-antibody precipitate, which appear as white arcs in the transparent gel of the electrophoresis medium. Each region corresponds to a different human serum protein.

Serum albumin is much less heterogeneous (i.e., contains fewer distinct proteins) than are the globulins; in fact, it is one of the few serum proteins that can be obtained in a crystalline form. Serum albumin combines easily with many acidic dyes (e.g., Congo red and methyl orange); with bilirubin, the yellow bile pigment; and with fatty acids. It seems to act, in living organisms, as a carrier for certain biological substances. Present in blood serum in relatively high concentration, serum albumin also acts as a protective colloid, a protein that stabilizes other proteins. Albumin (molecular weight of 68,000) has a single free sulfhydryl (―SH) group, which on oxidation forms a disulfide bond with the sulfhydryl group of another serum albumin molecule, thus forming a dimer. The isoelectric point of serum albumin is pH 4.7.

The alpha-globulin fraction of blood serum is a mixture of several conjugated proteins. The best known are an α-lipoprotein (combination of lipid and protein) and two mucoproteins (combinations of carbohydrate and protein). One mucoprotein is called orosomucoid, or α1-acid glycoprotein; the other is called haptoglobin because it combines specifically with globin, the protein component of hemoglobin. Haptoglobin contains about 20 percent carbohydrate. The beta-globulin fraction of serum contains, in addition to lipoproteins and mucoproteins, two metal-binding proteins, transferrin and ceruloplasmin, which bind iron and copper, respectively. They are the principal iron and copper carriers of the blood.

The gamma-globulins are the most heterogeneous globulins. Although most have a molecular weight of approximately 150,000, that of some, called macroglobulins, is as high as 800,000. Because typical antibodies are of the same size and exhibit the same electrophoretic behaviour as γ-globulins, they are called immunoglobulins. The designation IgM or gamma M (γM) is used for the macroglobulins; the designation IgG or gamma G (γG) is used for γ−globulins of molecular weight 150,000.

The four-chain structure of an antibody, or immunoglobulin, moleculeThe basic unit is composed of two identical light (L) chains and two identical heavy (H) chains, which are held together by disulfide bonds to form a flexible Y shape. Each chain is composed of a variable (V) region and a constant (C) region.

The four-chain structure of an antibody, or immunoglobulin, moleculeThe basic unit is composed of two identical light (L) chains and two identical heavy (H) chains, which are held together by disulfide bonds to form a flexible Y shape. Each chain is composed of a variable (V) region and a constant (C) region.

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 acid composition 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 enzyme rennin, 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 secrete around 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.Chemical compound diagram for biotin

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 acid and 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 heterogeneous mixtures. 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 acid in 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 designation mucoprotein 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 lipid portion 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.

Metalloproteins

Proteins in which heavy metal ions are bound directly to some of the side chains of histidine, cysteine, or some other amino acid are called metalloproteins. Two metalloproteins, transferrin and ceruloplasmin, occur in the globulin fractions of blood serum; they act as carriers of iron and copper, respectively. Transferrin has a molecular weight of about 80,000 and consists of two identical subunits, each of which contains one ferric ion (Fe3+) that seems to be bound to tyrosine. Several genetic variants of transferrin are known to occur in humans. Another iron protein, ferritin, which contains 20 to 22 percent iron, is the form in which iron is stored in animals; it has been obtained in crystalline form from liver and spleen. A molecule consisting of 20 subunits, its molecular weight is approximately 480,000. The iron can be removed by reduction from the ferric (Fe3+) to the ferrous (Fe2+) state. The iron-free protein, apoferritin, is synthesized in the body before the iron is incorporated.

Green plants and some photosynthetic and nitrogen-fixing bacteria (i.e., bacteria that convert atmospheric nitrogen, N2, into amino acids and proteins) contain various ferredoxins. They are small proteins containing 50 to 100 amino acids and a chain of iron and disulfide units (FeS2), in which some of the sulfur atoms are contributed by cysteine; others are sulfide ions (S2−). The number of FeS2 units per ferredoxin molecule varies from five in the ferredoxin of spinach to 10 in the ferredoxin of certain bacteria. Ferredoxins act as electron carriers in photosynthesis and in nitrogen fixation.

Ceruloplasmin is a copper-containing globulin that has a molecular weight of 151,000; the molecule consists of eight subunits, each containing one copper ion. Ceruloplasmin is the principal carrier of copper in organisms, although copper can also be transported by the iron-containing globulin transferrin. Another copper-containing protein, copper-zinc superoxide dismutase (formerly known as erythrocuprein), has been isolated from red blood cells; it has also been found in the liver and in the brain. The molecule, which consists of two subunits of similar size, contains copper ions and zinc ions. Because of their copper content, ceruloplasmin and copper-zinc superoxide dismutase possess catalytic activity in oxidation-reduction reactions.

Many animal enzymes contain zinc ions, which are usually bound to the sulfur of cysteine. Horse kidneys contain the protein metallothionein, which contain zinc and cadmium; both are bound to sulfur. A vanadium-protein complex (hemovanadin) has been found in surprisingly high amounts in yellowish-green cells (vanadocytes) of tunicates, which are marine invertebrates.

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 oxygen molecule. 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.

Nucleoproteins

When a protein solution is mixed with a solution of a nucleic acid, the phosphoric acid component of the nucleic acid combines with the positively charged ammonium groups (―NH3+) of the protein to form a protein–nucleic acid complex. The nucleus of a cell contains predominantly deoxyribonucleic acid (DNA) and the cytoplasm predominantly ribonucleic acid (RNA); both parts of the cell also contain protein. Protein–nucleic acid complexes, therefore, form in living cells.

The only nucleoproteins for which some evidence for specificity exists are nucleoprotamines, nucleohistones, and some RNA and DNA viruses. The nucleoprotamines are the form in which protamines occur in the sperm cells of fish; the histones of the thymus and of pea seedlings and other plant material apparently occur predominantly as nucleohistones. Both nucleoprotamines and nucleohistones contain only DNA.

Some of the simplest viruses consist of a specific RNA, which is coated by protein. One of the best known RNA viruses, tobacco mosaic virus (TMV), has the shape of a rod. RNA comprises only 5.1 percent of the mass of the virus. The complete sequence of the virus protein, which consists of about 2,130 identical peptide chains, each containing 158 amino acids, has been determined. The protein is arranged in a spiral around the RNA core.

Schematic structure of the tobacco mosaic virus. The cutaway section shows the helical ribonucleic acid associated with protein molecules in a ratio of three nucleotides per protein molecule.

Schematic structure of the tobacco mosaic virus. The cutaway section shows the helical ribonucleic acid associated with protein molecules in a ratio of three nucleotides per protein molecule.

DNA has been found in most bacterial viruses (bacteriophages) and in some animal viruses. As in TMV, the core of DNA is surrounded by protein. Phage protein is a mixture of enzymes and therefore cannot be considered as the protein portion of only one nucleoprotein.

Respiratory proteins

Hemoglobin

Hemoglobin is the oxygen carrier in all vertebrates and some invertebrates. In oxyhemoglobin (HbO2), which is bright red, the ferrous ion (Fe2+) is bound to the four nitrogen atoms of porphyrin; the other two substituents are an oxygen molecule and the histidine of globin, the protein component of hemoglobin. Deoxyhemoglobin (deoxy-Hb), as its name implies, is oxyhemoglobin minus oxygen (i.e., reduced hemoglobin); it is purple in colour. Oxidation of the ferrous ion of hemoglobin yields a ferric compound, methemoglobin, sometimes called hemiglobin or ferrihemoglobin. The oxygen of oxyhemoglobin can be displaced by carbon monoxide, for which hemoglobin has a much greater affinity, preventing oxygen from reaching the body tissues.

The hemoglobins of all mammals, birds, and many other vertebrates are tetramers of two α- and two β-chains. The molecular weight of the tetramer is 64,500; the molecular weight of the α- and β-chains is approximately 16,100 each, and the four subunits are linked to each other by noncovalent interactions. If hemin (the ferric porphyrin component) is removed from globin (the protein component), two molecules of globin, each consisting of one α- and one β-chain, are obtained; the molecular weight of globin is 32,200. In contrast to hemoglobin, globin is an unstable protein that is easily denatured. If native globin is incubated with a solution of hemin at pH values of 8 to 9, native hemoglobin is reconstituted. Myoglobin, the red pigment of mammalian muscles, is a monomer with a molecular weight of 16,000.

The mammalian hemoglobins differ from each other in their amino acid composition and therefore in their secondary and tertiary structure. Rat and horse hemoglobins crystallize very easily, but those of humans, cattle, and sheep, because they are more soluble, are difficult to crystallize. The shape of hemoglobin crystals varies in different species; moreover, decomposition and denaturation occur at different rates in different species. It was also found that the blood of human newborns contains two different hemoglobins: about 20 percent of their hemoglobin is an adult hemoglobin (hemoglobin A) and 80 percent is a fetal hemoglobin (hemoglobin F). Hemoglobin F persists in the infant for the first seven months of life. The same hemoglobin F has also been found in the blood of patients suffering from thalassemia, an anemia with a high incidence in regions surrounding the Mediterranean Sea. Hemoglobin F contains, as does hemoglobin A, two α-chains; the two β-chains, however, have been replaced by two quite different γ-chains. When the technique of electrophoresis was first applied to the hemoglobin of blacks suffering from sickle cell anemia in 1949, a new hemoglobin (hemoglobin S) was discovered. More than 200 different human hemoglobins have been discovered since. They differ from normal hemoglobin A in the amino acid composition of either the α- or the β-chain.

The hemoglobins of some of the lowest fishes are monomers containing one iron atom per molecule. Hemoglobin-like respiratory proteins have been found in some invertebrates. The red hemoglobin of insects, mollusks, and protozoans is called erythrocruorin. It differs from vertebrate hemoglobin by its high molecular weight.

Although green plants contain no hemoglobin, a red protein, called leghemoglobin, has been discovered in the root nodules of leguminous plants. It seems to be produced by the nitrogen-fixing bacteria of the root nodules and may be involved in the reduction of atmospheric nitrogen to ammonia and amino acids.

Other respiratory proteins

A green respiratory protein, chlorocruorin, has been found in the blood of marine worms in the genera Serpula and Spirographis. It has the same high molecular weight as erythrocruorin but differs from hemoglobin in its prosthetic group. A red metalloprotein, hemerythrin, acts as a respiratory protein in marine worms of the phylum Sipuncula. The molecule consists of eight subunits with a molecular weight of 13,500 each. Hemerythrin contains no porphyrins and therefore is not a heme protein.

A metalloprotein containing copper is the respiratory protein of crustaceans (shrimps, crabs, etc.) and of some gastropods (snails). The protein, called hemocyanin, is pale yellow when not combined with oxygen, and blue when combined with oxygen. The molecular weights of hemocyanins vary from 300,000 to 9,000,000. Each animal investigated thus far apparently has a species-specific hemocyanin.

Protein hormones

Some hormones that are products of endocrine glands 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 steroid hormones. 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.

Hormones of the thyroid gland

Thyroglobulin, the active groups of which are two molecules of the iodine-containing compound thyroxine, has a molecular weight of 670,000. Thyroglobulin also contains thyroxine with two and three iodine atoms instead of four and tyrosine with one and two iodine atoms. Injection of the hormone causes an increase in metabolism; lack of it results in a slowdown.

Another hormone, calcitonin, which lowers the calcium level of the blood, occurs in the thyroid gland. The amino acid sequences of calcitonin from pig, beef, and salmon differ from human calcitonin in some amino acids. All of them, however, have the half-cystines (C) and the prolinamide (P) in the same position.Proteins. Formula 7: The amino acid sequence of human calcitonin. At the left end the line represents the disulfide bond. At the right end (CONH2) indicates that the C terminal proline is present as prolinamide.

Parathyroid hormone (parathormone), produced in small glands that are embedded in or lie behind the thyroid gland, is essential for maintaining the calcium level of the blood. A decrease in its production results in hypocalcemia (a reduction of calcium levels in the bloodstream below the normal range). Bovine parathormone has a molecular weight of 8,500; it contains no cystine or cysteine and is rich in aspartic acid, glutamic acid, or their amides.

Hormones of the pancreas

Although the amino acid structure of insulin has been known since 1949, repeated attempts to synthesize it gave very poor yields because of the failure of the two peptide chains to combine forming the correct disulfide bridge. The ease of the biosynthesis of insulin is explained by the discovery in the pancreas of proinsulin, from which insulin is formed. The single peptide chain of proinsulin loses a peptide consisting of 33 amino acids and called the connecting peptide, or C peptide, during its conversion to insulin. The disulfide bridges of proinsulin connect the A and B chains.Proteins. Formula 8: The amino acid sequence of porcine proinsulin. The arrows indicate the direction from the N terminus of the B chain to the C terminus of the A chain. ...

In aqueous solutions, insulin exists predominantly as a complex of six subunits, each of which contains an A and a B chain. The insulins of several species have been isolated and analyzed; their amino acid sequences have been found to differ somewhat, but all apparently contain the same disulfide bridges between the two chains.

Although the injection of insulin lowers the blood sugar, administration of glucagon, another pancreas hormone, raises the blood sugar level. Glucagon consists of a straight peptide chain of 29 amino acids. It has been synthesized; the synthetic product has the full biological activity of natural glucagon. The structure of glucagon is free of cystine and isoleucine.

The pituitary gland has an anterior lobe, a posterior lobe, and an intermediate portion; they differ in cellular structure and in the structure and action of the hormones they form. The posterior lobe produces two similar hormones, oxytocin and vasopressin. The former causes contraction of the pregnant uterus; the latter raises the blood pressure. Both are octapeptides formed by a ring of five amino acids (the two cystine halves count as one amino acid) and a side chain of three amino acids. The two cystine halves are linked to each other by a disulfide bond, and the C terminal amino acid is glycinamide. The structure has been established and confirmed. Human vasopressin differs from oxytocin in that isoleucine is replaced by phenylalanine and leucine by arginine.Proteins. Formula 10: Amino acid sequence of the two similar hormones of the posterior lobe of the pituitary gland. (A) Oxytocin. (B) Human vasopressin. The solid line represents the disulfide bond between the two halves of cystine.

The intermediate part of the pituitary gland produces the melanocyte-stimulating hormone (MSH), which causes expansion of the pigmented melanophores (cells) in the skin of frogs and other batrachians. Two hormones, called α-MSH and β-MSH, have been prepared from hog pituitary glands. The first, α-MSH, consists of 13 amino acids; its N terminal serine is acetylated (i.e., the acetyl group, CH3CO, of acetic acid is attached), and its C terminal valine residue is present as valinamide. The second, β-MSH, contains in its 18 amino acids many of those occurring in α-MSH.Proteins. Formula 11: The amino acid sequence of hormones produced by the intermediate part of the pituitary gland. The amino acid sequence M.E.H.F.R.W.G. occurs in all melanocyte-stimulating hormones and in adrenocorticotropic homones (corticotropins).

The anterior pituitary lobe produces several protein hormones—a thyroid-stimulating hormone (thyrotropin), molecular weight 28,000; a lactogenic hormone, molecular weight 22,500; a growth hormone, molecular weight 21,500; a luteinizing hormone, molecular weight 30,000; and a follicle-stimulating hormone, molecular weight 29,000. The thyroid-stimulating hormone consists of α and β subunits with a composition similar to the subunits of luteinizing hormone. When separated, neither of the two subunits has hormonal activity; when combined, however, they regain about 50 percent of the original activity. The lactogenic hormone (prolactin) from sheep pituitary glands contains 190 amino acids. Their sequence has been elucidated; a similar peptide chain of 188 amino acids that has been synthesized not only has 10 percent of the biolog