Protein

Article by:

Ewalt, Karla L. The Scripps Research Institute, La Jolla, California.

Schimmel, Paul Skaggs Institute for Chemical Biology, Scripps Research Institute, La Jolla, California.

Perlmann, Gertrude E. Formerly, Rockefeller University, New York, New York.

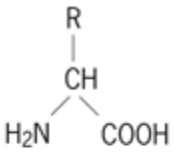
Manning, James M. Department of Biochemistry, Rockefeller University, New York, New York.

Publication year: 2014

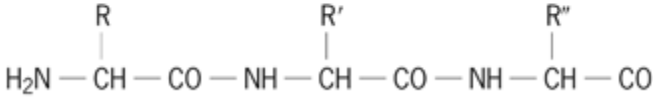
DOI: http://dx.doi.org/10.1036/1097-8542.550200 (http://dx.doi.org/10.1036/1097-8542.550200)

Content

A polymeric compound made up of various monomeric units called amino acids. Amino acids are joined together in a chain by peptide (amide) bonds between the α-carboxyl groups and the α-amino groups of adjacent amino acids. The first amino acid in a protein usually contains a free α-amino group, as shown below.



Amino acid



Peptide

Proteins generally contain from 50 to 1000 amino acids per chain. Small chains of up to 50 amino acids are usually referred to as peptides or polypeptides. *See also:* **Amino acids (/content/amino-acids/028100)**; **Peptide (/content/peptide/497210)**

Occurrence

Proteins are central to the processes of life. They are fundamental components of all biological systems, performing a wide variety of structural and functional roles. For example, proteins are primary constituents of structures such as hair, tendons, muscle, skin, and cartilage. Several hormones, such as insulin and growth hormone, are proteins. The substances responsible for oxygen and electron transport (hemoglobin and cytochromes, respectively) are conjugated proteins that contain a metalloporphyrin as the prosthetic group. Chromosomes are highly complex nucleoproteins, that is, proteins conjugated with nucleic acids. Viruses are also nucleoprotein in nature.

Protein enzymes are the catalysts of nearly all biochemical transformations. Pepsin and rennin are examples of digestive enzymes involved in breaking down food. DNA polymerases are enzymes that duplicate DNA for cell division, other enzymes are needed to repair damaged DNA, and gene expression is carried out by RNA polymerases. The chemical reactions used

in metabolic pathways for carbohydrates (citric acid cycle), lipids, amino acids, and energy production (oxidative phosphorylation) are catalyzed by protein enzymes. All living things contain proteins because they serve as the molecular tools and machines of life. *See also:* **Enzyme (/content/enzyme/236000)**

Of the more than 200 amino acids that have been discovered either in the free state or in small peptides, only 20 amino acids, or derivatives of these 20, are present in mammalian proteins. They are alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and valine. Each has a unique side chain (represented by R, R′, R″ in the peptide structure above) that defines the chemical properties of the amino acid. Except for glycine, amino acids have an inherent asymmetry referred to as chirality. In all natural proteins, they are of the L-configuration. Amino acids with the mirror-image configuration are D-amino acids. Both D- and L-amino acids occur in the free state and in small peptides. The amino acids can be grouped according to similarities in size, shape, charge, aromaticity, and polarity. Amino acids with nonpolar, hydrophobic side chains are alanine, isoleucine, leucine, methionine, tryptophan, phenyl-alanine, proline, and valine. Another group with polar, uncharged side chains is asparagine, cysteine, glutamine, serine, threonine, and tyrosine. Finally, five amino acids have side chains that can be charged at physiological pH values. Arginine, lysine, and histidine can be protonated to bear a positive charge, whereas aspartic acid and glutamic acid can be deprotonated to bear a negative charge. Commonly occurring derivatives of these 20 amino acids are cystine, hydroxyproline, and hydroxylysine. Cystine results from the oxidation of two cysteines to form a disulfide bridge. Hydroxyproline and hydroxy-lysine are formed by the enzymatic hydroxylation of specific proline and lysine, respectively. Modification of these two amino acids takes place while the polypeptide chain of collagen is being synthesized on the ribosome. Other examples of biological modification of parent amino acids are *O*-phosphoserine in phosvitin, ^{3}*N*-methylhistidine in actin, and ε-*N*-methyllysine in histones.

Biosynthesis

The linear arrangement of amino acids in a protein is termed its sequence (primary structure). The sequence is highly specific and characteristic for each particular protein. It is determined by the DNA sequence of each protein's gene that is expressed in the form of messenger RNA. Elucidation of the mechanism by which proteins are built up from free amino acids has been one of the key problems of molecular biology. *See also:* **Deoxyribonucleic acid (DNA) (/content/deoxyribonucleic-acid-dna/186500)**; **Molecular biology (/content/molecular-biology/430300)**; **Ribonucleic acid (RNA) (/content/ribonucleic-acid-rna/589000)**

Although a few proteins such as collagen are stable indefinitely in adulthood, most body proteins are in a continual process of degradation and synthesis (turnover). For example, the half-life of serum proteins in humans is about 10 days. Each amino acid is activated by adenosine triphosphate (ATP) by a specific enzyme called an aminoacyl-tRNA synthetase. These enzymes establish the relationship between specific nucleotides and amino acids, which is the basis for the genetic code. This activated amino acid is covalently attached to a transfer RNA that contains a triplet of nucleotides (anticodon) at one end. Each triplet of nucleotides is unique for one amino acid. Protein synthesis, the joining of amino acids to form polypeptides, takes place on the ribosome, a large ribonucleoprotein complex that translates mRNA into proteins. On the ribosome, aminoacyl-tRNA binds specifically to a segment on mRNA through bonding between the anticodon on the tRNA and a complementary codon of three nucleotides on the mRNA. Two activated aminoacyl-tRNA complexes bind adjacent mRNA codons and then react together to form a peptide bond. The process continues until the synthesis of the entire protein is complete. *See also:* **Ribosomes (/content/ribosomes/589200)**

Structure

The task of determining the sequences of proteins was a major preoccupation of many scientists in the twentieth century. Frederick Sanger was awarded a Nobel Prize in 1958 for determining the first protein sequence—insulin, a 51-amino acid

protein. Sanger's work proved for the first time that a protein was a pure chemical entity comprising a single amino acid sequence and not a collection of closely related amino acid sequences (sometimes referred to as statistical proteins). Later developments in DNA sequencing led to a Nobel Prize for Walter Gilbert and Frederick Sanger (his second). The relative ease of DNA sequencing combined with many additional technical innovations has permitted the whole-scale sequencing of genomes. Thus, most protein sequences are now deduced from the corresponding gene sequence. With this wealth of sequence information, researchers in the fields of protein chemistry and molecular and cellular biology are focused on understanding the functional roles and properties of proteins.

Proteins of similar function from different species have a common genetic origin and thus have related sequences. When differences exist in the sequence, changes are usually conservative, such as the replacement of one small hydrophobic amino acid by a different small hydrophobic amino acid. The degree of similarity between one protein in two different animals has been used to evaluate the relative genetic distance of the two animals. This type of comparison was made for the β-globin chain of hemoglobin, the blood protein that carries oxygen and carbon dioxide, for chimpanzees, monkeys, orangutans, gorillas, and humans. The relatively close relationship for all these animals is reflected by their nearly identical amino acid sequences for β-globin. However, some amino acid differences exist between β-globin chains. For example, chimpanzee β-globin is identical to gorilla β-globin at 146 of the 147 amino acids. The single difference is an arginine in one sequence and a lysine in the other. In this example, arginine and lysine are both positively charged amino acids and β-globin functions with either. Thus, this change is conservative. *See also:* **Hemoglobin (/content/hemoglobin/313800)**; **Proteins, evolution of (/content/proteins-evolution-of/550400)**

Nonconservative changes can alter the ability of a protein to function, as is evident by the single amino acid difference between normal and sickle human β-globin. Sickle β-globin contains valine at a position normally occupied by glutamic acid. This single amino acid change alters the overall structure of hemoglobin which, in turn, alters the shape of the red blood cell. When a person inherits the gene for sickle β-globin from both parents, it results in the disease sickle cell anemia. *See also:* **Sickle cell disease (/content/sickle-cell-disease/621650)**

Proteins are not stretched-out polymers; rather, each adopts a specific extended or compact and organized structure called its native structure. It is still not completely understood how proteins “fold” into their structures, nor can we accurately predict the complete structure from its amino acid sequence. Pioneering studies that won Christian Anfinsen the Nobel Prize in 1972 showed that proteins fold themselves—that all the information needed to reach the correct three-dimensional structure is contained in the linear sequence of amino acids. Thus, after synthesis on the ribosome, a polypeptide chain folds to its native structure. In order to prevent the misfolding of certain proteins, chaperone proteins intervene to facilitate proper folding. Proteins that become misfolded or that have lost their native structure pose a major problem in cells. Diseases such as cystic fibrosis and Alzheimer's disease are characterized by misfolded proteins. In cystic fibrosis, a mutation in the chloride ion channel protein prevents the final stages of protein folding and leads to deficiencies in ion transport across cell membranes. In Alzheimer's, a protein known as amyloid precursor protein takes on a misfolded shape that causes the protein to clump together, ultimately forming plaques in the brain. *See also:* **Molecular chaperone (/content/molecular-chaperone/801470)**

The polypeptide backbone of protein can fold in several ways by means of hydrogen bonds between the carbonyl oxygen and the amide nitrogen (**Fig. 1**, inset). Structural elements created by backbone hydrogen bonding interactions in the polypeptide are called secondary structures and include features like α-helices, β-sheets, and turns. In helices, the backbone is coiled in a regular fashion that brings peptide bonds separated by several amino acids into close spatial approximation. The stability of a helix is attributed to hydrogen bonds between these peptide bonds. The configuration of an α-helix, which is the most abundant secondary structural element in globular proteins, is described by a backbone amide hydrogen of one peptide unit bonded to a carbonyl oxygen (four amino acids away) that is part of another peptide unit (Fig. 1). Helical structures are often made up of peptide segments of around 10 amino acids, which corresponds to about three turns of helix. In keratin, the

fibrous protein found in hair and nails, much longer helices are formed and pack with helices from other keratin polypeptides to give tensile strength. *See also:* **Hydrogen bond (/content/hydrogen-bond/328800)**

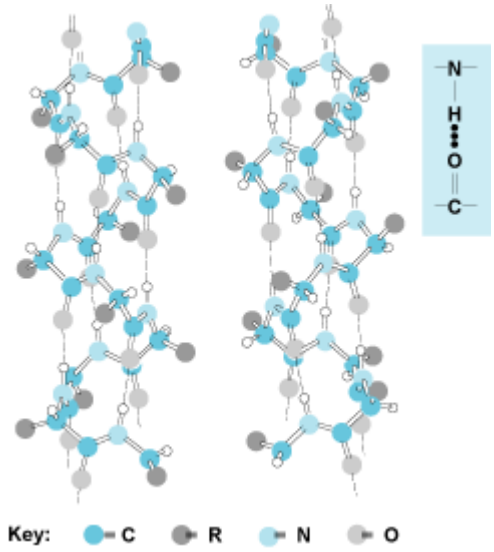


Fig. 1 The α-helix proposed by L. Pauling and R. B. Corey. The repeating —NH—CO—C units form the backbone which spirals up in a left-handed or a right-handed fashion. (The right-handed configuration is adopted in natural proteins with L-amino acids.) Hydrogen bonds are indicated by the broken lines. Note that the side chains (R) are all directed out from the helix. The pitch of the helix (axial displacement per complete turn of the helix) is about 0.54 nm, and there are 3.7 amino acids contained in one complete turn. The structure at the right shows the hydrogen bond. (*After J. T. Edsall and J. Wyman, Biophysical Chemistry, vol. 1, Academic Press, 1958*)

In addition to α-helices, polypeptides form β-sheet structures that are made of two or more segments that run parallel or antiparallel to each other and connect through backbone hydrogen bonds. The amino acid side chains stick out on alternating sides of the β-sheet surface. The human porin protein is made almost entirely of β-strands arranged to form a central pore that allows sugars, amino acids, ions, and other small metabolites in and out of cells.

The third level of folding in a protein (tertiary structure) comes through interactions between different parts of the molecule. At this level of structure, various secondary structure elements are brought together and interact through many types of associations. Hydrogen bonds between different amino acids and peptide bonds, hydrophobic interactions between nonpolar side chains of amino acids such as phenylalanine and leucine, and salt bridges such as those between positively charged lysyl side chains and negatively charged aspartyl side chains all contribute to the tertiary structure specific to a given protein. Disulfide bridges formed between two cysteines at different linear locations in the molecule can stabilize parts of a three-dimensional structure by introducing a covalent bond as a cross-link. The result is a unique architecture that is predetermined by the particular sequence of amino acids in the protein.

Finally, some proteins contain more than one polypeptide chain per molecule. This feature is referred to as the quaternary structure. There is usually a high degree of interaction between each subunit, for example, between the α- and β-globin chains of hemoglobin. The tetrameric structure of hemoglobin exemplifies how four individual polypeptide chains (two α- and two β-chains) function cooperatively. The binding of oxygen to one subunit causes a conformational change in the protein that facilitates the binding of oxygen to the other subunits. Thus, the cooperative binding of oxygen to hemoglobin is coordinated to facilitate oxygen binding in the lungs and its release in the tissues where oxygen pressure is slightly lower.

Among proteins with a known three-dimensional structure, there are a few recurring structures that predominate. These structural superfolds include closely related proteins, which have a high level of identical amino acids and perform similar functions, as well as unrelated proteins, which have independent amino acid sequences and function despite overall similarity

in architecture. Nine superfolds dominate the known protein structures. These are the globin, trefoil, up-down, immunoglobulin, αβ sandwich, jelly roll, doubly wound, UB αβ roll, and TIM barrel folds (**Fig. 2**). Hemoglobin and myoglobin are related and are representative of members of the globin fold family, which is characterized by eight α-helices arranged in a Greek key motif (the topological signature of many β-barrels and a majority of β-sandwich structures). Antibodies contain several repeating segments that have eight-stranded β-sheet structures belonging to the immunoglobulin fold family. Several enzymes that carry out the chemical reactions of metabolism are members of the TIM barrel fold family. This fold is made of eight β-strand/α-helix structures wrapped into a barrel. The namesake enzyme TIM (triose-phosphate isomerase) is used in the breakdown of sugars. Another member of the TIM barrel fold family is ribulose 1,5-bisphosphate carboxylase oxygenase (Rubisco), a plant protein that fixes carbon dioxide from the air into organic carbon needed to sustain life and is the most abundant protein on Earth. While many proteins are found to adopt one of the superfolds, many others have a unique architecture.

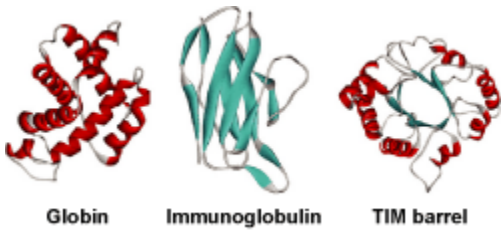


Fig. 2 Three commonly occurring protein folds.

Preparation

Many of the structural and functional studies of proteins are performed with purified preparations of proteins. Historically, proteins were obtained from their tissue of origin, but modern DNA technology allows proteins to be produced in high yield in bacteria, yeast, or isolated cells from insects and mammals. The isolation of a protein from any of these sources is accomplished by a variety of techniques. The primary concern is that the protein be isolated in its native form—in other words, not denatured or degraded during preparation.

All purification methods seek to exploit differences in solubility, charge, size, and resin-binding specificity in order to enrich the solution for the desired protein. The first step in preparation is usually to mechanically disrupt the tissue or cells containing the protein. Next, any of several methods are applied, often sequentially, to purify the protein. A few of the commonly applied methods are described.

Ion-exchange chromatography

Ion-exchange resins and modified cellulose derivatives are powerful tools for protein purification. A solid resin is chemically modified to introduce charged groups for anion (negative charge) or cation (positive charge) binding. Proteins, which typically contain many positive and negative charges on their surfaces, will interact with a specific resin with different binding affinities. The strength of the protein–resin interaction can be modulated by altering the pH of the solution and thus the protonation state of the protein, as well as the ionic strength of the buffer. Buffers, mixed to produce a gradient of pH and ionic strength, are used to differentially bind and elute proteins from the resin. Ion-exchange chromatography has the capacity to handle gram quantities of proteins. Low-molecular-weight proteins, especially those with markedly acidic (positively charged) or basic (negatively charged) properties, such as ribonuclease, histones, and lysozyme, can be separated from most other proteins in a mixture in a single experiment on an ion-exchange column. *See also:* **Chromatography (/content/chromatography/134300)**

Affinity chromatography

The resolving power of ion-exchange chromatography can be remarkably enhanced by the covalent attachment of a small molecule or of a protein that has an affinity for the target. For example, streptavidin is a protein from *Streptomyces* bacteria that tightly binds biotin. It is readily purified by binding to biotin-conjugated resins, which has thus led to its use as a handle for purification of other proteins. Therefore, through protein engineering, streptavidin is covalently linked to a second protein of interest. The two linked proteins can be purified by biotin affinity chromatography. Immunoglobulins, or antibodies, are tightly bound by protein A, made by *Staphylococcus* bacteria. Purification of immunoglobulins by affinity chromatography on protein A–coupled resin is an important method of isolating commercially produced immunoglobulins for medical therapies in diseases ranging from arthritis to cancer. In some cases, total purification in one step can be achieved with affinity chromatography. *See also:* **Protein engineering (/content/protein-engineering/801480)**

Preparative ultracentrifugation

Proteins of high molecular weight can be concentrated by applying a high gravitational field to the protein, contained in a tube, for a prolonged time. An ultracentrifuge produces a high gravitational field in which the protein sediments through the field, concentrating protein at the bottom of the tube. However, it is possible to achieve significant purification only for the heaviest or the lightest components in a mixture, and the degree of purification will depend on the range of sedimentation velocities represented. Solutions with densities less than that of the solvent allow proteins to float to the surface under the influence of a strong gravitational field. This principle has been successfully applied to the study of serum lipoproteins which, by virtue of their lipid content, have solution densities less than those of other, non-lipid-containing serum proteins. Sufficient salt is added to the serum to raise the solution densities of non-lipid-containing proteins above those of the lipoproteins. Centrifugation to equilibrium brings the lipoproteins to the surface, where they can be collected. This technique has been refined by the use of repeated centrifugation steps at progressively raised salt densities, allowing subfractions of lipoproteins of different solution density classes to be obtained on a preparative scale. *See also:* **Ultracentrifuge (/content/ultracentrifuge/719200)**

Gel filtration

Gel filtration is a technique for the purification of proteins that separates mainly on the basis of molecular weight. Dextran beads, just micrometers in diameter, are used to exclude proteins of high molecular weight but to permit smaller proteins to diffuse in and out. With a column of such beads, a continuous redistribution and separation occurs as the protein mixture moves through the column. Beads of different porosities are available to separate proteins of different molecular weights. For example, with one type of dextran bead, proteins with molecular weights of about 10,000 can be easily separated from those whose molecular weights are greater than 50,000. *See also:* **Filtration (/content/filtration/257300)**

Analysis

Methods for protein analysis include amino acid analysis, protein sequencing, mass spectrometry, isoelectric focusing, gel electrophoresis, and specific activity.

Amino acid analysis

Because of the relatively high molecular weights of proteins, the classical methods of organic chemistry are not adequate to establish composition or structure. A more meaningful empirical formula for a protein is given in terms of the total number of amino acids and the relative proportion of each amino acid. Thus, a protein can be hydrolyzed in hydrochloric acid to liberate the individual amino acids, which can be quantitatively estimated by amino acid analysis. These data, together with an independent measurement of molecular weight (that can be converted into the total number of amino acids), yield the amino acid composition, that is, the number and kind of each amino acid. In those few cases in which a protein is known to lack a particular amino acid (for example, tryptophan is absent from ribonuclease, and isoleucine is absent from human adult

hemoglobin), the amino acid composition can be used to estimate the purity of a sample.

Protein sequencing

Developed by Pehr Edman in the 1950s, sequential degradation (one amino acid at a time from one end of the chain) of proteins and peptides was used for decades to determine the sequence. The procedure was done by coupling the free N-terminal α-amino acid with a chemical (particularly phenylisothiocyanate) that removes the amino acid (under appropriate conditions) and leaves the remainder of the polypeptide chain intact. It is then possible to identify the cleaved derivative of the first amino acid and repeat the process on each truncated peptide that is successively generated.

Protein sequencing is rarely used anymore to determine the complete sequence of amino acids in a protein, because deduced protein sequences are available from the DNA sequences of many organisms. However, it remains a frequently used method to determine protein identity. Identity can be uniquely established from the sequence of the first six to ten amino acids and comparison of that sequence to those predicted from the sequences of the genes in the organism from which the protein was isolated.

Not all proteins generate useful sequence information from the amino terminus, based on reaction with phenylisothiocyanate. Sometimes the amino terminus of some proteins is blocked, or covalently modified, so the Edman method cannot work. To circumvent this problem, proteins can be randomly broken down into smaller polypeptides or peptides. The sequences of the individual peptides are then determined. However, the sequences of these pieces have to be combined in the correct order, and this can be challenging. Alternatively, stepwise removal of amino acids from the carboxyl end by enzymes called carboxypeptidases is used to determine the amino acid sequence at the opposite end of the polypeptide. *See also:* **Protein degradation (/content/protein-degradation/801460)**

Mass spectrometry

Powerful new applications of mass spectrometry have catalyzed its use in biology. Mass spectrometry has become an important tool for protein sequencing and identification, and for localization of modifications that occur after the protein has been synthesized. Mass-spectrometric analysis yields the molecular weight of the protein. The masses of the peptides in a proteolytic digest of a protein can also be determined. Then by comparison to protein sequence databases, the protein can be identified.

Isoelectric focusing

The net charge of a protein is largely a function of the relative number of basic (lysine, histidine, and arginine) and dicarboxylic amino acids (aspartic and glutamic acids). This net charge strongly influences the solubility of a protein at different pH values because the solubility depends in part on the proportion of polar groups. When the hydrogen-ion concentration is high (low pH), the net charge is positive; when the hydrogen-ion concentration is low (high pH), the net charge is negative. The pH at which the net charge of the protein is zero is defined as the isoelectric point (pI). Isoelectric focusing—the focusing by electrophoresis of individual proteins according to their respective pI's—is usually carried out in a vertical column that contains a sucrose concentration gradient or in a semisolid phase of dextran. Synthetic polymers of amino acids (polyamino acids) of varying isoelectric points are added to the protein mixture that is to be resolved. In an electric field, each synthetic polymer migrates to the point where it has zero net charge. Thus, a stable gradient of these polyamino acids is established that, in effect, is a pH gradient. The added proteins then are separated according to the intrinsic pI (isoelectric point) of each one. *See also:* **Isoelectric point (/content/isoelectric-point/355000)**; **pH (/content/ph/504000)**

Gel electrophoresis

Protein samples are analyzed for purity and size by various methods of polyacryl-amide gel electrophoresis (PAGE). Under native conditions, that is, in the absence of denaturing agents, proteins migrate according to their charge density and size and shape. At slightly basic pH, most proteins carry a net negative charge. The lower the charge density (less charge per molecular mass), the slower a protein migrates. The frictional force of the gel matrix creates a sieving effect, retarding the movement of proteins according to size.

Under denaturing conditions, mixtures of proteins are separated by PAGE based on size. Sodium dodecyl sulfate (SDS) is added to denature the protein. (Regardless of the shape of the native, undenatured protein, all proteins assume a rod shape when denatured. As a consequence, the shape differences between proteins are eliminated.) SDS is negatively charged and has an inherent tendency to stick to the protein backbone in an amount proportional to a protein's molecular weight. Thus, each SDS–protein complex migrates through the gel according to molecular weight. The size of a protein is calculated from the distance it migrates in the gel, when compared to proteins of known molecular mass.

Two-dimensional electrophoresis combines the resolving power of isoelectric focusing and denaturing PAGE to separate complex mixtures of proteins. To separate proteins that migrate at a single pI, SDS–PAGE is run at a right angle to the direction of the first (isoelectric focusing) electrophoresis. By these methods, thousands of proteins can be separated into a two-dimensional array of “spots” composed of single pure proteins. This technique is widely used in the field of research known as proteomics.

Specific activity

Function is the ultimate outcome of structure. Therefore, the function or activity of a protein should be measured whenever possible. The activity of a protein refers to the specific function of the native protein. In the case of an enzyme, the activity is the catalysis of a chemical transformation of substrate into product. Enzyme activities are usually reported in activity units such as micromoles of product formed per minute. The specific activity of an enzyme is the total activity divided by the amount of protein. During the purification of an enzyme, specific activity increases because contaminating proteins are removed.

For example, luciferase is an enzyme found in the light-producing organ of fireflies, which create light to signal other fireflies. The enzyme catalyzes a chemical reaction that converts luciferin, adenosine triphosphate, and oxygen (O_2) into oxyluciferin, adenosine monophosphate, pyrophosphate, and light. Thus, one way to measure the activity of firefly luciferase is to measure the amount of light produced.

A pure, completely native enzyme will have the highest specific activity. As the protein becomes degraded or denatured, the specific activity will be reduced. Mutations that affect (negatively or positively) the functioning of an enzyme can be assessed by measuring the specific activity and comparing it to that of the wild-type enzyme. Those mutations that disrupt activity are important because their locations in the structure can identify the residues needed for catalysis.

Karla L. Ewalt
Paul Schimmel
Gertrude E. Perlmann
James M. Manning

Bibliography

C. Branden and J. Tooze, *Introduction to Protein Structure*, Garland Publishing, New York, 1999

G. A. Petsko and D. Ringe, *Protein Structure and Function*, New Science Press, London, 2004

B. E. Tropp, *Molecular Biology: Genes to Proteins*, Jones and Bartlett, Sudbury, MA, 2008

D. Voet, *Biochemistry*, 4th ed., John Wiley & Sons, Hoboken, NJ, 2011

D. Whitford, *Proteins: Structure and Function*, John Wiley & Sons, Hoboken, NJ, 2005

Additional Readings

M. K. Campbell and S. O. Farrell, *Biochemistry*, 7th ed., Brooks Cole, Belmont, CA, 2012

S. Darmanis et al., Sensitive plasma protein analysis by microparticle-based proximity ligation assays, *Mol. Cell. Proteomics*, 9(2):327–335, 2010 DOI: 10.1074/mcp.M900248-MCP200 (http://dx.doi.org/10.1074/mcp.M900248-MCP200)

M. M. Gromiha, *Protein Bioinformatics: From Sequence to Function*, Academic Press, New Delhi, India, 2010

A. B. Hughes (ed.), *Amino Acids, Peptides and Proteins in Organic Chemistry: Analysis and Function of Amino Acids and Peptides*, vol. 5, Wiley-VCH Verlag, Weinheim, Germany, 2012

C. Timm and C. M. Niemeyer, On-chip protein biosynthesis, *Angew. Chem. Int. Ed.*, 52(10):2652–2654, 2013 DOI: 10.1002/anie.201208880 (http://dx.doi.org/10.1002/anie.201208880)