

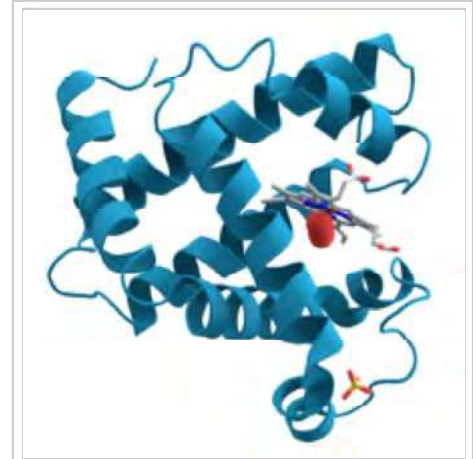


# Protein

From Wikipedia, the free encyclopedia

**Proteins** (/ˈproʊ.tiːnz/ or /ˈproʊti.nz/) are large biomolecules, or macromolecules, consisting of one or more long chains of amino acid residues. Proteins perform a vast array of functions within organisms, including catalysing metabolic reactions, DNA replication, responding to stimuli, and transporting molecules from one location to another. Proteins differ from one another primarily in their sequence of amino acids, which is dictated by the nucleotide sequence of their genes, and which usually results in protein folding into a specific three-dimensional structure that determines its activity.

A linear chain of amino acid residues is called a polypeptide. A protein contains at least one long polypeptide. Short polypeptides, containing less than 20–30 residues, are rarely considered to be proteins and are commonly called peptides, or sometimes oligopeptides. The individual amino acid residues are bonded together by peptide bonds and adjacent amino acid residues. The sequence of amino acid residues in a protein is defined by the sequence of a gene, which is encoded in the genetic code. In general, the genetic code specifies 20 standard amino acids; however, in certain organisms the genetic code can include selenocysteine and—in certain archaea—pyrrolysine. Shortly after or even during synthesis, the residues in a protein are often chemically modified by



A representation of the 3D structure of the protein myoglobin showing turquoise  $\alpha$ -helices. This protein was the first to have its structure solved by X-ray crystallography. Towards the right-center among the coils, a prosthetic group called a heme group (shown in gray) with a bound oxygen molecule (red).

post-translational modification, which alters the physical and chemical properties, folding, stability, activity, and ultimately, the function of the proteins. Sometimes proteins have non-peptide groups attached, which can be called prosthetic groups or cofactors. Proteins can also work together to achieve a particular function, and they often associate to form stable protein complexes.

Once formed, proteins only exist for a certain period of time and are then degraded and recycled by the cell's machinery through the process of protein turnover. A protein's lifespan is measured in terms of its half-life and covers a wide range. They can exist for minutes or years with an average lifespan of 1–2 days in mammalian cells. Abnormal and or misfolded proteins are degraded more rapidly either due to being targeted for destruction or due to being unstable.

Like other biological macromolecules such as polysaccharides and nucleic acids, proteins are essential parts of organisms and participate in virtually every process within cells. Many proteins are enzymes that catalyse biochemical reactions and are vital to metabolism. Proteins also have structural or mechanical functions, such as actin and myosin in muscle and the proteins in the cytoskeleton, which form a system of scaffolding that maintains cell shape. Other proteins are important in cell signaling, immune responses, cell adhesion, and the cell cycle. In animals, proteins are needed in the diet to provide the essential amino acids that cannot be synthesized. Digestion breaks the proteins down for use in the metabolism.

Proteins may be purified from other cellular components using a variety of techniques such as ultracentrifugation, precipitation, electrophoresis, and chromatography; the advent of genetic engineering has made possible a number of methods to facilitate purification. Methods commonly used to study protein structure and function include immunohistochemistry, site-directed mutagenesis, X-ray crystallography, nuclear magnetic resonance and mass spectrometry.



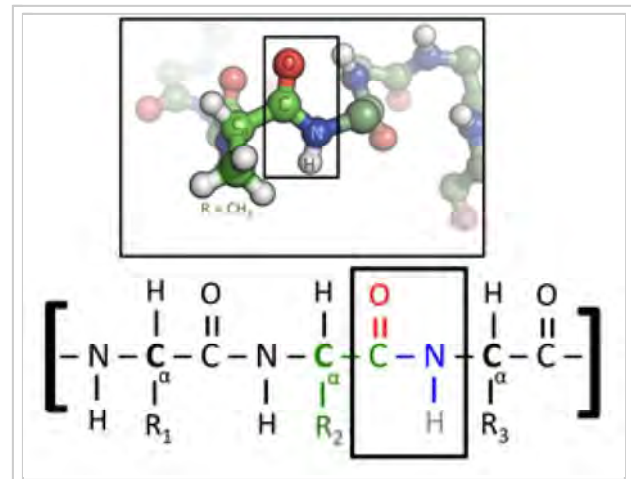
## Contents

- 1 Biochemistry
  - 1.1 Abundance in cells
- 2 Synthesis
  - 2.1 Biosynthesis
  - 2.2 Chemical synthesis
- 3 Structure
  - 3.1 Structure determination
- 4 Cellular functions
  - 4.1 Enzymes
  - 4.2 Cell signaling and ligand binding
  - 4.3 Structural proteins
- 5 Methods of study
  - 5.1 Protein purification
  - 5.2 Cellular localization
  - 5.3 Proteomics
  - 5.4 Bioinformatics
    - 5.4.1 Structure prediction and simulation
    - 5.4.2 Protein disorder and unstructure prediction
- 6 Nutrition
- 7 History and etymology
- 8 See also
- 9 References
- 10 Textbooks
- 11 External links
  - 11.1 Databases and projects
  - 11.2 Tutorials and educational websites

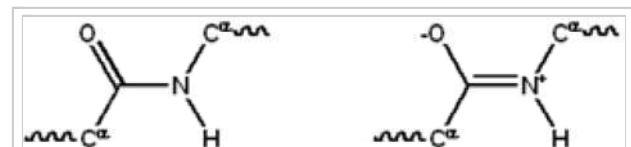
## Biochemistry

Most proteins consist of linear polymers built from series of up to 20 different L- $\alpha$ -amino acids. All proteinogenic amino acids possess common structural features, including an  $\alpha$ -carbon to which an amino group, a carboxyl group, and a variable side chain are bonded. Only proline differs from this basic structure as it contains an unusual ring to the N-end amine group, which forces the CO–NH amide moiety into a fixed conformation.<sup>[1]</sup> The side chains of the standard amino acids, detailed in the list of standard amino acids, have a great variety of chemical structures and properties; it is the combined effect of all of the amino acid side chains in a protein that ultimately determines its three-dimensional structure and its chemical reactivity.<sup>[2]</sup> The amino acids in a polypeptide chain are linked by peptide bonds. Once linked in the protein chain, an individual amino acid is called a *residue*, and the linked series of carbon, nitrogen, and oxygen atoms are known as the *main chain* or *protein backbone*.<sup>[3]</sup>

The peptide bond has two resonance forms that contribute some double-bond character and inhibit rotation around its axis, so that the alpha carbons are roughly coplanar. The other two dihedral angles in the peptide bond determine the local shape assumed by the protein backbone.<sup>[4]</sup> The end with a free amino group is known as the N-terminus or amino terminus, whereas the end of the



Chemical structure of the peptide bond (bottom) and the three-dimensional structure of a peptide bond between an alanine and an adjacent amino acid (top/inset)



Resonance structures of the peptide bond that links individual amino acids to form a protein polymer

protein with a free carboxyl group is known as the C-terminus or carboxy terminus (the sequence of the protein is written from N-terminus to C-terminus, from left to right).

The words *protein*, *polypeptide*, and *peptide* are a little ambiguous and can overlap in meaning. *Protein* is generally used to refer to the complete biological molecule in a stable conformation, whereas *peptide* is generally reserved for a short amino acid oligomers often lacking a stable three-dimensional structure. However, the boundary between the two is not well defined and usually lies near 20–30 residues.<sup>[5]</sup> *Polypeptide* can refer to any single linear chain of amino acids, usually regardless of length, but often implies an absence of a defined conformation.

## Abundance in cells

It has been estimated that average-sized bacteria contain about 2 million proteins per cell (e.g. *E. coli* and *Staphylococcus aureus*). Smaller bacteria, such as *Mycoplasma* or *spirochetes* contain fewer molecules, namely on the order of 50,000 to 1 million. By contrast, eukaryotic cells are larger and thus contain much more protein. For instance, yeast cells were estimated to contain about 50 million proteins and human cells on the order of 1 to 3 billion.<sup>[6]</sup> The concentration of individual protein copies ranges from a few molecules per cell up to 20 million.<sup>[7]</sup> Not all genes coding proteins are expressed in most cells and their number depends on for example cell type and external stimuli. For instance, of the 20,000 or so proteins encoded by the human genome, only 6,000 are detected in lymphoblastoid cells.<sup>[8]</sup> Moreover, the number of proteins the genome encodes correlates well with the organism complexity. Eukaryotes, bacteria, Archaea and viruses have on average 15145, 3200, 2358 and 42 proteins respectively coded in their genomes.<sup>[9]</sup>

# Synthesis

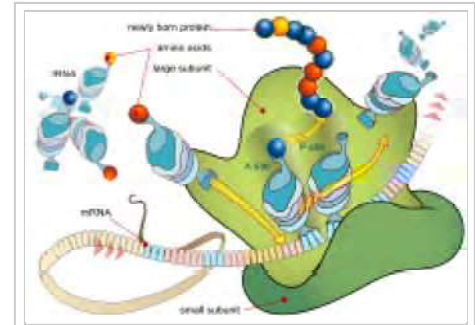
## Biosynthesis

Proteins are assembled from amino acids using information encoded in genes. Each protein has its own unique amino acid sequence that is specified by the nucleotide sequence of the gene encoding this protein. The genetic code is a set of three-nucleotide sets called codons and each three-nucleotide combination designates an amino acid, for example AUG

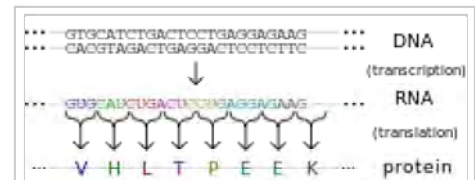
(adenine-uracil-guanine) is the code for methionine. Because DNA contains four nucleotides, the total number of possible codons is 64; hence, there is some redundancy in the genetic code, with some amino acids

specified by more than one codon.<sup>[10]</sup> Genes encoded in DNA are first transcribed into pre-messenger RNA (mRNA) by proteins such as RNA polymerase. Most organisms then process the pre-mRNA (also known as a *primary transcript*) using various forms of Post-transcriptional modification to form the mature mRNA, which is then used as

a template for protein synthesis by the ribosome. In prokaryotes the mRNA may either be used as soon as it is produced, or be bound by a ribosome after having moved away from the nucleoid. In contrast, eukaryotes make mRNA in the cell nucleus and then translocate it across the nuclear membrane into the cytoplasm, where protein synthesis then takes place. The rate of protein synthesis is higher in prokaryotes than eukaryotes and can reach up to 20 amino acids per second.<sup>[11]</sup>



A ribosome produces a protein using mRNA as template



The DNA sequence of a gene encodes the amino acid sequence of a protein

The process of synthesizing a protein from an mRNA template is known as translation. The mRNA is loaded onto the ribosome and is read three nucleotides at a time by matching each codon to its base pairing anticodon located on a transfer RNA molecule, which carries the amino acid corresponding to the codon it recognizes. The enzyme aminoacyl tRNA synthetase "charges" the tRNA molecules with the correct amino acids. The growing polypeptide is often termed the *nascent chain*. Proteins are always biosynthesized from N-terminus to C-terminus.<sup>[10]</sup>

The size of a synthesized protein can be measured by the number of amino acids it contains and by its total molecular mass, which is normally reported in units of *daltons* (synonymous with atomic mass units), or the derivative unit kilodalton (kDa). The average size of protein increases from Archaea, Bacteria to Eukaryote (283, 311, 438 residues and 31, 34, 49 kDa respectively) due bigger number of protein domains constituting proteins in higher organisms.<sup>[9]</sup> For instance, yeast proteins are on average 466 amino acids long and 53 kDa in mass.<sup>[5]</sup> The largest known proteins are the titins, a component of the muscle sarcomere, with a molecular mass of almost 3,000 kDa and a total length of almost 27,000 amino acids.<sup>[12]</sup>

## Chemical synthesis

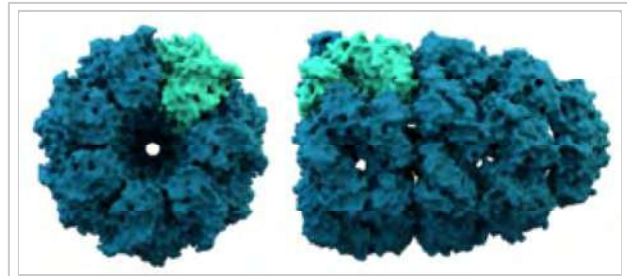
Short proteins can also be synthesized chemically by a family of methods known as peptide synthesis, which rely on organic synthesis techniques such as chemical ligation to produce peptides in high yield.<sup>[13]</sup> Chemical synthesis allows for the introduction of non-natural amino acids into polypeptide chains, such as attachment of fluorescent probes to amino acid side chains.<sup>[14]</sup> These methods are useful in laboratory biochemistry and cell biology, though generally not for commercial applications. Chemical synthesis is inefficient for polypeptides longer than about 300 amino acids, and the synthesized proteins may not readily assume their native tertiary structure. Most chemical synthesis methods proceed from C-terminus to N-terminus, opposite the biological reaction.<sup>[15]</sup>



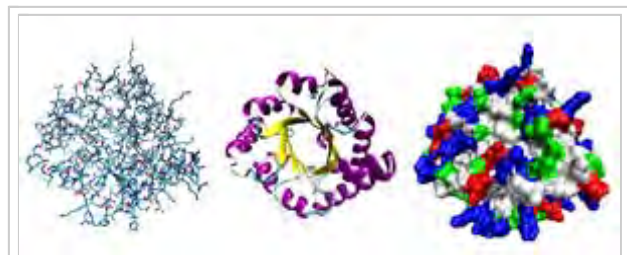
## Structure

Most proteins fold into unique 3-dimensional structures. The shape into which a protein naturally folds is known as its native conformation.<sup>[16]</sup> Although many proteins can fold unassisted, simply through the chemical properties of their amino acids, others require the aid of molecular chaperones to fold into their native states.<sup>[17]</sup> Biochemists often refer to four distinct aspects of a protein's structure:<sup>[18]</sup>

- *Primary structure*: the amino acid sequence. A protein is a polyamide.
- *Secondary structure*: regularly repeating local structures stabilized by hydrogen bonds. The most common examples are the  $\alpha$ -helix,  $\beta$ -sheet and turns. Because secondary structures are local, many regions of different secondary structure can be present in the same protein molecule.
- *Tertiary structure*: the overall shape of a single protein molecule; the spatial relationship of the secondary structures to one another. Tertiary structure is generally stabilized by nonlocal interactions, most commonly the formation of a hydrophobic core, but also through salt bridges, hydrogen



The crystal structure of the chaperonin, a huge protein complex. A single protein subunit is highlighted. Chaperonins assist protein folding.



Three possible representations of the three-dimensional structure of the protein triose phosphate isomerase. **Left**: All-atom representation colored by atom type. **Middle**: Simplified representation illustrating the backbone conformation, colored by secondary structure. **Right**: Solvent-accessible surface representation colored by residue type (acidic residues red, basic residues blue, polar residues green, nonpolar residues white).



bonds, disulfide bonds, and even posttranslational modifications. The term "tertiary structure" is often used as synonymous with the term *fold*. The tertiary structure is what controls the basic function of the protein.

- *Quaternary structure*: the structure formed by several protein molecules (polypeptide chains), usually called *protein subunits* in this context, which function as a single protein complex.

Proteins are not entirely rigid molecules. In addition to these levels of structure, proteins may shift between several related structures while they perform their functions. In the context of these functional rearrangements, these tertiary or quaternary structures are usually referred to as "conformations", and transitions between them are called *conformational changes*. Such changes are often induced by the binding of a substrate molecule to an enzyme's active site, or the physical region of the protein that participates in chemical catalysis. In solution proteins also undergo variation in structure through thermal vibration and the collision with other molecules.

[19]

Proteins can be informally divided into three main classes, which correlate with typical tertiary structures: globular proteins, fibrous proteins, and membrane proteins. Almost all globular proteins are soluble and many are enzymes. Fibrous proteins are often structural, such as collagen, the major component of connective tissue, or keratin, the protein component of hair and nails. Membrane proteins often serve as receptors or provide channels for polar or charged molecules to pass through the cell membrane.<sup>[20]</sup>



Molecular surface of several proteins showing their comparative sizes. From left to right are: immunoglobulin G (IgG, an antibody), hemoglobin, insulin (a hormone), adenylate kinase (an enzyme), and glutamine synthetase (an enzyme).

A special case of intramolecular hydrogen bonds within proteins, poorly shielded from water attack and hence promoting their own dehydration, are called dehydrons.<sup>[21]</sup>

## Structure determination

Discovering the tertiary structure of a protein, or the quaternary structure of its complexes, can provide important clues about how the protein performs its function. Common experimental methods of structure determination include X-ray crystallography and NMR spectroscopy, both of which can produce information at atomic resolution. However, NMR experiments are able to provide information from which a subset of distances between pairs of atoms can be estimated, and the final possible conformations for a protein are determined by solving a distance geometry problem. Dual polarisation interferometry is a quantitative analytical method for measuring the overall protein conformation and conformational changes due to interactions or other stimulus. Circular dichroism is another laboratory technique for determining internal  $\beta$ -sheet /  $\alpha$ -helical composition of proteins. Cryoelectron microscopy is used to produce lower-resolution structural information about very large protein complexes, including assembled viruses;<sup>[22]</sup> a variant known as electron crystallography can also produce high-resolution information in some cases, especially for two-dimensional crystals of membrane proteins.<sup>[23]</sup> Solved structures are usually deposited in the Protein Data Bank (PDB), a freely available resource from which structural data about thousands of proteins can be obtained in the form of Cartesian coordinates for each atom in the protein.<sup>[24]</sup>

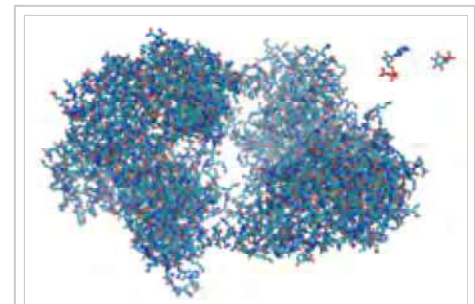
Many more gene sequences are known than protein structures. Further, the set of solved structures is biased toward proteins that can be easily subjected to the conditions required in X-ray crystallography, one of the major structure determination methods. In particular, globular proteins are comparatively easy to crystallize in preparation for X-ray crystallography. Membrane proteins, by contrast, are difficult to crystallize and are underrepresented in the PDB.<sup>[25]</sup> Structural genomics initiatives have attempted to remedy these deficiencies by

systematically solving representative structures of major fold classes. Protein structure prediction methods attempt to provide a means of generating a plausible structure for proteins whose structures have not been experimentally determined.<sup>[26]</sup>

## Cellular functions

Proteins are the chief actors within the cell, said to be carrying out the duties specified by the information encoded in genes.<sup>[5]</sup> With the exception of certain types of RNA, most other biological molecules are relatively inert elements upon which proteins act. Proteins make up half the dry weight of an *Escherichia coli* cell, whereas other macromolecules such as DNA and RNA make up only 3% and 20%, respectively.<sup>[27]</sup> The set of proteins expressed in a particular cell or cell type is known as its proteome.

The chief characteristic of proteins that also allows their diverse set of functions is their ability to bind other molecules specifically and tightly. The region of the protein responsible for binding another molecule is known as the binding site and is often a depression or "pocket" on the molecular surface. This binding ability is mediated by the tertiary structure of the protein, which defines the binding site pocket, and by the chemical properties of the surrounding amino acids' side chains. Protein binding can be extraordinarily tight and specific; for example, the ribonuclease inhibitor protein binds to human angiogenin with a sub-femtomolar dissociation constant ( $<10^{-15}$  M) but does not bind at all to its amphibian homolog onconase ( $>1$  M). Extremely minor chemical changes such as the addition of a single methyl group to a binding partner can sometimes suffice to nearly eliminate binding; for



The enzyme hexokinase is shown as a conventional ball-and-stick molecular model. To scale in the top right-hand corner are two of its substrates, ATP and glucose.

example, the aminoacyl tRNA synthetase specific to the amino acid valine discriminates against the very similar side chain of the amino acid isoleucine.  
[28]

Proteins can bind to other proteins as well as to small-molecule substrates. When proteins bind specifically to other copies of the same molecule, they can oligomerize to form fibrils; this process occurs often in structural proteins that consist of globular monomers that self-associate to form rigid fibers. Protein–protein interactions also regulate enzymatic activity, control progression through the cell cycle, and allow the assembly of large protein complexes that carry out many closely related reactions with a common biological function. Proteins can also bind to, or even be integrated into, cell membranes. The ability of binding partners to induce conformational changes in proteins allows the construction of enormously complex signaling networks.<sup>[29]</sup> As interactions between proteins are reversible, and depend heavily on the availability of different groups of partner proteins to form aggregates that are capable to carry out discrete sets of function, study of the interactions between specific proteins is a key to understand important aspects of cellular function, and ultimately the properties that distinguish particular cell types.<sup>[30][31]</sup>

## Enzymes

The best-known role of proteins in the cell is as enzymes, which catalyse chemical reactions. Enzymes are usually highly specific and accelerate only one or a few chemical reactions. Enzymes carry out most of the reactions involved in metabolism, as well as manipulating DNA in processes such as DNA replication, DNA repair, and transcription. Some enzymes act on other proteins to add or remove chemical groups in a process known as posttranslational modification. About 4,000 reactions are known to be catalysed by enzymes.<sup>[32]</sup> The rate acceleration conferred by enzymatic catalysis is often enormous—as much as  $10^{17}$ -fold increase in rate over the uncatalysed reaction in the case of orotate decarboxylase (78 million years without the enzyme, 18 milliseconds with the enzyme).<sup>[33]</sup>

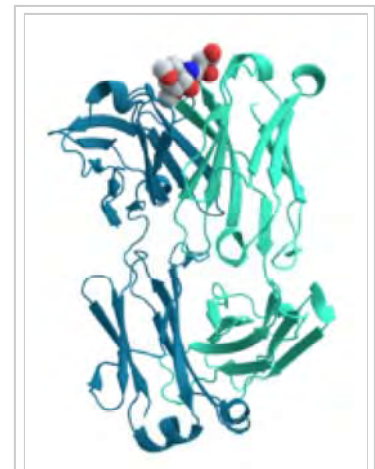
The molecules bound and acted upon by enzymes are called substrates. Although enzymes can consist of hundreds of amino acids, it is usually only a small fraction of the residues that come in contact with the substrate, and an even smaller fraction—three to four residues on average—that are directly involved in catalysis.<sup>[34]</sup> The region of the enzyme that binds the substrate and contains the catalytic residues is known as the active site.

Dirigent proteins are members of a class of proteins that dictate the stereochemistry of a compound synthesized by other enzymes.<sup>[35]</sup>

## Cell signaling and ligand binding

Many proteins are involved in the process of cell signaling and signal transduction. Some proteins, such as insulin, are extracellular proteins that transmit a signal from the cell in which they were synthesized to other cells in distant tissues. Others are membrane proteins that act as receptors whose main function is to bind a signaling molecule and induce a biochemical response in the cell. Many receptors have a binding site exposed on the cell surface and an effector domain within the cell, which may have enzymatic activity or may undergo a conformational change detected by other proteins within the cell.<sup>[36]</sup>

Antibodies are protein components of an adaptive immune system whose main function is to bind antigens, or foreign substances in the body, and target them for destruction. Antibodies can be secreted into the extracellular environment or anchored in the membranes of specialized B cells known as plasma cells. Whereas enzymes are limited in their binding affinity for their substrates by the necessity of conducting their reaction, antibodies have no such constraints. An antibody's binding affinity to its target is extraordinarily high.<sup>[37]</sup>



Ribbon diagram of a mouse antibody against cholera that binds a carbohydrate antigen

Many ligand transport proteins bind particular small biomolecules and transport them to other locations in the body of a multicellular organism. These proteins must have a high binding affinity when their ligand is present in high concentrations, but must also release the ligand when it is present at low concentrations in the target tissues. The canonical example of a ligand-binding protein is haemoglobin, which transports oxygen from the lungs to other organs and tissues in all vertebrates and has close homologs in every biological kingdom.<sup>[38]</sup> Lectins are sugar-binding proteins which are highly specific for their sugar moieties. Lectins typically play a role in biological recognition phenomena involving cells and proteins.<sup>[39]</sup> Receptors and hormones are highly specific binding proteins.

Transmembrane proteins can also serve as ligand transport proteins that alter the permeability of the cell membrane to small molecules and ions. The membrane alone has a hydrophobic core through which polar or charged molecules cannot diffuse. Membrane proteins contain internal channels that allow such molecules to enter and exit the cell. Many ion channel proteins are specialized to select for only a particular ion; for example, potassium and sodium channels often discriminate for only one of the two ions.<sup>[40]</sup>

## Structural proteins

Structural proteins confer stiffness and rigidity to otherwise-fluid biological components. Most structural proteins are fibrous proteins; for example, collagen and elastin are critical components of connective tissue such as cartilage, and keratin is found in hard or filamentous structures such as hair, nails, feathers, hooves, and some animal shells.<sup>[41]</sup> Some globular proteins can also play structural functions, for example, actin and tubulin are globular and soluble as monomers, but polymerize to form long, stiff fibers that make up the cytoskeleton, which allows the cell to maintain its shape and size.

Other proteins that serve structural functions are motor proteins such as myosin, kinesin, and dynein, which are capable of generating mechanical forces. These proteins are crucial for cellular motility of single celled

organisms and the sperm of many multicellular organisms which reproduce sexually. They also generate the forces exerted by contracting muscles<sup>[42]</sup> and play essential roles in intracellular transport.

## Methods of study

The activities and structures of proteins may be examined *in vitro*, *in vivo*, and *in silico*. ***In vitro*** studies of purified proteins in controlled environments are useful for learning how a protein carries out its function: for example, enzyme kinetics studies explore the chemical mechanism of an enzyme's catalytic activity and its relative affinity for various possible substrate molecules. By contrast, ***in vivo*** experiments can provide information about the physiological role of a protein in the context of a cell or even a whole organism. ***In silico*** studies use computational methods to study proteins.

## Protein purification

To perform *in vitro* analysis, a protein must be purified away from other cellular components. This process usually begins with cell lysis, in which a cell's membrane is disrupted and its internal contents released into a solution known as a crude lysate. The resulting mixture can be purified using ultracentrifugation, which fractionates the various cellular components into fractions containing soluble proteins; membrane lipids and proteins; cellular organelles, and nucleic acids. Precipitation by a method known as salting out can concentrate the proteins from this lysate. Various types of chromatography are then used to isolate the protein or proteins of interest based on properties such as molecular weight, net charge and binding affinity.

<sup>[43]</sup> The level of purification can be monitored using various types of gel electrophoresis if the desired protein's molecular weight and isoelectric point are known, by spectroscopy if the protein has distinguishable spectroscopic features, or by enzyme assays if the protein has enzymatic activity.

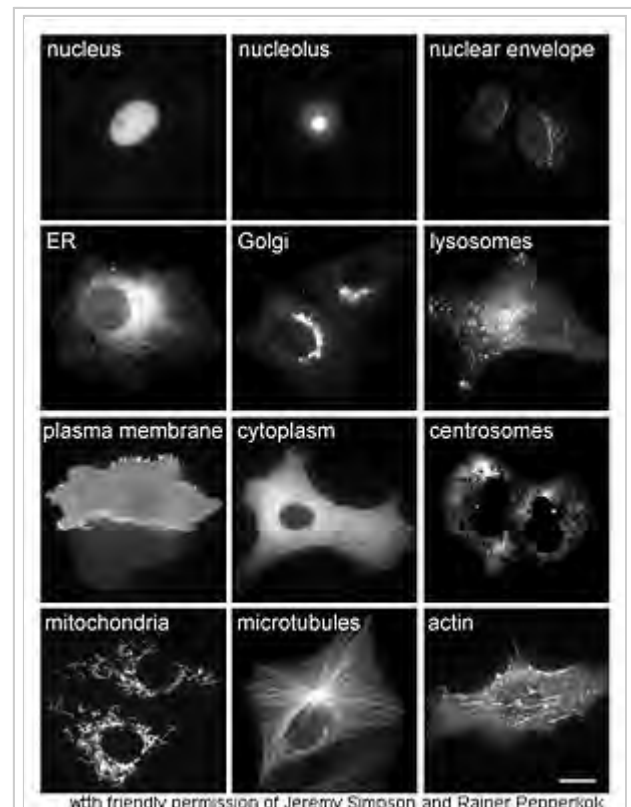
Additionally, proteins can be isolated according their charge using electrofocusing.<sup>[44]</sup>



For natural proteins, a series of purification steps may be necessary to obtain protein sufficiently pure for laboratory applications. To simplify this process, genetic engineering is often used to add chemical features to proteins that make them easier to purify without affecting their structure or activity. Here, a "tag" consisting of a specific amino acid sequence, often a series of histidine residues (a "His-tag"), is attached to one terminus of the protein. As a result, when the lysate is passed over a chromatography column containing nickel, the histidine residues ligate the nickel and attach to the column while the untagged components of the lysate pass unimpeded. A number of different tags have been developed to help researchers purify specific proteins from complex mixtures.<sup>[45]</sup>

## Cellular localization

The study of proteins *in vivo* is often concerned with the synthesis and localization of the protein within the cell. Although many intracellular proteins are synthesized in the cytoplasm and membrane-bound or secreted proteins in the endoplasmic reticulum, the specifics of how proteins are targeted to specific organelles or cellular structures is often unclear. A useful technique for assessing cellular localization uses genetic engineering to express in a cell a fusion protein or chimera consisting of the natural protein of interest linked to a "reporter" such as green fluorescent protein (GFP).<sup>[46]</sup> The fused protein's position within the



Proteins in different cellular compartments and structures tagged with green fluorescent protein (here, white)

cell can be cleanly and efficiently visualized using microscopy,<sup>[47]</sup> as shown in the figure opposite.

Other methods for elucidating the cellular location of proteins requires the use of known compartmental markers for regions such as the ER, the Golgi, lysosomes or vacuoles, mitochondria, chloroplasts, plasma membrane, etc. With the use of fluorescently tagged versions of these markers or of antibodies to known markers, it becomes much simpler to identify the localization of a protein of interest. For example, indirect immunofluorescence will allow for fluorescence colocalization and demonstration of location. Fluorescent dyes are used to label cellular compartments for a similar purpose.<sup>[48]</sup>

Other possibilities exist, as well. For example, immunohistochemistry usually utilizes an antibody to one or more proteins of interest that are conjugated to enzymes yielding either luminescent or chromogenic signals that can be compared between samples, allowing for localization information. Another applicable technique is cofractionation in sucrose (or other material) gradients using isopycnic centrifugation.<sup>[49]</sup> While this technique does not prove colocalization of a compartment of known density and the protein of interest, it does increase the likelihood, and is more amenable to large-scale studies.

Finally, the gold-standard method of cellular localization is immunoelectron microscopy. This technique also uses an antibody to the protein of interest, along with classical electron microscopy techniques. The sample is prepared for normal electron microscopic examination, and then treated with an antibody to the protein of interest that is conjugated to an extremely electro-dense material, usually gold. This allows for the localization of both ultrastructural details as well as the protein of interest.<sup>[50]</sup>

Through another genetic engineering application known as site-directed mutagenesis, researchers can alter the protein sequence and hence its structure, cellular localization, and susceptibility to regulation. This technique

even allows the incorporation of unnatural amino acids into proteins, using modified tRNAs,<sup>[51]</sup> and may allow the rational design of new proteins with novel properties.<sup>[52]</sup>

## Proteomics

The total complement of proteins present at a time in a cell or cell type is known as its proteome, and the study of such large-scale data sets defines the field of proteomics, named by analogy to the related field of genomics. Key experimental techniques in proteomics include 2D electrophoresis,<sup>[53]</sup> which allows the separation of a large number of proteins, mass spectrometry,<sup>[54]</sup> which allows rapid high-throughput identification of proteins and sequencing of peptides (most often after in-gel digestion), protein microarrays,<sup>[55]</sup> which allow the detection of the relative levels of a large number of proteins present in a cell, and two-hybrid screening, which allows the systematic exploration of protein–protein interactions.<sup>[56]</sup> The total complement of biologically possible such interactions is known as the interactome.<sup>[57]</sup> A systematic attempt to determine the structures of proteins representing every possible fold is known as structural genomics.<sup>[58]</sup>

## Bioinformatics

A vast array of computational methods have been developed to analyze the structure, function, and evolution of proteins.

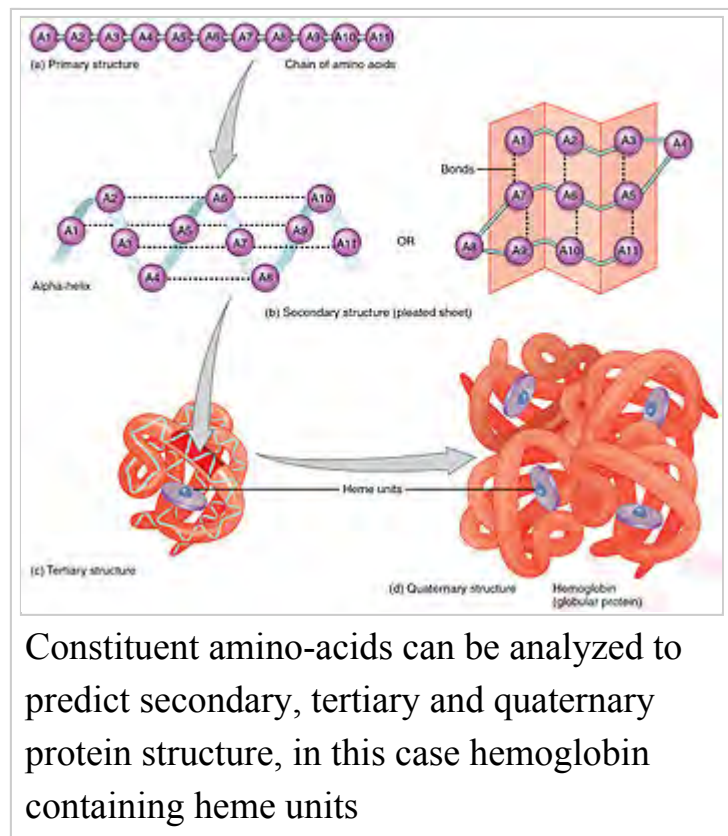
The development of such tools has been driven by the large amount of genomic and proteomic data available for a variety of organisms, including the human genome. It is simply impossible to study all proteins experimentally, hence only a few are subjected to laboratory experiments while computational tools are used to extrapolate to similar proteins. Such homologous proteins can be efficiently identified in distantly related organisms by sequence alignment. Genome and gene sequences can be searched by a variety of tools for certain properties. Sequence profiling tools

can find restriction enzyme sites, open reading frames in nucleotide sequences, and predict secondary structures. Phylogenetic trees can be constructed and evolutionary hypotheses developed using special software like ClustalW regarding the ancestry of modern organisms and the genes they express. The field of bioinformatics is now indispensable for the analysis of genes and proteins.

## Structure prediction and simulation

Complementary to the field of structural genomics, protein structure prediction seeks to develop efficient ways to provide plausible models for proteins whose structures have not yet been determined experimentally.<sup>[59]</sup> The most successful type of structure prediction, known as homology modeling, relies on the existence of a "template" structure with sequence similarity to the protein being modeled; structural genomics' goal is to provide sufficient representation in solved structures to model most of those that remain.<sup>[60]</sup>

Although producing accurate models remains a challenge when only distantly related template structures are available, it has been suggested that sequence alignment is the bottleneck in this process, as quite accurate models can be produced if a "perfect" sequence alignment is known.<sup>[61]</sup> Many structure prediction methods have served to inform the emerging field of protein engineering, in which novel protein folds have already been designed.<sup>[62]</sup> A



Constituent amino-acids can be analyzed to predict secondary, tertiary and quaternary protein structure, in this case hemoglobin containing heme units

more complex computational problem is the prediction of intermolecular interactions, such as in molecular docking and protein–protein interaction prediction.<sup>[63]</sup>

The processes of protein folding and binding can be simulated using such technique as molecular mechanics, in particular, molecular dynamics and Monte Carlo, which increasingly take advantage of parallel and distributed computing (Folding@home project,<sup>[64]</sup> molecular modeling on GPU). The folding of small  $\alpha$ -helical protein domains such as the villin headpiece<sup>[65]</sup> and the HIV accessory protein<sup>[66]</sup> have been successfully simulated *in silico*, and hybrid methods that combine standard molecular dynamics with quantum mechanics calculations have allowed exploration of the electronic states of rhodopsins.<sup>[67]</sup>

### **Protein disorder and unstructure prediction**

Many proteins (in Eucaryota ~33%) contain large unstructured but biologically functional segments and can be classified as intrinsically disordered proteins.<sup>[68]</sup> Predicting and analysing protein disorder is, therefore, an important part of protein structure characterisation.<sup>[69]</sup>

## **Nutrition**

Most microorganisms and plants can biosynthesize all 20 standard amino acids, while animals (including humans) must obtain some of the amino acids from the diet.<sup>[27]</sup> The amino acids that an organism cannot synthesize on its own are referred to as essential amino acids. Key enzymes that synthesize certain amino acids are not present in animals — such as aspartokinase, which catalyses the first step in the synthesis of lysine, methionine, and threonine from aspartate. If amino acids are present in the environment, microorganisms can conserve energy by taking up the amino acids from their surroundings and downregulating their biosynthetic pathways.

In animals, amino acids are obtained through the consumption of foods containing protein. Ingested proteins are then broken down into amino acids through digestion, which typically involves denaturation of the protein through exposure to acid and hydrolysis by enzymes called proteases. Some ingested amino acids are used for protein biosynthesis, while others are converted to glucose through gluconeogenesis, or fed into the citric acid cycle. This use of protein as a fuel is particularly important under starvation conditions as it allows the body's own proteins to be used to support life, particularly those found in muscle.<sup>[70]</sup> Amino acids are also an important dietary source of nitrogen.

## History and etymology

Proteins were recognized as a distinct class of biological molecules in the eighteenth century by Antoine Fourcroy and others, distinguished by the molecules' ability to coagulate or flocculate under treatments with heat or acid.<sup>[71]</sup> Noted examples at the time included albumin from egg whites, blood serum albumin, fibrin, and wheat gluten.

Proteins were first described by the Dutch chemist Gerardus Johannes Mulder and named by the Swedish chemist Jöns Jacob Berzelius in 1838.<sup>[72][73]</sup> Mulder carried out elemental analysis of common proteins and found that nearly all proteins had the same empirical formula,  $C_{400}H_{620}N_{100}O_{120}P_1S_1$ .<sup>[74]</sup> He came to the erroneous conclusion that they might be composed of a single type of (very large) molecule. The term "protein" to describe these molecules was proposed by Mulder's associate Berzelius; protein is derived from the Greek word *πρωτεῖος* (*proteios*), meaning "primary",<sup>[75]</sup> "in the lead", or "standing in front",<sup>[76]</sup> + *-in*. Mulder went on to identify the products of protein degradation such as the amino acid leucine for which he found a (nearly correct) molecular weight of 131 Da.<sup>[74]</sup>

Early nutritional scientists such as the German Carl von Voit believed that protein was the most important nutrient for maintaining the structure of the body, because it was generally believed that "flesh makes flesh."<sup>[77]</sup> Karl Heinrich Ritthausen extended known protein forms with the identification of glutamic acid. At the Connecticut Agricultural Experiment Station a detailed review of the vegetable proteins was compiled by Thomas Burr Osborne. Working with Lafayette Mendel and applying Liebig's law of the minimum in feeding laboratory rats, the nutritionally essential amino acids were established. The work was continued and communicated by William Cumming Rose. The understanding of proteins as polypeptides came through the work of Franz Hofmeister and Hermann Emil Fischer. The central role of proteins as enzymes in living organisms was not fully appreciated until 1926, when James B. Sumner showed that the enzyme urease was in fact a protein.<sup>[78]</sup>

The difficulty in purifying proteins in large quantities made them very difficult for early protein biochemists to study. Hence, early studies focused on proteins that could be purified in large quantities, e.g., those of blood, egg white, various toxins, and digestive/metabolic enzymes obtained from slaughterhouses. In the 1950s, the Armour Hot Dog Co. purified 1 kg of pure bovine pancreatic ribonuclease A and made it freely available to scientists; this gesture helped ribonuclease A become a major target for biochemical study for the following decades.<sup>[74]</sup>

Linus Pauling is credited with the successful prediction of regular protein secondary structures based on hydrogen bonding, an idea first put forth by William Astbury in 1933.<sup>[79]</sup> Later work by Walter Kauzmann on denaturation,<sup>[80][81]</sup> based partly on previous studies by Kaj Linderstrøm-Lang,<sup>[82]</sup> contributed an understanding of protein folding and structure mediated by hydrophobic interactions.



John Kendrew with model of myoglobin in progress



The first protein to be sequenced was insulin, by Frederick Sanger, in 1949. Sanger correctly determined the amino acid sequence of insulin, thus conclusively demonstrating that proteins consisted of linear polymers of amino acids rather than branched chains, colloids, or cyclols.<sup>[83]</sup> He won the Nobel Prize for this achievement in 1958.<sup>[84]</sup>




The first protein structures to be solved were hemoglobin and myoglobin, by Max Perutz and Sir John Cowdery Kendrew, respectively, in 1958.<sup>[85][86]</sup> As of 2016, the Protein Data Bank has over 115,000 atomic-resolution structures of proteins.<sup>[87]</sup> In more recent times, cryo-electron microscopy of large macromolecular assemblies<sup>[88]</sup> and computational protein structure prediction of small protein domains<sup>[89]</sup> are two methods approaching atomic resolution.

## See also

- DNA-binding protein
- Macromolecule
- Intein
- List of proteins
- Proteopathy
- Proteopedia
- Proteolysis
- Protein sequence space
- Protein superfamily




## References

1. Nelson DL, Cox MM (2005). *Lehninger's Principles of Biochemistry* (4th ed.). New York, New York: W. H. Freeman and Company.
2. Gutteridge A, Thornton JM (2005). "Understanding nature's catalytic toolkit". *Trends in Biochemical Sciences*. **30** (11): 622–29. doi:10.1016/j.tibs.2005.09.006. PMID 16214343.
3. Murray *et al.*, p. 19.
4. Murray *et al.*, p. 31.
5. Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipurksy SL, Darnell J (2004). *Molecular Cell Biology* (5th ed.). New York, New York: WH Freeman and Company.

6. Milo, Ron (2013-12-01). "What is the total number of protein molecules per cell volume? A call to rethink some published values". *BioEssays*. **35** (12): 1050–1055. doi:10.1002/bies.201300066. ISSN 1521-1878. PMC 3910158 . PMID 24114984.
7. Beck, M.; Schmidt, A.; Malmstroem, J.; Claassen, M.; Ori, A.; Szymborska, A.; Herzog, F.; Rinner, O.; Ellenberg, J. (2011). "The quantitative proteome of a human cell line.". *Mol Syst Biol*. **7**: 549. doi:10.1038/msb.2011.82. PMID 22068332.
8. Wu, Linfeng; Candille, Sophie I.; Choi, Yoonha; Xie, Dan; Jiang, Lihua; Li-Pook-Than, Jennifer; Tang, Hua; Snyder, Michael. "Variation and genetic control of protein abundance in humans". *Nature*. **499** (7456): 79–82. doi:10.1038/nature12223. PMC 3789121 . PMID 23676674.
9. Kozłowski, Lukasz P. (2016). "Proteome-pI: proteome isoelectric point database". *Nucleic Acids Research*: gkw978. doi:10.1093/nar/gkw978. ISSN 0305-1048.
10. van Holde and Mathews, pp. 1002–42.
11. Dobson CM (2000). "The nature and significance of protein folding". In Pain RH. *Mechanisms of Protein Folding*. Oxford, Oxfordshire: Oxford University Press. pp. 1–28. ISBN 0-19-963789-X.
12. Fulton A, Isaacs W (1991). "Titin, a huge, elastic sarcomeric protein with a probable role in morphogenesis". *BioEssays*. **13** (4): 157–61. doi:10.1002/bies.950130403. PMID 1859393.
13. Bruckdorfer T, Marder O, Albericio F (2004). "From production of peptides in milligram amounts for research to multi-tons quantities for drugs of the future". *Current Pharmaceutical Biotechnology*. **5** (1): 29–43. doi:10.2174/1389201043489620. PMID 14965208.
14. Schwarzer D, Cole P (2005). "Protein semisynthesis and expressed protein ligation: chasing a protein's tail". *Current Opinion in Chemical Biology*. **9** (6): 561–69. doi:10.1016/j.cbpa.2005.09.018. PMID 16226484.
15. Kent SB (2009). "Total chemical synthesis of proteins". *Chemical Society Reviews*. **38** (2): 338–51. doi:10.1039/b700141j. PMID 19169452.
16. Murray *et al.*, p. 36.
17. Murray *et al.*, p. 37.
18. Murray *et al.*, pp. 30–34.
19. van Holde and Mathews, pp. 368–75.
20. van Holde and Mathews, pp. 165–85.
21. Fernández A, Scott R (2003). "Dehydron: a structurally encoded signal for protein interaction". *Biophysical Journal*. **85** (3): 1914–28. Bibcode:2003BpJ....85.1914F. doi:10.1016/S0006-3495(03)74619-0. PMC 1303363 . PMID 12944304.
22. Branden and Tooze, pp. 340–41.




23. Gonen T, Cheng Y, Sliz P, Hiroaki Y, Fujiyoshi Y, Harrison SC, Walz T (2005). "Lipid-protein interactions in double-layered two-dimensional AQP0 crystals". *Nature*. **438** (7068): 633–38. Bibcode:2005Natur.438..633G. doi:10.1038/nature04321. PMC 1350984 . PMID 16319884.
24. Standley DM, Kinjo AR, Kinoshita K, Nakamura H (2008). "Protein structure databases with new web services for structural biology and biomedical research". *Briefings in Bioinformatics*. **9** (4): 276–85. doi:10.1093/bib/bbn015. PMID 18430752.
25. Walian P, Cross TA, Jap BK (2004). "Structural genomics of membrane proteins". *Genome Biology*. **5** (4): 215. doi:10.1186/gb-2004-5-4-215. PMC 395774 . PMID 15059248.
26. Sleator RD. (2012). "Prediction of protein functions". *Methods in Molecular Biology*. *Methods in Molecular Biology*. **815**: 15–24. doi:10.1007/978-1-61779-424-7\_2. ISBN 978-1-61779-423-0. PMID 22130980.
27. Voet D, Voet JG. (2004). *Biochemistry* Vol 1 3rd ed. Wiley: Hoboken, NJ.
28. Sankaranarayanan R, Moras D (2001). "The fidelity of the translation of the genetic code". *Acta Biochimica Polonica*. **48** (2): 323–35. PMID 11732604.
29. van Holde and Mathews, pp. 830–49.
30. Copland JA, Sheffield-Moore M, Koldzic-Zivanovic N, Gentry S, Lamprou G, Tzortzatou-Stathopoulou F, Zoumpourlis V, Urban RJ, Vlahopoulos SA (2009). "Sex steroid receptors in skeletal differentiation and epithelial neoplasia: is tissue-specific intervention possible?". *BioEssays*. **31** (6): 629–41. doi:10.1002/bies.200800138. PMID 19382224.
31. Samarin S, Nusrat A (2009). "Regulation of epithelial apical junctional complex by Rho family GTPases". *Frontiers in Bioscience*. **14** (14): 1129–42. doi:10.2741/3298. PMID 19273120.
32. Bairoch A (2000). "The ENZYME database in 2000" (PDF). *Nucleic Acids Research*. **28** (1): 304–305. doi:10.1093/nar/28.1.304. PMC 102465 . PMID 10592255. Archived from the original (PDF) on June 1, 2011.
33. Radzicka A, Wolfenden R (1995). "A proficient enzyme". *Science*. **267** (5194): 90–93. Bibcode:1995Sci...267...90R. doi:10.1126/science.7809611. PMID 7809611.
34. EBI External Services (2010-01-20). "The Catalytic Site Atlas at The European Bioinformatics Institute". Ebi.ac.uk. Retrieved 2011-01-16.
35. Pickel B, Schaller A (2013). "Dirigent proteins: molecular characteristics and potential biotechnological applications". *Applied Microbiology and Biotechnology*. **97** (19): 8427–8438. doi:10.1007/s00253-013-5167-4. PMID 23989917.
36. Branden and Tooze, pp. 251–81.
37. van Holde and Mathews, pp. 247–50.
38. van Holde and Mathews, pp. 220–29.

39. Rüdiger H, Siebert HC, Solís D, Jiménez-Barbero J, Romero A, von der Lieth CW, Diaz-Mariño T, Gabius HJ (2000). "Medicinal chemistry based on the sugar code: fundamentals of lectinology and experimental strategies with lectins as targets". *Current Medicinal Chemistry*. **7** (4): 389–416. doi:10.2174/0929867003375164. PMID 10702616.
40. Branden and Tooze, pp. 232–34.
41. van Holde and Mathews, pp. 178–81.
42. van Holde and Mathews, pp. 258–64; 272.
43. Murray *et al.*, pp. 21–24.
44. Hey J, Posch A, Cohen A, Liu N, Harbers A (2008). "Fractionation of complex protein mixtures by liquid-phase isoelectric focusing". *Methods in Molecular Biology*. *Methods in Molecular Biology*<sup>TM</sup>. **424**: 225–39. doi:10.1007/978-1-60327-064-9\_19. ISBN 978-1-58829-722-8. PMID 18369866.
45. Terpe K (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems". *Applied Microbiology and Biotechnology*. **60** (5): 523–33. doi:10.1007/s00253-002-1158-6. PMID 12536251.
46. Stepanenko OV, Verkhusha VV, Kuznetsova IM, Uversky VN, Turoverov KK (2008). "Fluorescent proteins as biomarkers and biosensors: throwing color lights on molecular and cellular processes". *Current Protein & Peptide Science*. **9** (4): 338–69. doi:10.2174/138920308785132668. PMC 2904242 . PMID 18691124.
47. Yuste R (2005). "Fluorescence microscopy today". *Nature Methods*. **2** (12): 902–904. doi:10.1038/nmeth1205-902. PMID 16299474.
48. Margolin W (2000). "Green fluorescent protein as a reporter for macromolecular localization in bacterial cells". *Methods (San Diego, Calif.)*. **20** (1): 62–72. doi:10.1006/meth.1999.0906. PMID 10610805.
49. Walker JH, Wilson K (2000). *Principles and Techniques of Practical Biochemistry*. Cambridge, UK: Cambridge University Press. pp. 287–89. ISBN 0-521-65873-X.
50. Mayhew TM, Lucocq JM (2008). "Developments in cell biology for quantitative immunoelectron microscopy based on thin sections: a review". *Histochemistry and Cell Biology*. **130** (2): 299–313. doi:10.1007/s00418-008-0451-6. PMC 2491712 . PMID 18553098.
51. Hohsaka T, Sisido M (2002). "Incorporation of non-natural amino acids into proteins". *Current Opinion in Chemical Biology*. **6** (6): 809–15. doi:10.1016/S1367-5931(02)00376-9. PMID 12470735.
52. Cedrone F, Ménez A, Quéméneur E (2000). "Tailoring new enzyme functions by rational redesign". *Current Opinion in Structural Biology*. **10** (4): 405–10. doi:10.1016/S0959-440X(00)00106-8. PMID 10981626.

53. Görg A, Weiss W, Dunn MJ (2004). "Current two-dimensional electrophoresis technology for proteomics". *Proteomics*. **4** (12): 3665–85. doi:10.1002/pmic.200401031. PMID 15543535.
54. Conrotto P, Souchelnytskyi S (2008). "Proteomic approaches in biological and medical sciences: principles and applications". *Experimental Oncology*. **30** (3): 171–80. PMID 18806738.
55. Joos T, Bachmann J (2009). "Protein microarrays: potentials and limitations". *Frontiers in Bioscience*. **14** (14): 4376–85. doi:10.2741/3534. PMID 19273356.
56. Koegl M, Uetz P (2007). "Improving yeast two-hybrid screening systems". *Briefings in Functional Genomics & Proteomics*. **6** (4): 302–12. doi:10.1093/bfgp/elm035. PMID 18218650.
57. Plewczyński D, Ginalski K (2009). "The interactome: predicting the protein–protein interactions in cells". *Cellular & Molecular Biology Letters*. **14** (1): 1–22. doi:10.2478/s11658-008-0024-7. PMID 18839074.
58. Zhang C, Kim SH (2003). "Overview of structural genomics: from structure to function". *Current Opinion in Chemical Biology*. **7** (1): 28–32. doi:10.1016/S1367-5931(02)00015-7. PMID 12547423.
59. Zhang Y (2008). "Progress and challenges in protein structure prediction". *Current Opinion in Structural Biology*. **18** (3): 342–48. doi:10.1016/j.sbi.2008.02.004. PMC 2680823 . PMID 18436442.
60. Xiang Z (2006). "Advances in homology protein structure modeling". *Current Protein and Peptide Science*. **7** (3): 217–27. doi:10.2174/138920306777452312. PMC 1839925 . PMID 16787261.
61. Zhang Y, Skolnick J (2005). "The protein structure prediction problem could be solved using the current PDB library". *Proceedings of the National Academy of Sciences of the United States of America*. **102** (4): 1029–34. Bibcode:2005PNAS..102.1029Z. doi:10.1073/pnas.0407152101. PMC 545829 . PMID 15653774.
62. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (2003). "Design of a novel globular protein fold with atomic-level accuracy". *Science*. **302** (5649): 1364–68. Bibcode:2003Sci...302.1364K. doi:10.1126/science.1089427. PMID 14631033.
63. Ritchie DW (2008). "Recent progress and future directions in protein–protein docking". *Current Protein and Peptide Science*. **9** (1): 1–15. doi:10.2174/138920308783565741. PMID 18336319.
64. Scheraga HA, Khalili M, Liwo A (2007). "Protein-folding dynamics: overview of molecular simulation techniques". *Annual Review of Physical Chemistry*. **58**: 57–83. Bibcode:2007ARPC...58...57S. doi:10.1146/annurev.physchem.58.032806.104614. PMID 17034338.
65. Zagrovic B, Snow CD, Shirts MR, Pande VS (2002). "Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed

- computing". *Journal of Molecular Biology*. **323** (5): 927–37. doi:10.1016/S0022-2836(02)00997-X. PMID 12417204.
66. Herges T, Wenzel W (2005). "In silico folding of a three helix protein and characterization of its free-energy landscape in an all-atom force field". *Physical Review Letters*. **94** (1): 018101. Bibcode:2005PhRvL..94a8101H. doi:10.1103/PhysRevLett.94.018101. PMID 15698135.
67. Hoffmann M, Wanko M, Strodel P, König PH, Frauenheim T, Schulten K, Thiel W, Tajkhorshid E, Elstner M (2006). "Color tuning in rhodopsins: the mechanism for the spectral shift between bacteriorhodopsin and sensory rhodopsin II". *Journal of the American Chemical Society*. **128** (33): 10808–18. doi:10.1021/ja062082i. PMID 16910676.
68. Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. (2004). "Prediction and functional analysis of native disorder in proteins from the three kingdoms of life". *Journal of Molecular Biology*. **337** (3): 635–45. doi:10.1016/j.jmb.2004.02.002. PMID 15019783.
69. Peter Tompa; Alan Fersht (18 November 2009). *Structure and Function of Intrinsically Disordered Proteins*. CRC Press. ISBN 978-1-4200-7893-0.
70. Brosnan J (June 2003). "Interorgan amino acid transport and its regulation". *Journal of Nutrition*. **133** (6 Suppl 1): 2068S–72S. PMID 12771367.
71. Thomas Burr Osborne (1909): *The Vegetable Proteins* (<https://archive.org/details/vegetableprotein00osbouoft>), History pp 1 to 6, from archive.org
72. Bulletin des Sciences Physiques et Naturelles en Néerlande (1838). pg 104. SUR LA COMPOSITION DE QUELQUES SUBSTANCES ANIMALES (<https://archive.org/details/bulletindesscien00leyd>)
73. Hartley Harold (1951). "Origin of the Word 'Protein.' ". *Nature*. **168** (4267): 244–244. doi:10.1038/168244a0.
74. Perrett D (2007). "From 'protein' to the beginnings of clinical proteomics". *Proteomics: Clinical Applications*. **1** (8): 720–38. doi:10.1002/prca.200700525. PMID 21136729.
75. *New Oxford Dictionary of English*
76. Reynolds JA, Tanford C (2003). *Nature's Robots: A History of Proteins (Oxford Paperbacks)*. New York, New York: Oxford University Press. p. 15. ISBN 0-19-860694-X.
77. Bischoff TLW, Voit, C (1860). *Die Gesetze der Ernaehrung des Pflanzenfressers durch neue Untersuchungen festgestellt* (in German). Leipzig, Heidelberg.
78. Sumner JB (1926). "The isolation and crystallization of the enzyme urease. Preliminary paper" (PDF). *Journal of Biological Chemistry*. **69** (2): 435–41.
79. Pauling L, Corey RB, Branson HR (1951). "The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain" (PDF). *Proceedings of the National Academy of Sciences of the United States of America*.



- 37** (5): 235–40. Bibcode:1951PNAS...37..235P. doi:10.1073/pnas.37.5.235. PMC 1063348 . PMID 14834145.
80. Kauzmann W (1956). "Structural factors in protein denaturation". *Journal of Cellular Physiology. Supplement*. **47** (Suppl 1): 113–31. doi:10.1002/jcp.1030470410. PMID 13332017.
81. Kauzmann W (1959). "Some factors in the interpretation of protein denaturation". *Advances in Protein Chemistry*. *Advances in Protein Chemistry*. **14**: 1–63. doi:10.1016/S0065-3233(08)60608-7. ISBN 978-0-12-034214-3. PMID 14404936.
82. Kalman SM, Linderstrom-Lang K, Ottesen M, Richards FM (1955). "Degradation of ribonuclease by subtilisin". *Biochimica et Biophysica Acta*. **16** (2): 297–99. doi:10.1016/0006-3002(55)90224-9. PMID 14363272.
83. Sanger F (1949). "The terminal peptides of insulin". *Biochemical Journal*. **45** (5): 563–74. PMC 1275055 . PMID 15396627.
84. Sanger F. (1958), *Nobel lecture: The chemistry of insulin* (PDF), Nobelprize.org
85. Muirhead H, Perutz M (1963). "Structure of hemoglobin. A three-dimensional fourier synthesis of reduced human hemoglobin at 5.5 Å resolution". *Nature*. **199** (4894): 633–38. Bibcode:1963Natur.199..633M. doi:10.1038/199633a0. PMID 14074546.
86. Kendrew J, Bodo G, Dintzis H, Parrish R, Wyckoff H, Phillips D (1958). "A three-dimensional model of the myoglobin molecule obtained by X-ray analysis". *Nature*. **181** (4610): 662–66. Bibcode:1958Natur.181..662K. doi:10.1038/181662a0. PMID 13517261.
87. "RCSB Protein Data Bank". Retrieved 2014-04-05.
88. Zhou ZH (2008). "Towards atomic resolution structural determination by single-particle cryo-electron microscopy". *Current Opinion in Structural Biology*. **18** (2): 218–28. doi:10.1016/j.sbi.2008.03.004. PMC 2714865 . PMID 18403197.
89. Keskin O, Tuncbag N, Gursoy A (2008). "Characterization and prediction of protein interfaces to infer protein-protein interaction networks". *Current Pharmaceutical Biotechnology*. **9** (2): 67–76. doi:10.2174/138920108783955191. PMID 18393863.

## Textbooks

- Branden C, Tooze J (1999). *Introduction to Protein Structure*. New York: Garland Pub. ISBN 0-8153-2305-0.
- Murray RF, Harper HW, Granner DK, Mayes PA, Rodwell VW (2006). *Harper's Illustrated Biochemistry*. New York: Lange Medical Books/McGraw-Hill. ISBN 0-07-146197-3.



- Van Holde KE, Mathews CK (1996). *Biochemistry*. Menlo Park, California: Benjamin/Cummings Pub. Co., Inc. ISBN 0-8053-3931-0.

## External links

### Databases and projects

- The Protein Naming Utility (<http://www.jcvi.org/pn-utility>)
- Human Protein Atlas (<http://www.proteinatlas.org/>)
- NCBI Entrez Protein database (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=protein>)
- NCBI Protein Structure database (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=structure>)
- Human Protein Reference Database (<http://www.hprd.org/>)
- Human Proteinpedia (<http://www.humanproteinpedia.org/>)
- Folding@Home (Stanford University) (<http://folding.stanford.edu/>)
- Comparative Toxicogenomics Database (<http://ctd.mdibl.org/>) curates protein–chemical interactions, as well as gene/protein–disease relationships and chemical-disease relationships.
- Bioinformatic Harvester (<http://harvester.fzk.de/>) A Meta search engine (29 databases) for gene and protein information.
- Protein Databank in Europe (<http://www.pdbe.org/>) (see also PDBeQuips (<http://www.pdbe.org/quips>), short articles and tutorials on interesting PDB structures)
- Research Collaboratory for Structural Bioinformatics (<http://www.rcsb.org/>) (see also Molecule of the Month ([http://www.rcsb.org/pdb/static.do?p=education\\_discussion/molecule\\_of\\_the\\_month/index.html](http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/index.html)), presenting short accounts on selected proteins from the PDB)



Wikimedia Commons has media related to ***Proteins***.



Look up ***protein*** in Wiktionary, the free dictionary.

- Proteopedia – Life in 3D (<http://www.proteopedia.org/>): rotatable, zoomable 3D model with wiki annotations for every known protein molecular structure.
- UniProt the Universal Protein Resource (<http://www.expasy.uniprot.org/>)
- neXtProt – Exploring the universe of human proteins (<http://www.nextprot.org/>): human-centric protein knowledge resource
- Multi-Omics Profiling Expression Database: MOPED (<https://www.proteinspire.org/MOPED/mopedviews/mopedAbout.jsf/>) human and model organism protein/gene knowledge and expression data

## Tutorials and educational websites

- "An Introduction to Proteins" ([http://web.stanford.edu/group/hopes/cgi-bin/hopes\\_test/an-introduction-to-proteins/](http://web.stanford.edu/group/hopes/cgi-bin/hopes_test/an-introduction-to-proteins/)) from HOPES (Huntington's Disease Outreach Project for Education at Stanford)
- Proteins: Biogenesis to Degradation – The Virtual Library of Biochemistry and Cell Biology (<http://www.biochemweb.org/proteins.shtml>)
- Alphabet of Protein Structures (<http://www.bioch.ox.ac.uk/howarth/alphabet.htm>)

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Protein&oldid=754983427>"

Categories: [Molecular biology](#) | [Proteins](#) | [Proteomics](#)

- 
- This page was last modified on 15 December 2016, at 16:07.
  - Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.