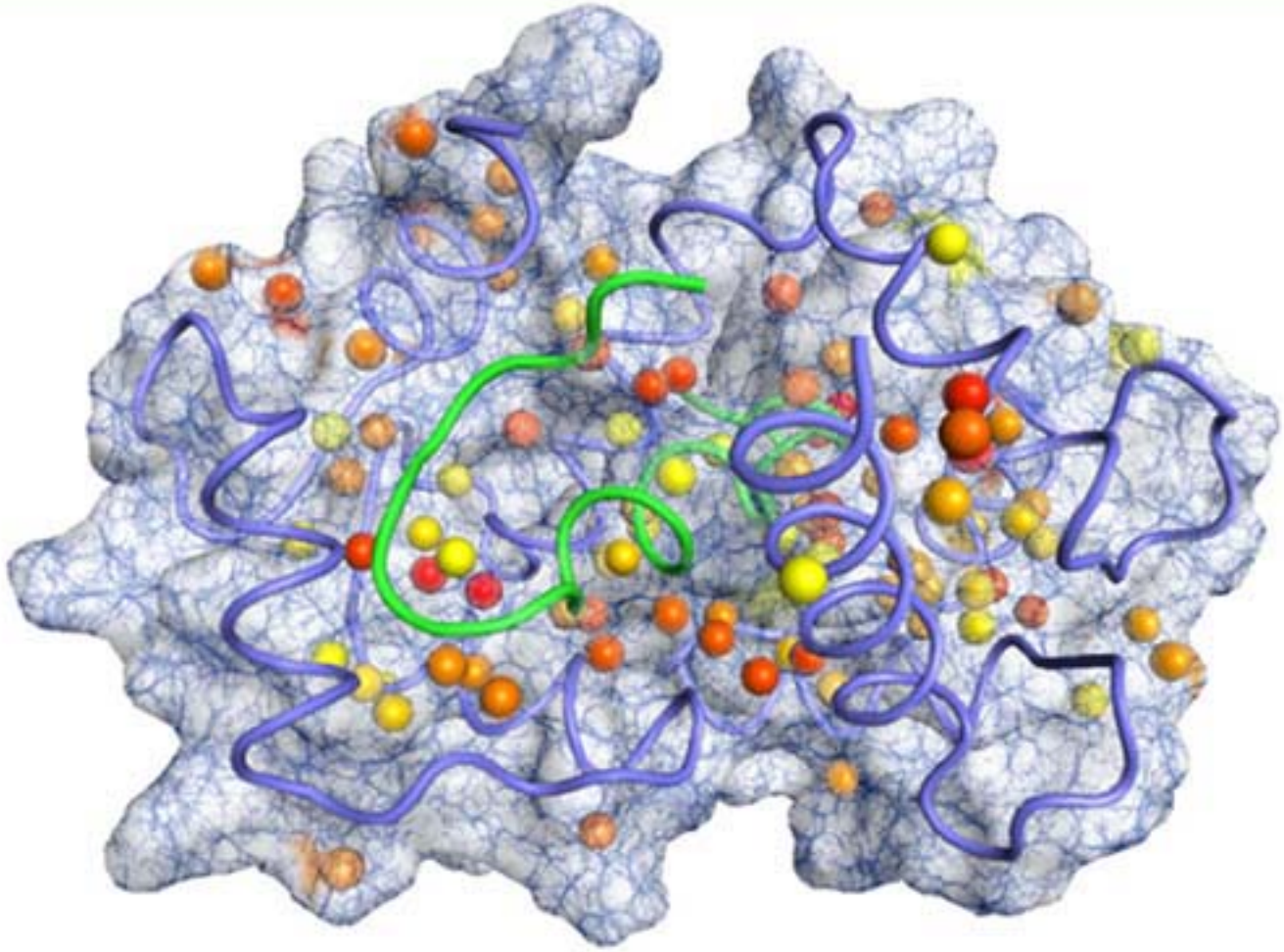
Protein Biology and its Applications



Gilberto Hutton

First Edition, 2012

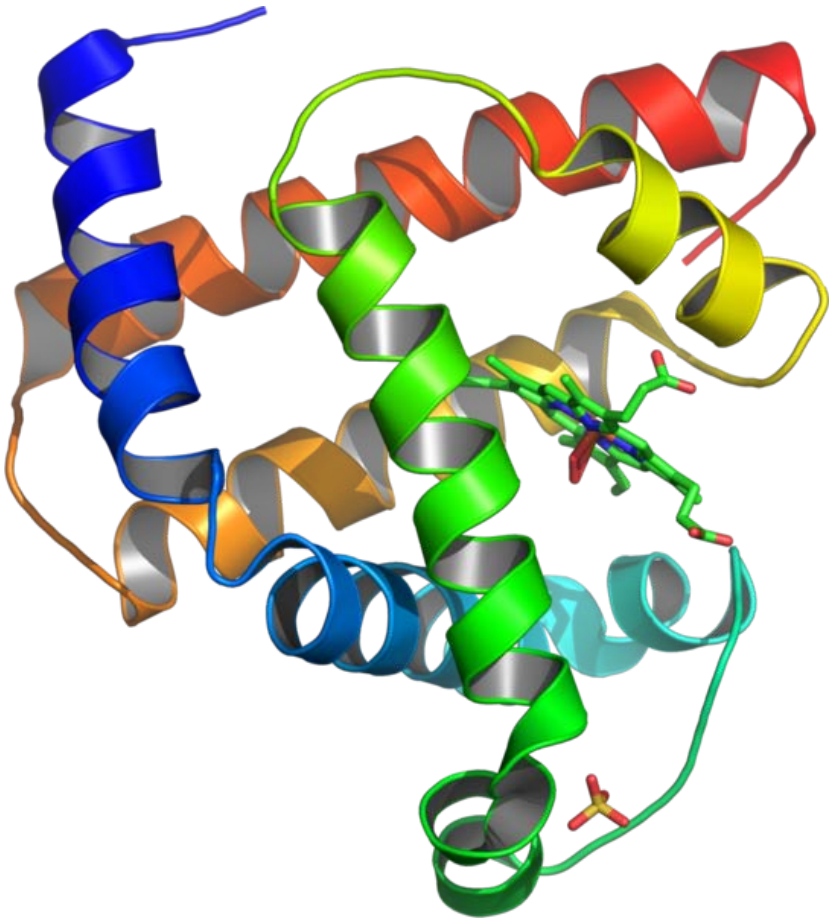
ISBN 978-81-323-4220-5

Published by:
White Word Publications
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1

Protein



A representation of the 3D structure of the protein myoglobin showing colored 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 is shown colored largely in green.

Proteins are biochemical compounds consisting of one or more polypeptides typically folded into a globular or fibrous form in a biologically functional way. A polypeptide is a single linear polymer chain of amino acids bonded together by peptide bonds between the

carboxyl and amino groups of adjacent amino acid residues. The sequence of amino acids 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 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 complexes.

One of the most distinguishing features of polypeptides is their ability to fold into a globular state, or "structure". The extent to which proteins fold into a defined structure varies widely. Some proteins fold into a highly rigid structure with small fluctuations and are therefore considered to be single structure. Other proteins undergo large rearrangements from one conformation to another. This conformational change is often associated with a signaling event. Thus, the structure of a protein serves as a medium through which to regulate either the function of a protein or activity of an enzyme. Not all proteins requiring a folding process in order to function, as some function in an unfolded state.

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 catalyze 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. Proteins are also necessary in animals' diets, since animals cannot synthesize all the amino acids they need and must obtain essential amino acids from food. Through the process of digestion, animals break down ingested protein into free amino acids that are then used in metabolism.

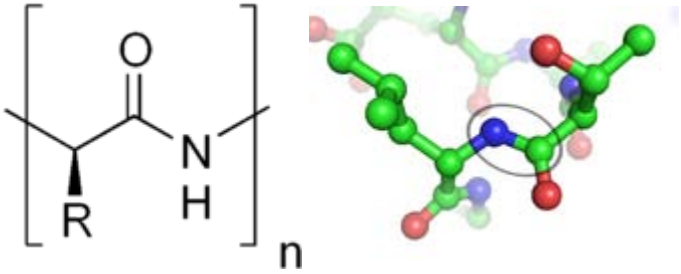
Proteins were first described by the Dutch chemist Gerhardus Johannes Mulder and named by the Swedish chemist Jöns Jakob Berzelius in 1838. Early nutritional scientists such as the German Carl von Voit believed that protein was the most important nutrient for maintaining the structure of the body, because it was generally believed that "flesh makes flesh." The central role of proteins as enzymes in living organisms was however not fully appreciated until 1926, when James B. Sumner showed that the enzyme urease was in fact a protein. The first protein to be sequenced was insulin, by Frederick Sanger, who won the Nobel Prize for this achievement in 1958. The first protein structures to be solved were hemoglobin and myoglobin, by Max Perutz and Sir John Cowdery Kendrew, respectively, in 1958. The three-dimensional structures of both proteins were first determined by X-ray diffraction analysis; Perutz and Kendrew shared the 1962 Nobel Prize in Chemistry for these discoveries. 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, nuclear magnetic resonance and mass spectrometry. Distributed computing is a relatively new tool researchers are using to examine the infamously complex interactions that govern protein folding; the statistical analysis techniques employed to calculate a protein's probable tertiary structure from its amino acid sequence (primary structure) are well-suited for the distributed computing environment, which has made this otherwise prohibitively expensive and time consuming problem significantly more manageable.

Biochemistry

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

Most proteins consist of linear polymers built from series of up to 20 different L-α-amino acids. All proteinogenic amino acids possess common structural features, including an α-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. 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.



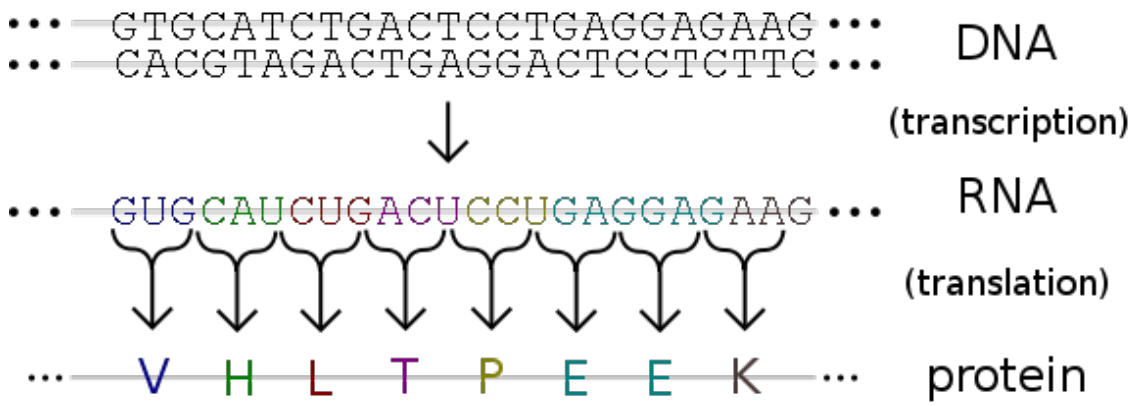
Chemical structure of the peptide bond (left) and a peptide bond between leucine and threonine (right)

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.* 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. The end of the protein with a free carboxyl group is known as the C-

terminus or carboxy terminus, whereas the end with a free amino group is known as the N-terminus or amino terminus.

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. *Polypeptide* can refer to any single linear chain of amino acids, usually regardless of length, but often implies an absence of a defined conformation.

Synthesis



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

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. 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.

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

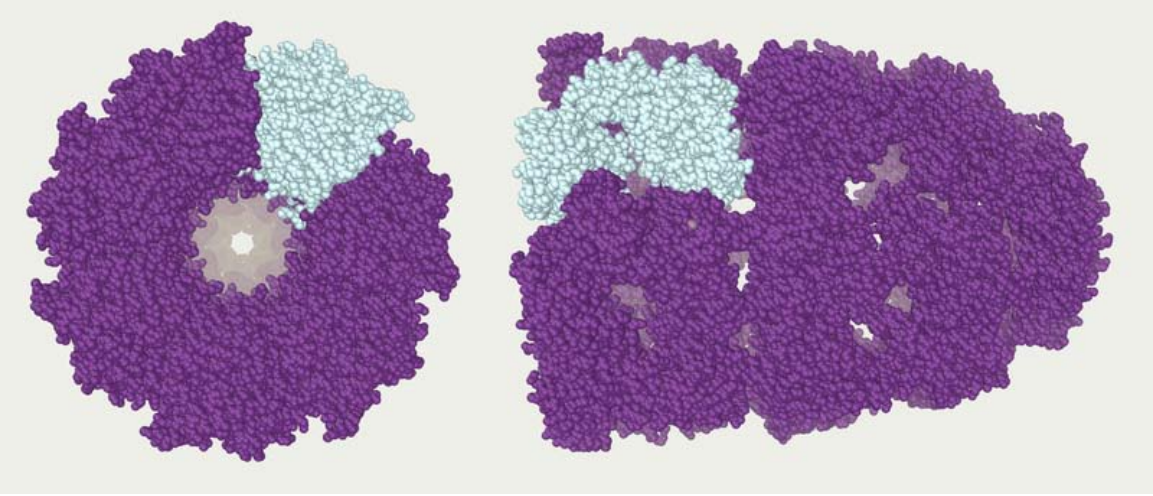
matching each codon to its base pairing anticodon located on a transfer RNA molecule, which carries the amino acid corresponding to the codon it recognizes. The enzyme aminoacyl tRNA synthetase "charges" the tRNA molecules with the correct amino acids. The growing polypeptide is often termed the *nascent chain*. Proteins are always biosynthesized from N-terminus to C-terminus.

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). Yeast proteins are on average 466 amino acids long and 53 kDa in mass. 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.

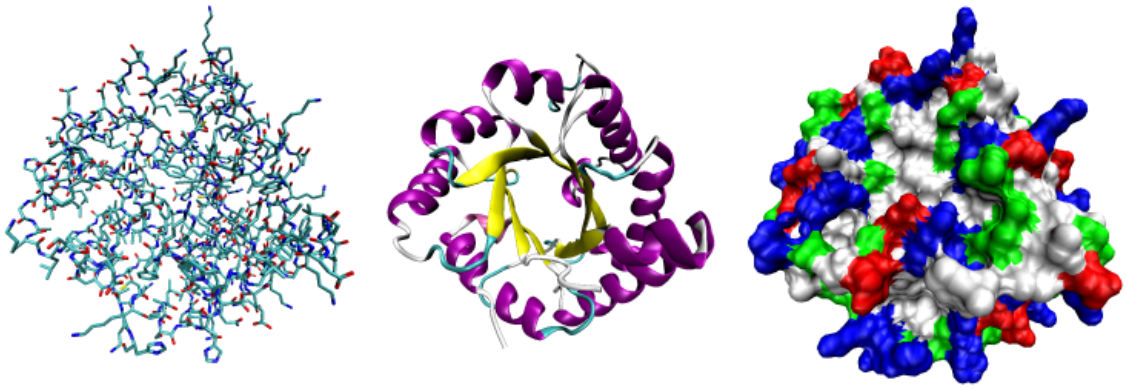
Chemical synthesis

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. 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. 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.

Structure



The crystal structure of the chaperonin. Chaperonins assist protein folding.



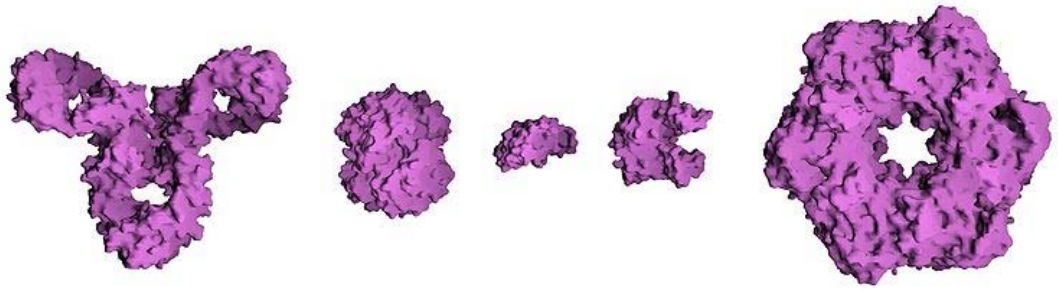
Three possible representations of the three-dimensional structure of the protein triose phosphate isomerase. Left: all-atom representation colored by atom type. Middle: Simplified representation illustrating the backbone conformation, colored by secondary structure. Right: Solvent-accessible surface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

Most proteins fold into unique 3-dimensional structures. The shape into which a protein naturally folds is known as its native conformation. 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. Biochemists often refer to four distinct aspects of a protein's structure:

- *Primary structure*: the amino acid sequence.
- *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 bonds, disulfide bonds, and even post-translational 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.



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).

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.

A special case of intramolecular hydrogen bonds within proteins, poorly shielded from water attack and hence promoting their own dehydration, are called dehydrons.

Structure determination

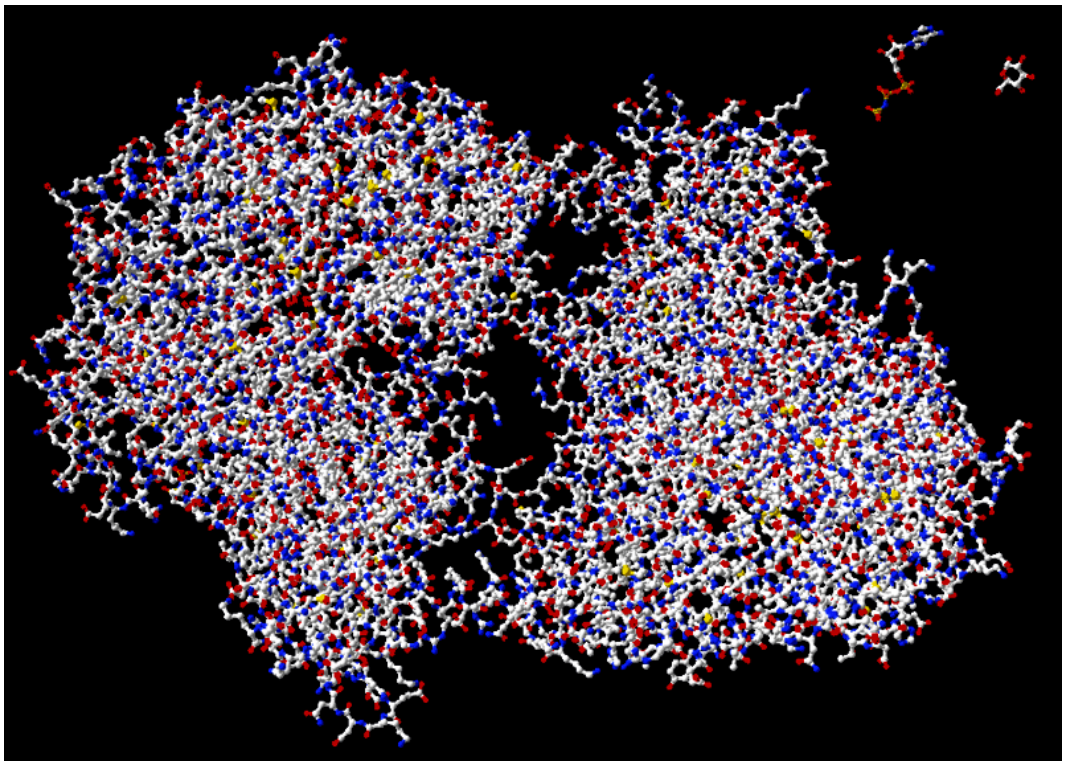
Discovering the tertiary structure of a protein, or the quaternary structure of its complexes, can provide important clues about how the protein performs its function. Common experimental methods of structure determination include X-ray crystallography and NMR spectroscopy, both of which can produce 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/ helical composition of proteins. Cryoelectron microscopy is used to produce lower-resolution structural information about very large protein complexes, including assembled viruses; a variant known as electron crystallography can also produce high-resolution information in some cases , especially for two-dimensional crystals of membrane proteins. 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.

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. 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.

Cellular functions

Proteins are the chief actors within the cell, said to be carrying out the duties specified by the information encoded in genes. 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. The set of proteins expressed in a particular cell or cell type is known as its proteome.



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.

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

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. Importantly, 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.

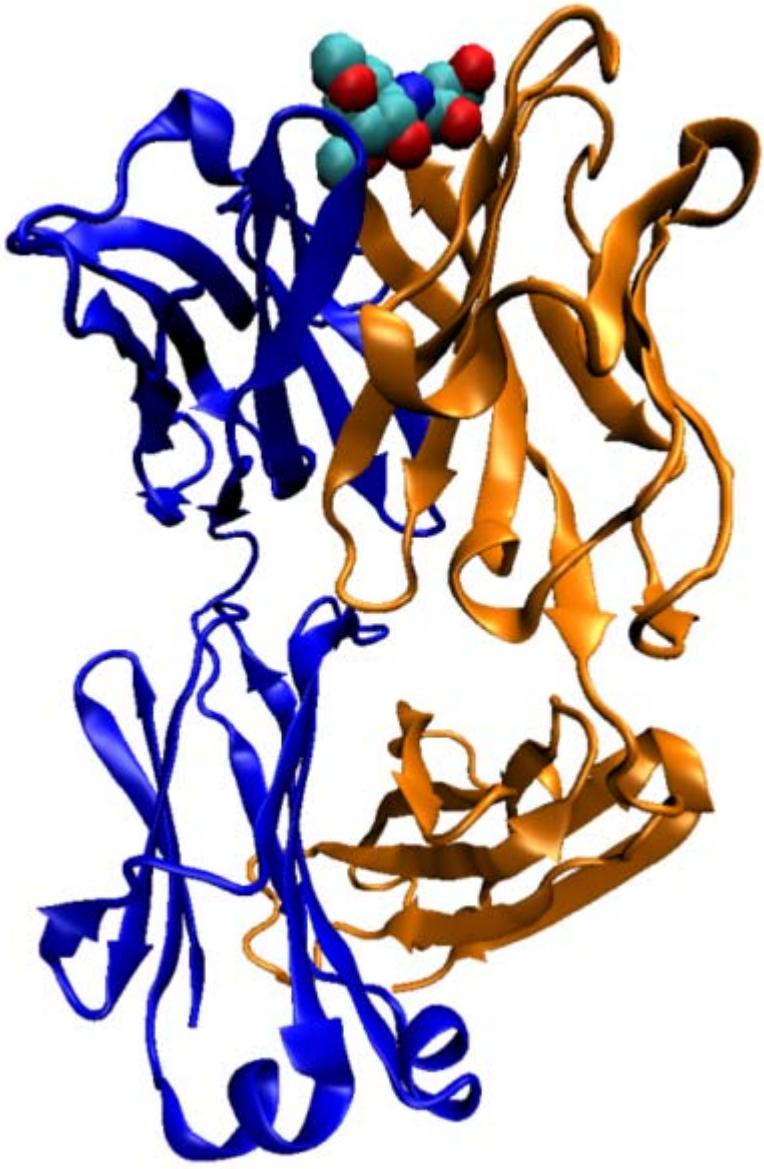
Enzymes

The best-known role of proteins in the cell is as enzymes, which catalyze 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 post-translational modification. About 4,000 reactions are known to be catalyzed by enzymes. The rate acceleration conferred by enzymatic catalysis is often enormous—as much as 10^{17}-fold increase in rate over the uncatalyzed reaction in the case of orotate decarboxylase (78 million years without the enzyme, 18 milliseconds with the enzyme).

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. The region of the

enzyme that binds the substrate and contains the catalytic residues is known as the active site.

Cell signaling and ligand binding



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

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.

Antibodies are protein components of 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.

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. 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. 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.

Structural proteins

Structural proteins confer stiffness and rigidity to otherwise-fluid biological components. Most structural proteins are fibrous proteins; for example, actin and tubulin are globular and soluble as monomers, but polymerize to form long, stiff fibers that comprise the cytoskeleton, which allows the cell to maintain its shape and size. 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.

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.

Methods of study

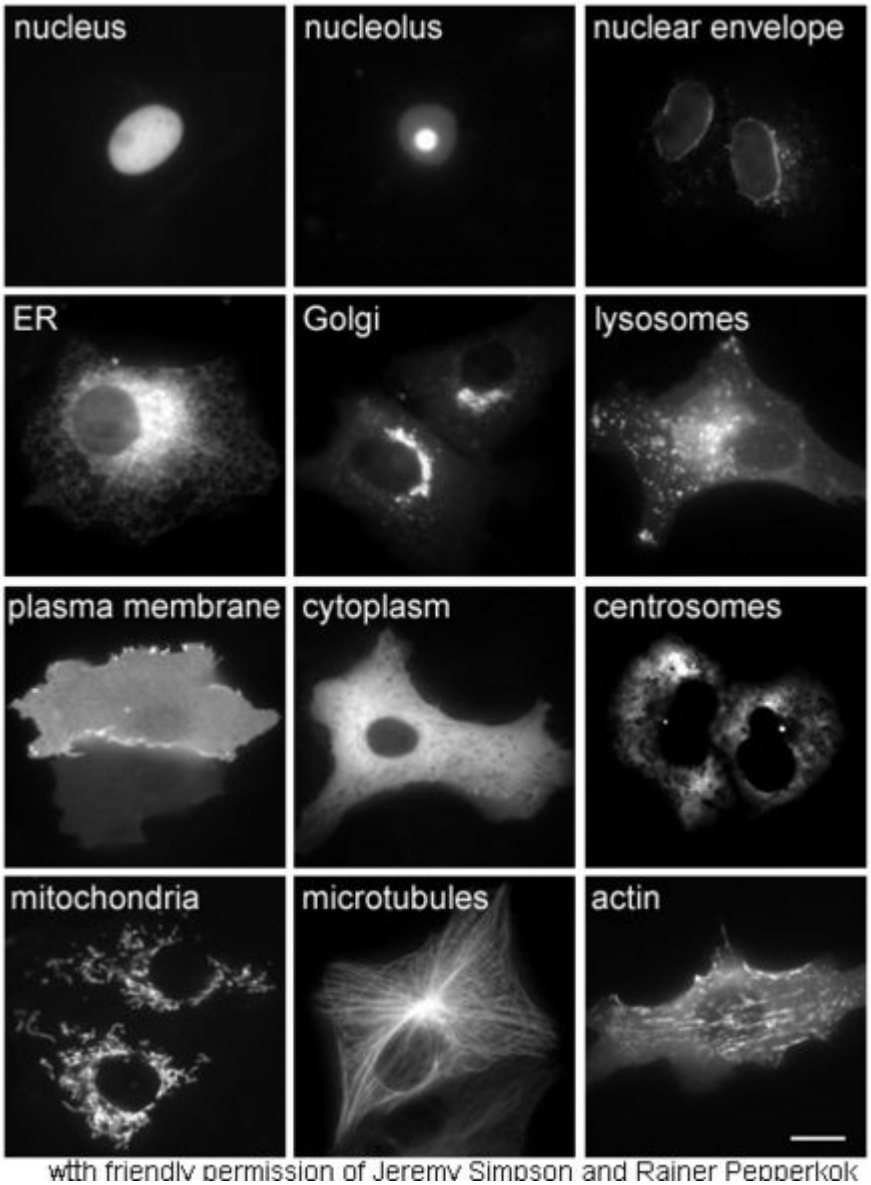
As some of the most commonly studied biological molecules, the activities and structures of proteins are examined both *in vitro* and *in vivo*. *In vitro* studies of purified proteins in controlled environments are useful for learning how a protein carries out its function: for example, enzyme kinetics 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 on proteins' activities within cells or even within whole organisms can provide complementary information about where a protein functions and how it is regulated.

Protein purification

In order 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. 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.

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.

Cellular localization



With friendly permission of Jeremy Simpson and Rainer Pepperkok

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

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). The fused protein's

position within the cell can be cleanly and efficiently visualized using microscopy, 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/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.

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. 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.

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, and may allow the rational design of new proteins with novel properties.

Proteomics and bioinformatics

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, which allows the separation of a large number of proteins, mass spectrometry, which allows rapid high-throughput identification of proteins and sequencing of peptides (most often after in-gel digestion), protein microarrays, 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. The total complement of biologically possible such interactions is known as the interactome. A systematic attempt to determine the structures of proteins representing every possible fold is known as structural genomics.

The large amount of genomic and proteomic data available for a variety of organisms, including the human genome, allows researchers to efficiently identify homologous proteins in distantly related organisms by sequence alignment. Sequence profiling tools can perform more specific sequence manipulations such as restriction enzyme maps, open reading frame analyses for nucleotide sequences, and secondary structure prediction. From this data 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 seeks to assemble, annotate, and analyze genomic and proteomic data, applying computational techniques to biological problems such as gene finding and cladistics.

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. 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. 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. Many structure prediction methods have served to inform the emerging field of protein engineering, in which novel protein folds have already been designed. A more complex computational problem is the prediction of intermolecular interactions, such as in molecular docking and protein–protein interaction prediction.

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; molecular modeling on GPU). The folding of small alpha-helical protein domains such as the villin headpiece and the HIV accessory protein 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.

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. 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 catalyzes 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. 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. Noted examples at the time included albumin from egg whites, blood serum albumin, fibrin, and wheat gluten. Dutch chemist Gerhardus Johannes Mulder carried out elemental analysis of common proteins and found that nearly all proteins had the same empirical formula, $C_{400}H_{620}N_{100}O_{120}P_1S_1$. 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 in 1838 by Mulder's associate Jöns Jakob Berzelius; protein is derived from the Greek word πρωτεῖος (*proteios*), meaning "primary", "in the lead", or "standing in front". 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.

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.

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. Later work by Walter Kauzmann on denaturation, based partly on previous studies by Kaj Linderstrøm-Lang, contributed an understanding of protein folding and structure mediated by hydrophobic interactions. In 1949 Fred 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. The first atomic-resolution structures of proteins were solved by X-ray crystallography in the 1960s and by NMR in the 1980s. As of 2009, the Protein Data Bank has over 55,000 atomic-resolution structures of proteins. In more recent times, cryo-electron microscopy of large macromolecular assemblies and computational protein structure prediction of small protein domains are two methods approaching atomic resolution.

Chapter 2

Enzyme

Human glyoxalase I. Two zinc ions that are needed for the enzyme to catalyze its reaction are shown as purple spheres, and an enzyme inhibitor called *S*-hexylglutathione is shown as a space-filling model, filling the two active sites.

Enzymes are proteins that catalyze (*i.e.*, increase or decrease the rates of) chemical reactions. In enzymatic reactions, the molecules at the beginning of the process are called substrates, and they are converted into different molecules, called the products. Almost all processes in a biological cell need enzymes to occur at significant rates. Since enzymes are selective for their substrates and speed up only a few reactions from among many possibilities, the set of enzymes made in a cell determines which metabolic pathways occur in that cell.

Like all catalysts, enzymes work by lowering the activation energy ($E_a^{\ddagger}$) for a reaction, thus dramatically increasing the rate of the reaction. As a result, products are formed faster and reactions reach their equilibrium state more rapidly. Most enzyme reaction rates are millions of times faster than those of comparable un-catalyzed reactions. As with all catalysts, enzymes are not consumed by the reactions they catalyze, nor do they alter the equilibrium of these reactions. However, enzymes do differ from most other

catalysts by being much more specific. Enzymes are known to catalyze about 4,000 biochemical reactions. A few RNA molecules called ribozymes also catalyze reactions, with an important example being some parts of the ribosome. Synthetic molecules called artificial enzymes also display enzyme-like catalysis.

Enzyme activity can be affected by other molecules. Inhibitors are molecules that decrease enzyme activity; activators are molecules that increase activity. Many drugs and poisons are enzyme inhibitors. Activity is also affected by temperature, chemical environment (*e.g.*, pH), and the concentration of substrate. Some enzymes are used commercially, for example, in the synthesis of antibiotics. In addition, some household products use enzymes to speed up biochemical reactions (*e.g.*, enzymes in biological washing powders break down protein or fat stains on clothes; enzymes in meat tenderizers break down proteins into smaller molecules, making the meat easier to chew).

Etymology and history

Eduard Buchner

As early as the late 17th and early 18th centuries, the digestion of meat by stomach secretions and the conversion of starch to sugars by plant extracts and saliva were known. However, the mechanism by which this occurred had not been identified.

In the 19th century, when studying the fermentation of sugar to alcohol by yeast, Louis Pasteur came to the conclusion that this fermentation was catalyzed by a vital force contained within the yeast cells called "ferments", which were thought to function only within living organisms. He wrote that "alcoholic fermentation is an act correlated with the life and organization of the yeast cells, not with the death or putrefaction of the cells."

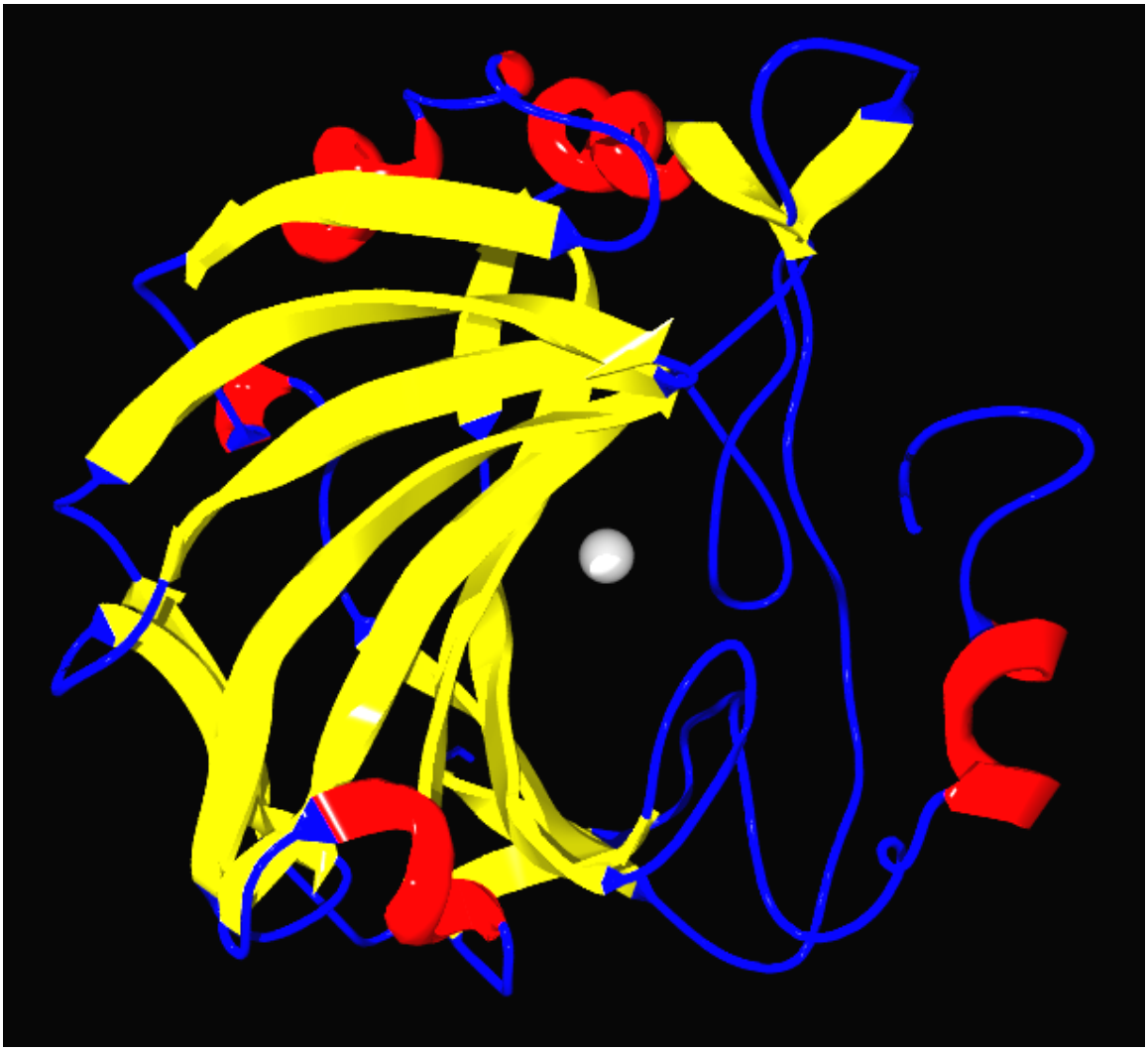
In 1877, German physiologist Wilhelm Kühne (1837–1900) first used the term *enzyme*, which comes from Greek *ενζυμον*, "in leaven", to describe this process. The word *enzyme* was used later to refer to nonliving substances such as pepsin, and the word *ferment* was used to refer to chemical activity produced by living organisms.

In 1897, Eduard Buchner submitted his first paper on the ability of yeast extracts that lacked any living yeast cells to ferment sugar. In a series of experiments at the University of Berlin, he found that the sugar was fermented even when there were no living yeast cells in the mixture. He named the enzyme that brought about the fermentation of sucrose "zymase". In 1907, he received the Nobel Prize in Chemistry "for his biochemical research and his discovery of cell-free fermentation". Following Buchner's example, enzymes are usually named according to the reaction they carry out. Typically, to generate the name of an enzyme, the suffix *-ase* is added to the name of its substrate (*e.g.*, lactase is the enzyme that cleaves lactose) or the type of reaction (*e.g.*, DNA polymerase forms DNA polymers).

Having shown that enzymes could function outside a living cell, the next step was to determine their biochemical nature. Many early workers noted that enzymatic activity was associated with proteins, but several scientists (such as Nobel laureate Richard Willstätter) argued that proteins were merely carriers for the true enzymes and that proteins *per se* were incapable of catalysis. However, in 1926, James B. Sumner showed that the enzyme urease was a pure protein and crystallized it; Sumner did likewise for the enzyme catalase in 1937. The conclusion that pure proteins can be enzymes was definitively proved by Northrop and Stanley, who worked on the digestive enzymes pepsin (1930), trypsin and chymotrypsin. These three scientists were awarded the 1946 Nobel Prize in Chemistry.

This discovery that enzymes could be crystallized eventually allowed their structures to be solved by x-ray crystallography. This was first done for lysozyme, an enzyme found in tears, saliva and egg whites that digests the coating of some bacteria; the structure was solved by a group led by David Chilton Phillips and published in 1965. This high-resolution structure of lysozyme marked the beginning of the field of structural biology and the effort to understand how enzymes work at an atomic level of detail.

Structures and mechanisms



Ribbon diagram showing human carbonic anhydrase II. The grey sphere is the zinc cofactor in the active site. Diagram drawn from PDB 1MOO.

Enzymes are generally globular proteins and range from just 62 amino acid residues in size, for the monomer of 4-oxalocrotonate tautomerase, to over 2,500 residues in the animal fatty acid synthase. A small number of RNA-based biological catalysts exist, with the most common being the ribosome; these are referred to as either RNA-enzymes or ribozymes. The activities of enzymes are determined by their three-dimensional structure. However, although structure does determine function, predicting a novel enzyme's activity just from its structure is a very difficult problem that has not yet been solved.

Most enzymes are much larger than the substrates they act on, and only a small portion of the enzyme (around 3–4 amino acids) is directly involved in catalysis. The region that contains these catalytic residues, binds the substrate, and then carries out the reaction is known as the active site. Enzymes can also contain sites that bind cofactors, which are

needed for catalysis. Some enzymes also have binding sites for small molecules, which are often direct or indirect products or substrates of the reaction catalyzed. This binding can serve to increase or decrease the enzyme's activity, providing a means for feedback regulation.

Like all proteins, enzymes are long, linear chains of amino acids that fold to produce a three-dimensional product. Each unique amino acid sequence produces a specific structure, which has unique properties. Individual protein chains may sometimes group together to form a protein complex. Most enzymes can be denatured—that is, unfolded and inactivated—by heating or chemical denaturants, which disrupt the three-dimensional structure of the protein. Depending on the enzyme, denaturation may be reversible or irreversible.

Structures of enzymes in complex with substrates or substrate analogs during a reaction may be obtained using Time resolved crystallography methods.

Specificity

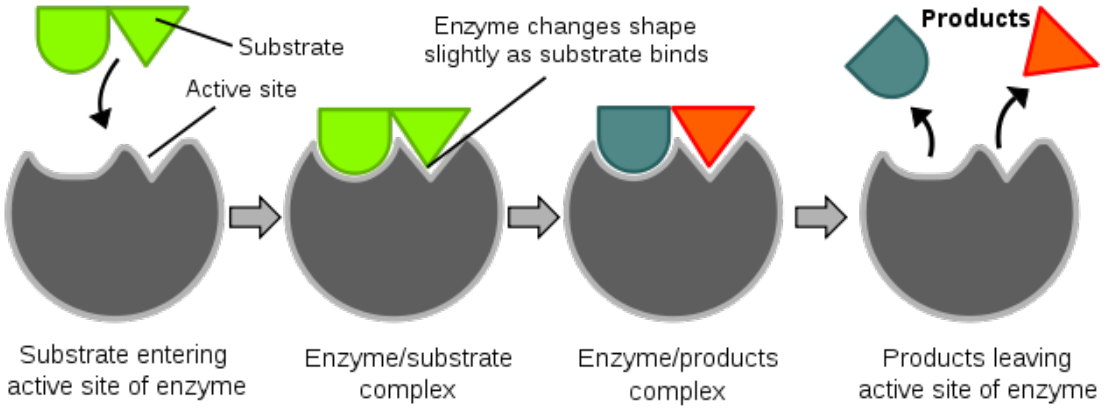
Enzymes are usually very specific as to which reactions they catalyze and the substrates that are involved in these reactions. Complementary shape, charge and hydrophilic/hydrophobic characteristics of enzymes and substrates are responsible for this specificity. Enzymes can also show impressive levels of stereospecificity, regioselectivity and chemoselectivity.

Some of the enzymes showing the highest specificity and accuracy are involved in the copying and expression of the genome. These enzymes have "proof-reading" mechanisms. Here, an enzyme such as DNA polymerase catalyzes a reaction in a first step and then checks that the product is correct in a second step. This two-step process results in average error rates of less than 1 error in 100 million reactions in high-fidelity mammalian polymerases. Similar proofreading mechanisms are also found in RNA polymerase, aminoacyl tRNA synthetases and ribosomes.

Some enzymes that produce secondary metabolites are described as promiscuous, as they can act on a relatively broad range of different substrates. It has been suggested that this broad substrate specificity is important for the evolution of new biosynthetic pathways.

"Lock and key" model

Enzymes are very specific, and it was suggested by the Nobel laureate organic chemist Emil Fischer in 1894 that this was because both the enzyme and the substrate possess specific complementary geometric shapes that fit exactly into one another. This is often referred to as "the lock and key" model. However, while this model explains enzyme specificity, it fails to explain the stabilization of the transition state that enzymes achieve.



Diagrams to show the induced fit hypothesis of enzyme action

In 1958, Daniel Koshland suggested a modification to the lock and key model: since enzymes are rather flexible structures, the active site is continually reshaped by interactions with the substrate as the substrate interacts with the enzyme. As a result, the substrate does not simply bind to a rigid active site; the amino acid side chains which make up the active site are molded into the precise positions that enable the enzyme to perform its catalytic function. In some cases, such as glycosidases, the substrate molecule also changes shape slightly as it enters the active site. The active site continues to change until the substrate is completely bound, at which point the final shape and charge is determined. Induced fit may enhance the fidelity of molecular recognition in the presence of competition and noise via the conformational proofreading mechanism .

Mechanisms

Enzymes can act in several ways, all of which lower $\Delta G^{\ddagger}$:

- Lowering the activation energy by creating an environment in which the transition state is stabilized (e.g. straining the shape of a substrate—by binding the transition-state conformation of the substrate/product molecules, the enzyme distorts the bound substrate(s) into their transition state form, thereby reducing the amount of energy required to complete the transition).
- Lowering the energy of the transition state, but without distorting the substrate, by creating an environment with the opposite charge distribution to that of the transition state.
- Providing an alternative pathway. For example, temporarily reacting with the substrate to form an intermediate ES complex, which would be impossible in the absence of the enzyme.
- Reducing the reaction entropy change by bringing substrates together in the correct orientation to react. Considering $\Delta H^{\ddagger}$ alone overlooks this effect.
- Increases in temperatures speed up reactions. Thus, temperature increases help the enzyme function and develop the end product even faster. However, if heated too

much, the enzyme’s shape deteriorates and the enzyme becomes denatured. Some enzymes like thermolabile enzymes work best at low temperatures.

Interestingly, this entropic effect involves destabilization of the ground state, and its contribution to catalysis is relatively small.

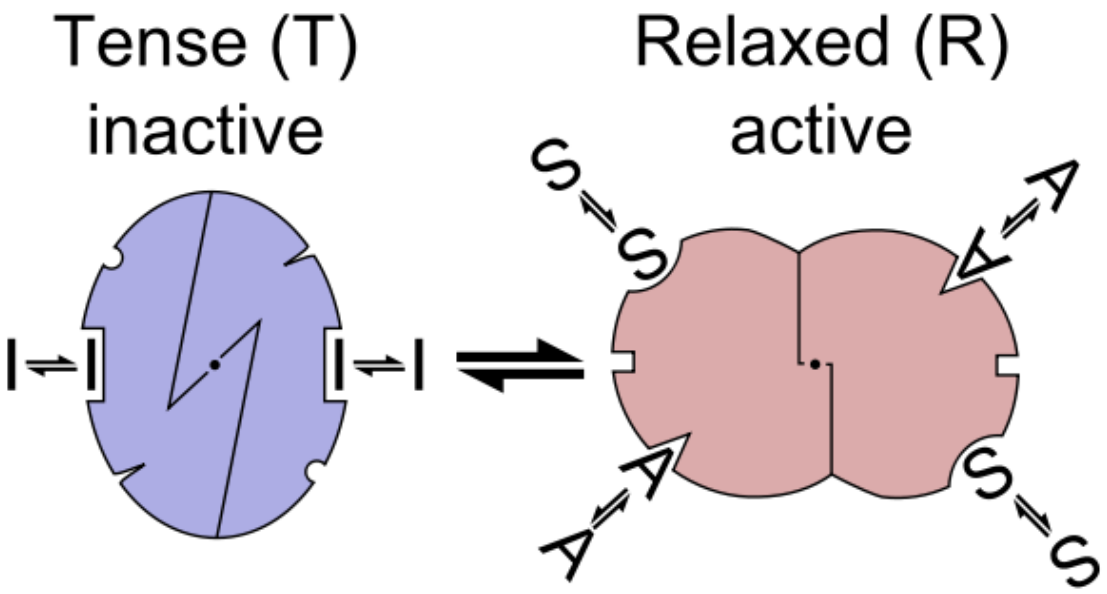
Transition State Stabilization

The understanding of the origin of the reduction of $\Delta G^{\ddagger}$ requires one to find out how the enzymes can stabilize its transition state more than the transition state of the uncatalyzed reaction. Apparently, the most effective way for reaching large stabilization is the use of electrostatic effects, in particular, by having a relatively fixed polar environment that is oriented toward the charge distribution of the transition state. Such an environment does not exist in the uncatalyzed reaction in water.

Dynamics and function

The internal dynamics of enzymes is linked to their mechanism of catalysis. Internal dynamics are the movement of parts of the enzyme's structure, such as individual amino acid residues, a group of amino acids, or even an entire protein domain. These movements occur at various time-scales ranging from femtoseconds to seconds. Networks of protein residues throughout an enzyme's structure can contribute to catalysis through dynamic motions. Protein motions are vital to many enzymes, but whether small and fast vibrations, or larger and slower conformational movements are more important depends on the type of reaction involved. However, although these movements are important in binding and releasing substrates and products, it is not clear if protein movements help to accelerate the chemical steps in enzymatic reactions. These new insights also have implications in understanding allosteric effects and developing new drugs.

Allosteric modulation



Allosteric transition of an enzyme between R and T states, stabilized by an agonist, an inhibitor and a substrate (the MWC model)

Allosteric sites are sites on the enzyme that bind to molecules in the cellular environment. The sites form weak, noncovalent bonds with these molecules, causing a change in the conformation of the enzyme. This change in conformation translates to the active site, which then affects the reaction rate of the enzyme. Allosteric interactions can both inhibit and activate enzymes and are a common way that enzymes are controlled in the body.

Cofactors and coenzymes

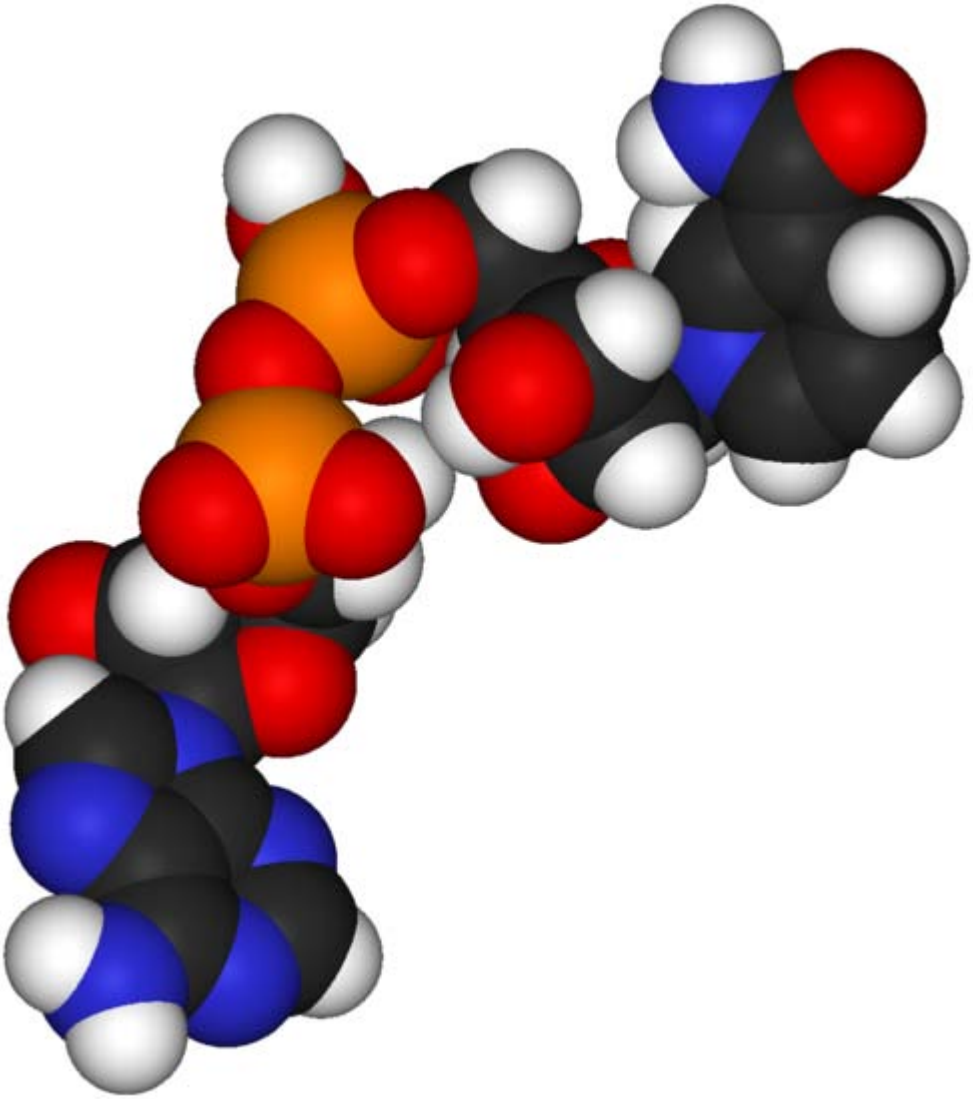
Cofactors

Some enzymes do not need any additional components to show full activity. However, others require non-protein molecules called cofactors to be bound for activity. Cofactors can be either inorganic (*e.g.*, metal ions and iron-sulfur clusters) or organic compounds (e.g., flavin and heme). Organic cofactors can be either prosthetic groups, which are tightly bound to an enzyme, or coenzymes, which are released from the enzyme's active site during the reaction. Coenzymes include NADH, NADPH and adenosine triphosphate. These molecules transfer chemical groups between enzymes.

An example of an enzyme that contains a cofactor is carbonic anhydrase, and is shown in the ribbon diagram above with a zinc cofactor bound as part of its active site. These tightly bound molecules are usually found in the active site and are involved in catalysis. For example, flavin and heme cofactors are often involved in redox reactions.

Enzymes that require a cofactor but do not have one bound are called *apoenzymes* or *apoproteins*. An apoenzyme together with its cofactor(s) is called a *holoenzyme* (this is the active form). Most cofactors are not covalently attached to an enzyme, but are very tightly bound. However, organic prosthetic groups can be covalently bound (*e.g.*, thiamine pyrophosphate in the enzyme pyruvate dehydrogenase). The term "holoenzyme" can also be applied to enzymes that contain multiple protein subunits, such as the DNA polymerases; here the holoenzyme is the complete complex containing all the subunits needed for activity.

Coenzymes



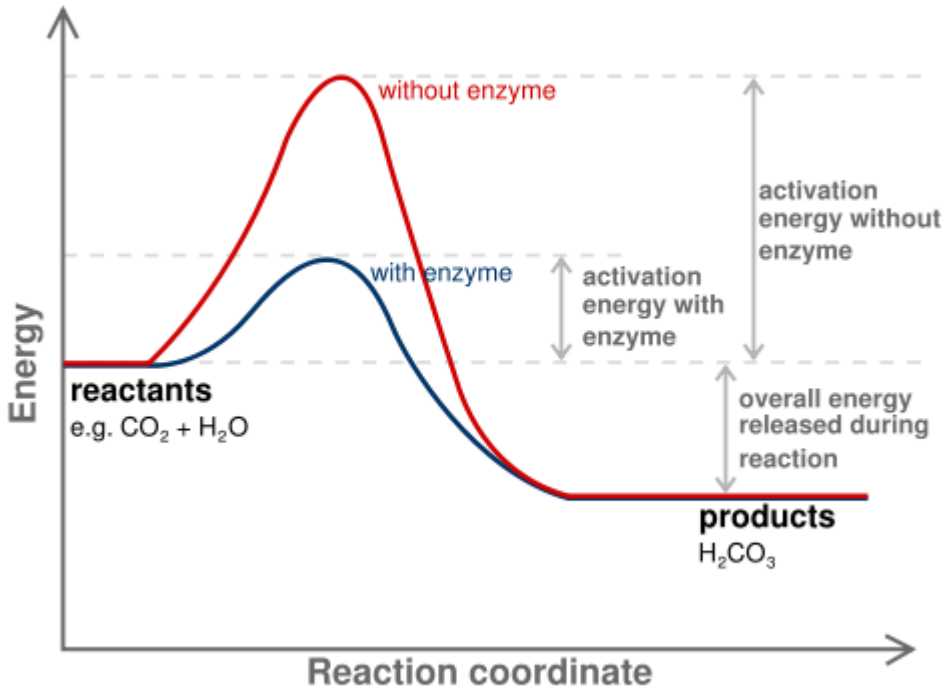
Space-filling model of the coenzyme NADH

Coenzymes are small organic molecules that can be loosely or tightly bound to an enzyme. Tightly bound coenzymes can be called allosteric groups. Coenzymes transport chemical groups from one enzyme to another. Some of these chemicals such as riboflavin, thiamine and folic acid are vitamins (compounds which cannot be synthesized by the body and must be acquired from the diet). The chemical groups carried include the hydride ion (H^-) carried by NAD or $NADP^+$, the phosphate group carried by adenosine triphosphate, the acetyl group carried by coenzyme A, formyl, methenyl or methyl groups carried by folic acid and the methyl group carried by S-adenosylmethionine.

Since coenzymes are chemically changed as a consequence of enzyme action, it is useful to consider coenzymes to be a special class of substrates, or second substrates, which are common to many different enzymes. For example, about 700 enzymes are known to use the coenzyme NADH.

Coenzymes are usually continuously regenerated and their concentrations maintained at a steady level inside the cell: for example, NADPH is regenerated through the pentose phosphate pathway and *S*-adenosylmethionine by methionine adenosyltransferase. This continuous regeneration means that even small amounts of coenzymes are used very intensively. For example, the human body turns over its own weight in ATP each day.

Thermodynamics



The energies of the stages of a chemical reaction. Substrates need a lot of energy to reach a transition state, which then decays into products. The enzyme stabilizes the transition state, reducing the energy needed to form products.

As all catalysts, enzymes do not alter the position of the chemical equilibrium of the reaction. Usually, in the presence of an enzyme, the reaction runs in the same direction as it would without the enzyme, just more quickly. However, in the absence of the enzyme, other possible uncatalyzed, "spontaneous" reactions might lead to different products, because in those conditions this different product is formed faster.

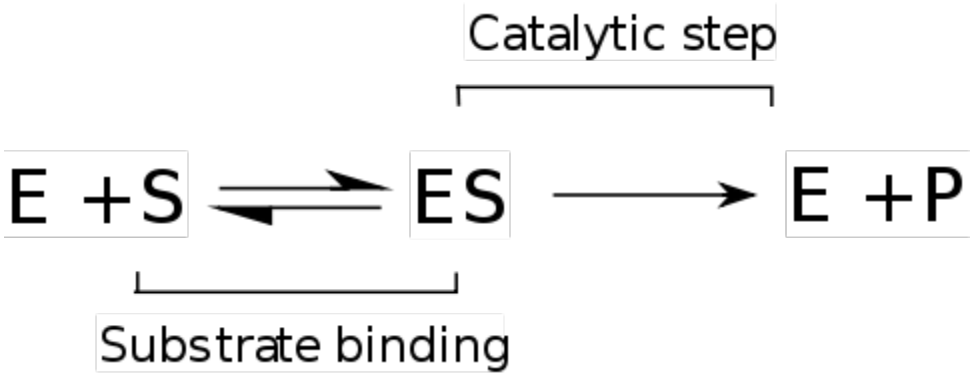
Furthermore, enzymes can couple two or more reactions, so that a thermodynamically favorable reaction can be used to "drive" a thermodynamically unfavorable one. For example, the hydrolysis of ATP is often used to drive other chemical reactions.

Enzymes catalyze the forward and backward reactions equally. They do not alter the equilibrium itself, but only the speed at which it is reached. For example, carbonic anhydrase catalyzes its reaction in either direction depending on the concentration of its reactants.

$$\mathrm{CO_2 + H_2O \xrightarrow{Carbonic\ anhydrase} H_2CO_3} \text{(in tissues; high } \mathrm{CO_2} \text{ concentration)}$$
$$\mathrm{H_2CO_3 \xrightarrow{Carbonic\ anhydrase} CO_2 + H_2O} \text{(in lungs; low } \mathrm{CO_2} \text{ concentration)}$$

Nevertheless, if the equilibrium is greatly displaced in one direction, that is, in a very exergonic reaction, the reaction is *effectively* irreversible. Under these conditions the enzyme will, in fact, only catalyze the reaction in the thermodynamically allowed direction.

Kinetics



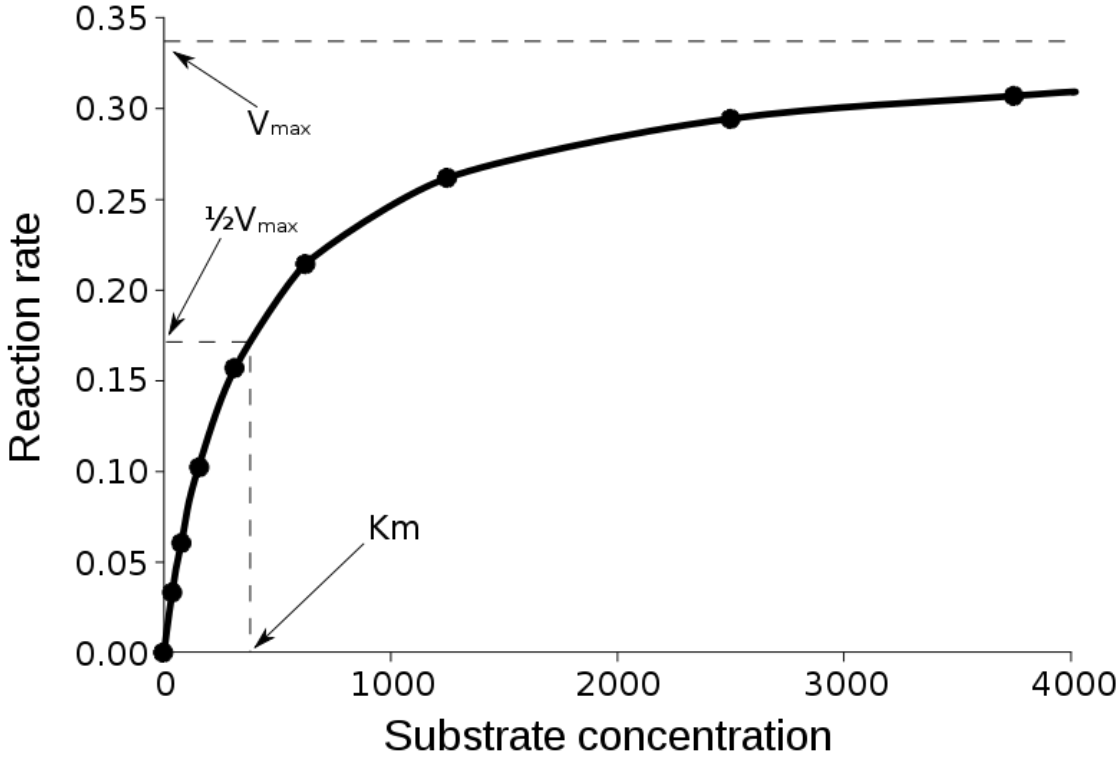
Mechanism for a single substrate enzyme catalyzed reaction. The enzyme (E) binds a substrate (S) and produces a product (P).

Enzyme kinetics is the investigation of how enzymes bind substrates and turn them into products. The rate data used in kinetic analyses are obtained from enzyme assays.

In 1902 Victor Henri proposed a quantitative theory of enzyme kinetics, but his experimental data were not useful because the significance of the hydrogen ion concentration was not yet appreciated. After Peter Lauritz Sørensen had defined the logarithmic pH-scale and introduced the concept of buffering in 1909 the German

chemist Leonor Michaelis and his Canadian postdoc Maud Leonora Menten repeated Henri's experiments and confirmed his equation which is referred to as Henri-Michaelis-Menten kinetics (sometimes also Michaelis-Menten kinetics). Their work was further developed by G. E. Briggs and J. B. S. Haldane, who derived kinetic equations that are still widely used today.

The major contribution of Henri was to think of enzyme reactions in two stages. In the first, the substrate binds reversibly to the enzyme, forming the enzyme-substrate complex. This is sometimes called the Michaelis complex. The enzyme then catalyzes the chemical step in the reaction and releases the product.



Saturation curve for an enzyme reaction showing the relation between the substrate concentration (S) and rate (v)

Enzymes can catalyze up to several million reactions per second. For example, the uncatalyzed decarboxylation of orotidine 5'-monophosphate has a half life of 78 million years. However, when the enzyme orotidine 5'-phosphate decarboxylase is added, the same process takes just 25 milliseconds. Enzyme rates depend on solution conditions and substrate concentration. Conditions that denature the protein abolish enzyme activity, such as high temperatures, extremes of pH or high salt concentrations, while raising substrate concentration tends to increase activity. To find the maximum speed of an enzymatic reaction, the substrate concentration is increased until a constant rate of product formation is seen. This is shown in the saturation curve on the right. Saturation happens because, as substrate concentration increases, more and more of the free enzyme

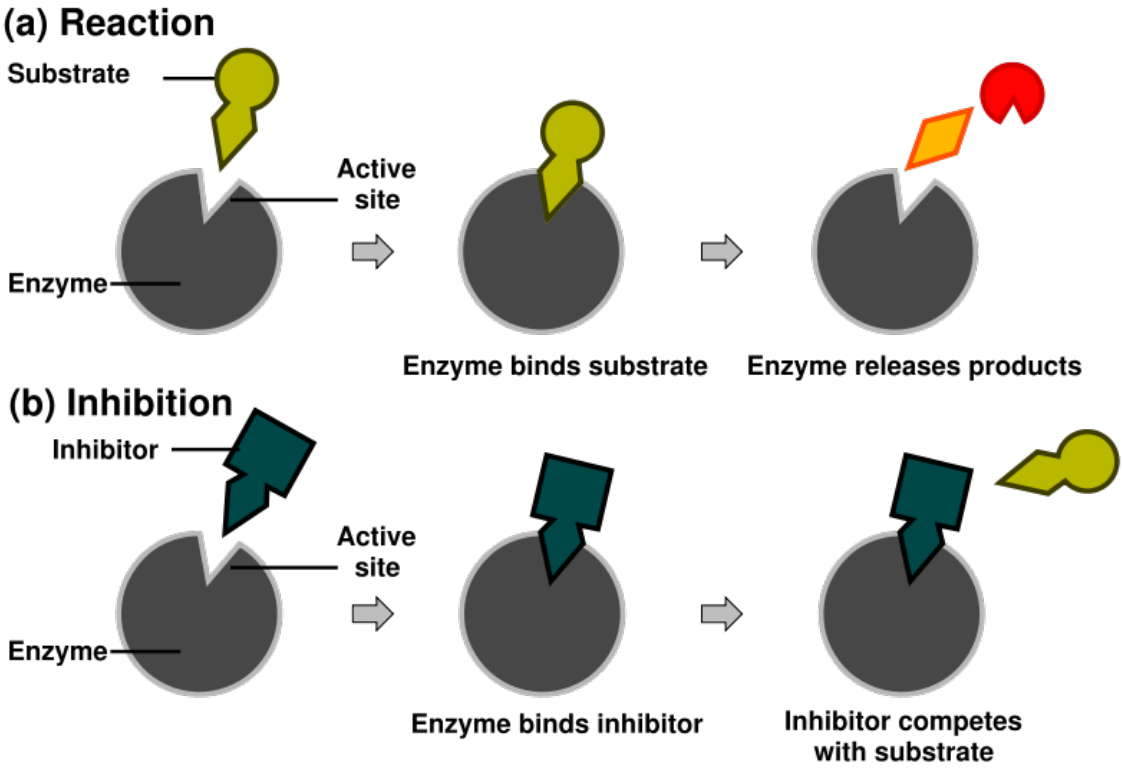
is converted into the substrate-bound ES form. At the maximum reaction rate (V_{max}) of the enzyme, all the enzyme active sites are bound to substrate, and the amount of ES complex is the same as the total amount of enzyme. However, V_{max} is only one kinetic constant of enzymes. The amount of substrate needed to achieve a given rate of reaction is also important. This is given by the Michaelis-Menten constant (K_m), which is the substrate concentration required for an enzyme to reach one-half its maximum reaction rate. Each enzyme has a characteristic K_m for a given substrate, and this can show how tight the binding of the substrate is to the enzyme. Another useful constant is k_{cat}, which is the number of substrate molecules handled by one active site per second.

The efficiency of an enzyme can be expressed in terms of k_{cat}/K_m. This is also called the specificity constant and incorporates the rate constants for all steps in the reaction. Because the specificity constant reflects both affinity and catalytic ability, it is useful for comparing different enzymes against each other, or the same enzyme with different substrates. The theoretical maximum for the specificity constant is called the diffusion limit and is about 10^8 to 10^9 (M^{-1} s^{-1}). At this point every collision of the enzyme with its substrate will result in catalysis, and the rate of product formation is not limited by the reaction rate but by the diffusion rate. Enzymes with this property are called *catalytically perfect* or *kinetically perfect*. Example of such enzymes are triose-phosphate isomerase, carbonic anhydrase, acetylcholinesterase, catalase, fumarase, β-lactamase, and superoxide dismutase.

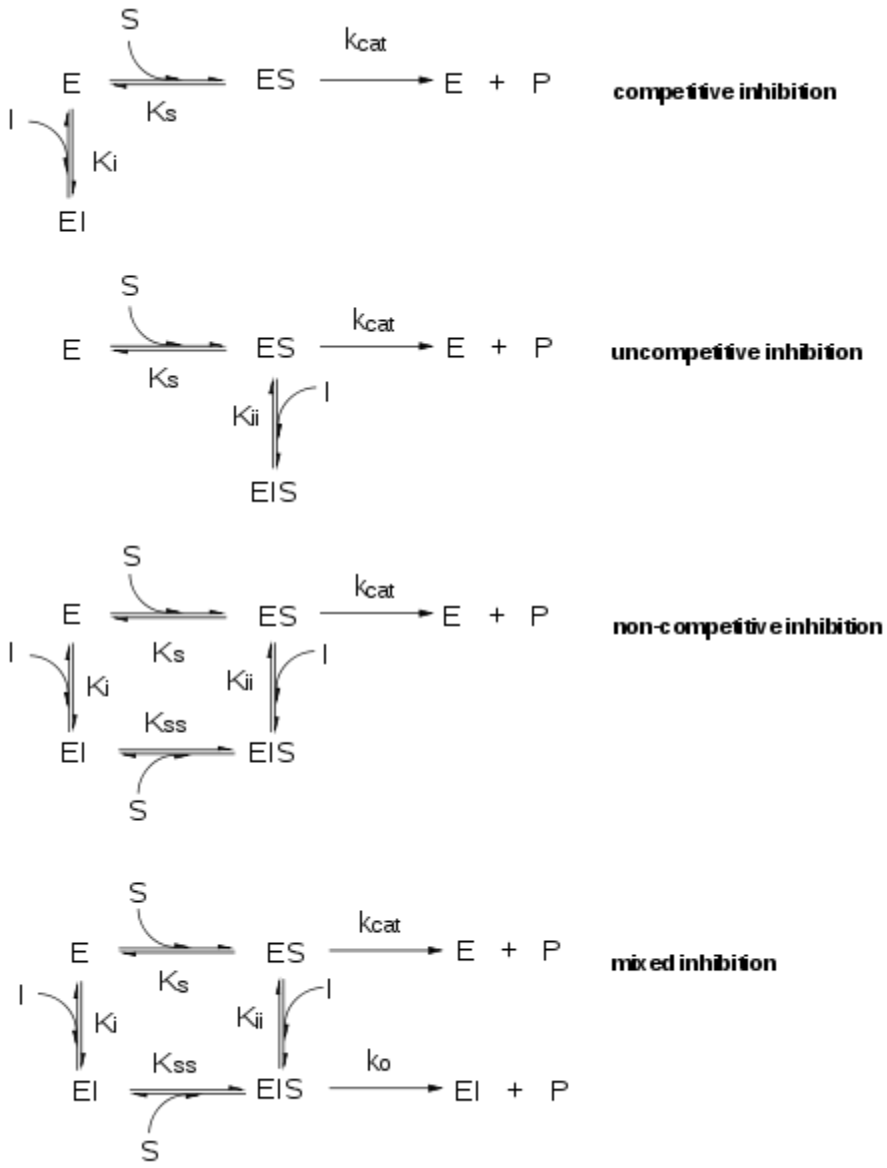
Michaelis-Menten kinetics relies on the law of mass action, which is derived from the assumptions of free diffusion and thermodynamically driven random collision. However, many biochemical or cellular processes deviate significantly from these conditions, because of macromolecular crowding, phase-separation of the enzyme/substrate/product, or one or two-dimensional molecular movement. In these situations, a fractal Michaelis-Menten kinetics may be applied.

Some enzymes operate with kinetics which are faster than diffusion rates, which would seem to be impossible. Several mechanisms have been invoked to explain this phenomenon. Some proteins are believed to accelerate catalysis by drawing their substrate in and pre-orienting them by using dipolar electric fields. Other models invoke a quantum-mechanical tunneling explanation, whereby a proton or an electron can tunnel through activation barriers, although for proton tunneling this model remains somewhat controversial. Quantum tunneling for protons has been observed in tryptamine. This suggests that enzyme catalysis may be more accurately characterized as "through the barrier" rather than the traditional model, which requires substrates to go "over" a lowered energy barrier.

Inhibition



Competitive inhibitors bind reversibly to the enzyme, preventing the binding of substrate. On the other hand, binding of substrate prevents binding of the inhibitor. Substrate and inhibitor compete for the enzyme.



Types of inhibition. This classification was introduced by W.W. Cleland.

Enzyme reaction rates can be decreased by various types of enzyme inhibitors.

Competitive inhibition

In competitive inhibition, the inhibitor and substrate compete for the enzyme (i.e., they can not bind at the same time). Often competitive inhibitors strongly resemble the real substrate of the enzyme. For example, methotrexate is a competitive inhibitor of the enzyme dihydrofolate reductase, which catalyzes the reduction of dihydrofolate to tetrahydrofolate. The similarity between the structures of folic acid and this drug are shown in the figure to the *right* bottom. Note that binding of the inhibitor need *not* be to the substrate binding site (as frequently stated), if binding of the inhibitor changes the

conformation of the enzyme to prevent substrate binding and *vice versa*. In competitive inhibition the maximal rate of the reaction is not changed, but higher substrate concentrations are required to reach a given maximum rate, increasing the apparent K_m.

Uncompetitive inhibition

In uncompetitive inhibition the inhibitor can not bind to the free enzyme, but only to the ES-complex. The EIS-complex thus formed is enzymatically inactive. This type of inhibition is rare, but may occur in multimeric enzymes.

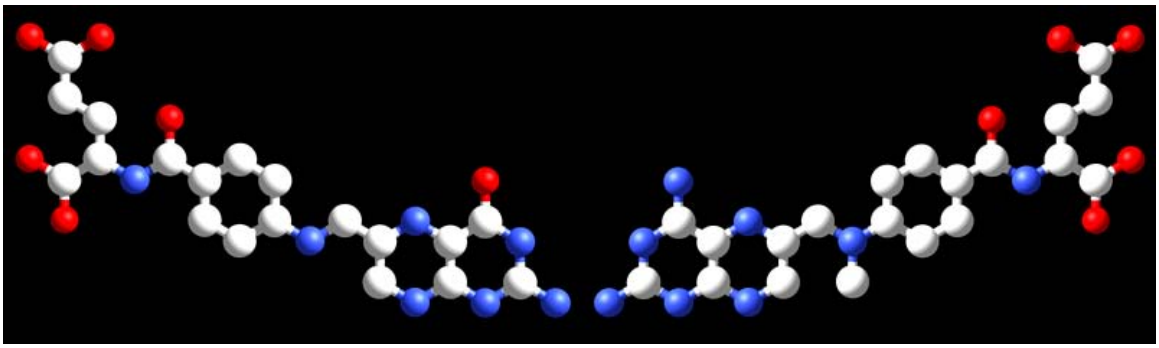
Non-competitive inhibition

Non-competitive inhibitors can bind to the enzyme at the binding site at the same time as the substrate,but not to the active site. Both the EI and EIS complexes are enzymatically inactive. Because the inhibitor can not be driven from the enzyme by higher substrate concentration (in contrast to competitive inhibition), the apparent V_{max} changes. But because the substrate can still bind to the enzyme, the K_m stays the same.

Mixed inhibition

This type of inhibition resembles the non-competitive, except that the EIS-complex has residual enzymatic activity.This type of inhibitor does not follow Michaelis-Menten equation.

In many organisms inhibitors may act as part of a feedback mechanism. If an enzyme produces too much of one substance in the organism, that substance may act as an inhibitor for the enzyme at the beginning of the pathway that produces it, causing production of the substance to slow down or stop when there is sufficient amount. This is a form of negative feedback. Enzymes which are subject to this form of regulation are often multimeric and have allosteric binding sites for regulatory substances. Their substrate/velocity plots are not hyperbolar, but sigmoidal (S-shaped).



The coenzyme folic acid (left) and the anti-cancer drug methotrexate (right) are very similar in structure. As a result, methotrexate is a competitive inhibitor of many enzymes that use folates.

Irreversible inhibitors react with the enzyme and form a covalent adduct with the protein. The inactivation is irreversible. These compounds include eflornithine a drug used to treat the parasitic disease sleeping sickness. Penicillin and Aspirin also act in this manner. With these drugs, the compound is bound in the active site and the enzyme then converts the inhibitor into an activated form that reacts irreversibly with one or more amino acid residues.

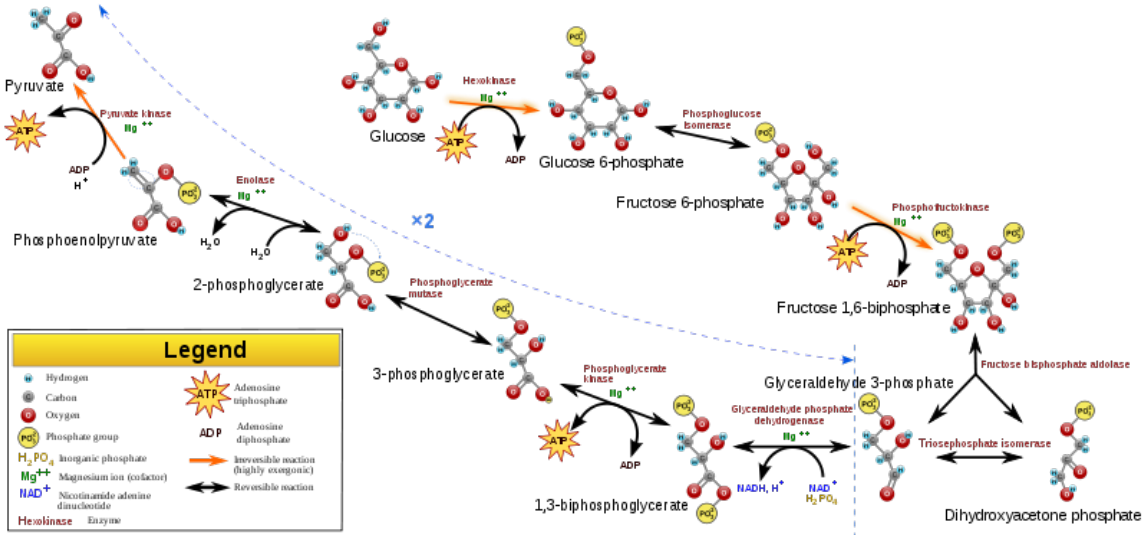
Uses of inhibitors

Since inhibitors modulate the function of enzymes they are often used as drugs. A common example of an inhibitor that is used as a drug is aspirin, which inhibits the COX-1 and COX-2 enzymes that produce the inflammation messenger prostaglandin, thus suppressing pain and inflammation. However, other enzyme inhibitors are poisons. For example, the poison cyanide is an irreversible enzyme inhibitor that combines with the copper and iron in the active site of the enzyme cytochrome c oxidase and blocks cellular respiration.

Biological function

Enzymes serve a wide variety of functions inside living organisms. They are indispensable for signal transduction and cell regulation, often via kinases and phosphatases. They also generate movement, with myosin hydrolysing ATP to generate muscle contraction and also moving cargo around the cell as part of the cytoskeleton. Other ATPases in the cell membrane are ion pumps involved in active transport. Enzymes are also involved in more exotic functions, such as luciferase generating light in fireflies. Viruses can also contain enzymes for infecting cells, such as the HIV integrase and reverse transcriptase, or for viral release from cells, like the influenza virus neuraminidase.

An important function of enzymes is in the digestive systems of animals. Enzymes such as amylases and proteases break down large molecules (starch or proteins, respectively) into smaller ones, so they can be absorbed by the intestines. Starch molecules, for example, are too large to be absorbed from the intestine, but enzymes hydrolyse the starch chains into smaller molecules such as maltose and eventually glucose, which can then be absorbed. Different enzymes digest different food substances. In ruminants which have herbivorous diets, microorganisms in the gut produce another enzyme, cellulase to break down the cellulose cell walls of plant fiber.



Glycolytic enzymes and their functions in the metabolic pathway of glycolysis

Several enzymes can work together in a specific order, creating metabolic pathways. In a metabolic pathway, one enzyme takes the product of another enzyme as a substrate. After the catalytic reaction, the product is then passed on to another enzyme. Sometimes more than one enzyme can catalyze the same reaction in parallel, this can allow more complex regulation: with for example a low constant activity being provided by one enzyme but an inducible high activity from a second enzyme.

Enzymes determine what steps occur in these pathways. Without enzymes, metabolism would neither progress through the same steps, nor be fast enough to serve the needs of the cell. Indeed, a metabolic pathway such as glycolysis could not exist independently of enzymes. Glucose, for example, can react directly with ATP to become phosphorylated at one or more of its carbons. In the absence of enzymes, this occurs so slowly as to be insignificant. However, if hexokinase is added, these slow reactions continue to take place except that phosphorylation at carbon 6 occurs so rapidly that if the mixture is tested a short time later, glucose-6-phosphate is found to be the only significant product. Consequently, the network of metabolic pathways within each cell depends on the set of functional enzymes that are present.

Control of activity

There are five main ways that enzyme activity is controlled in the cell.

1. **Enzyme production** (transcription and translation of enzyme genes) can be enhanced or diminished by a cell in response to changes in the cell's environment. This form of gene regulation is called enzyme induction and inhibition. For example, bacteria may become resistant to antibiotics such as penicillin because enzymes called beta-lactamases are induced that hydrolyse the crucial beta-lactam ring within the penicillin molecule. Another example are enzymes in the liver

called cytochrome P450 oxidases, which are important in drug metabolism. Induction or inhibition of these enzymes can cause drug interactions.

2. Enzymes can be **compartmentalized**, with different metabolic pathways occurring in different cellular compartments. For example, fatty acids are synthesized by one set of enzymes in the cytosol, endoplasmic reticulum and the Golgi apparatus and used by a different set of enzymes as a source of energy in the mitochondrion, through β-oxidation.
3. Enzymes can be regulated by **inhibitors and activators**. For example, the end product(s) of a metabolic pathway are often inhibitors for one of the first enzymes of the pathway (usually the first irreversible step, called *committed step*), thus regulating the amount of end product made by the pathways. Such a regulatory mechanism is called a negative feedback mechanism, because the amount of the end product produced is regulated by its own concentration. Negative feedback mechanism can effectively adjust the rate of synthesis of intermediate metabolites according to the demands of the cells. This helps allocate materials and energy economically, and prevents the manufacture of excess end products. The control of enzymatic action helps to maintain a stable internal environment in living organisms.
4. Enzymes can be regulated through **post-translational modification**. This can include phosphorylation, myristoylation and glycosylation. For example, in the response to insulin, the phosphorylation of multiple enzymes, including glycogen synthase, helps control the synthesis or degradation of glycogen and allows the cell to respond to changes in blood sugar. Another example of post-translational modification is the cleavage of the polypeptide chain. Chymotrypsin, a digestive protease, is produced in inactive form as chymotrypsinogen in the pancreas and transported in this form to the stomach where it is activated. This stops the enzyme from digesting the pancreas or other tissues before it enters the gut. This type of inactive precursor to an enzyme is known as a zymogen.
5. Some enzymes may become **activated when localized to a different environment** (e.g. from a reducing (cytoplasm) to an oxidizing (periplasm) environment, high pH to low pH etc.). For example, hemagglutinin in the influenza virus is activated by a conformational change caused by the acidic conditions, these occur when it is taken up inside its host cell and enters the lysosome.

Involvement in disease

Phenylalanine hydroxylase. Created from PDB 1KW0

Since the tight control of enzyme activity is essential for homeostasis, any malfunction (mutation, overproduction, underproduction or deletion) of a single critical enzyme can lead to a genetic disease. The importance of enzymes is shown by the fact that a lethal illness can be caused by the malfunction of just one type of enzyme out of the thousands of types present in our bodies.

One example is the most common type of phenylketonuria. A mutation of a single amino acid in the enzyme phenylalanine hydroxylase, which catalyzes the first step in the

degradation of phenylalanine, results in build-up of phenylalanine and related products. This can lead to mental retardation if the disease is untreated.

Another example is when germline mutations in genes coding for DNA repair enzymes cause hereditary cancer syndromes such as xeroderma pigmentosum. Defects in these enzymes cause cancer since the body is less able to repair mutations in the genome. This causes a slow accumulation of mutations and results in the development of many types of cancer in the sufferer.

Naming conventions

An enzyme's name is often derived from its substrate or the chemical reaction it catalyzes, with the word ending in ***-ase***. Examples are lactase, alcohol dehydrogenase and DNA polymerase. This may result in different enzymes, called isozymes, with the same function having the same basic name. Isoenzymes have a different amino acid sequence and might be distinguished by their optimal pH, kinetic properties or immunologically. Isoenzyme and isozyme are homologous proteins. Furthermore, the normal physiological reaction an enzyme catalyzes may not be the same as under artificial conditions. This can result in the same enzyme being identified with two different names. *E.g.* Glucose isomerase, used industrially to convert glucose into the sweetener fructose, is a xylose isomerase *in vivo*.

The International Union of Biochemistry and Molecular Biology have developed a nomenclature for enzymes, the **EC numbers**; each enzyme is described by a sequence of four numbers preceded by "EC". The first number broadly classifies the enzyme based on its mechanism.

The top-level classification is

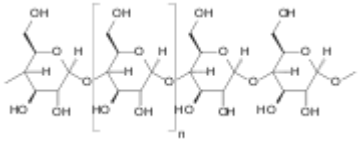
- EC 1 *Oxidoreductases*: catalyze oxidation/reduction reactions
- EC 2 *Transferases*: transfer a functional group (*e.g.* a methyl or phosphate group)
- EC 3 *Hydrolases*: catalyze the hydrolysis of various bonds
- EC 4 *Lyases*: cleave various bonds by means other than hydrolysis and oxidation
- EC 5 *Isomerases*: catalyze isomerization changes within a single molecule
- EC 6 *Ligases*: join two molecules with covalent bonds.

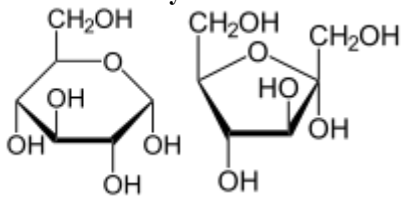
According to the naming conventions, enzymes are generally classified into six main family classes and many sub-family classes. Some web-servers, e.g., EzyPred and bioinformatics tools have been developed to predict which main family class and sub-family class an enzyme molecule belongs to according to its sequence information alone via the pseudo amino acid composition.

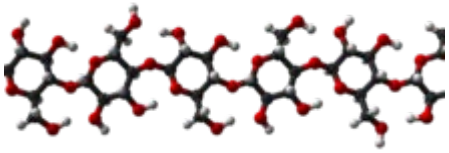
Industrial applications

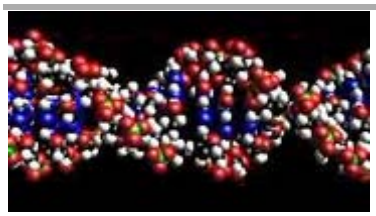
Enzymes are used in the chemical industry and other industrial applications when extremely specific catalysts are required. However, enzymes in general are limited in the

number of reactions they have evolved to catalyze and also by their lack of stability in organic solvents and at high temperatures. Consequently, protein engineering is an active area of research and involves attempts to create new enzymes with novel properties, either through rational design or *in vitro* evolution. These efforts have begun to be successful, and a few enzymes have now been desiged "from scratch" to catalyze reactions that do not occur in nature.

Application	Enzymes used	Uses
Food processing  Amylases catalyze the release of simple sugars from starch.	Amylases from fungi and plants	Production of sugars from starch, such as in making high-fructose corn syrup. In baking, catalyze breakdown of starch in the flour to sugar. Yeast fermentation of sugar produces the carbon dioxide that raises the dough.
	Proteases	Biscuit manufacturers use them to lower the protein level of flour.
Baby foods	Trypsin	To predigest baby foods
Brewing industry  Germinating barley used for malt	Enzymes from barley are released during the mashing stage of beer production.	They degrade starch and proteins to produce simple sugar, amino acids and peptides that are used by yeast for fermentation.
	Industrially produced barley enzymes	Widely used in the brewing process to substitute for the natural enzymes found in barley.
	Amylase, glucanases, proteases	Split polysaccharides and proteins in the malt.
	Betaglucanases and arabinoxylanases	Improve the wort and beer filtration characteristics.
	Amyloglucosidase and pullulanases	Low-calorie beer and adjustment of fermentability.
	Proteases	Remove cloudiness produced during

		storage of beers.
	Acetolactatedecarboxylase (ALDC)	Increases fermentation efficiency by reducing diacetyl formation.
Fruit juices	Cellulases, pectinases	Clarify fruit juices.
Dairy industry  Roquefort cheese	Rennin, derived from the stomachs of young ruminant animals (like calves and lambs)	Manufacture of cheese, used to hydrolyze protein
	Microbially produced enzyme	Now finding increasing use in the dairy industry
	Lipases	Is implemented during the production of Roquefort cheese to enhance the ripening of the blue-mould cheese.
	Lactases	Break down lactose to glucose and galactose.
Meat tenderizers	Papain	To soften meat for cooking
Starch industry  Glucose Fructose	Amylases, amyloglucosideases and glucoamylases	Converts starch into glucose and various syrups.
	Glucose isomerase	Converts glucose into fructose in production of high fructose syrups from starchy materials. These syrups have enhanced sweetening properties and lower calorific values than sucrose for the same level of sweetness.
Paper industry 	Amylases, Xylanases, Cellulases and ligninases	Degrade starch to lower viscosity, aiding sizing and coating paper. Xylanases reduce bleach required for decolorising; cellulases smooth fibers, enhance

A paper mill in South Carolina		water drainage, and promote ink removal; lipases reduce pitch and lignin-degrading enzymes remove lignin to soften paper.
Biofuel industry  Cellulose in 3D	Cellulases	Used to break down cellulose into sugars that can be fermented
	Ligninases	Use of lignin waste
Biological detergent	Primarily proteases, produced in an extracellular form from bacteria	Used for presoak conditions and direct liquid applications helping with removal of protein stains from clothes
	Amylases	Detergents for machine dish washing to remove resistant starch residues
	Lipases	Used to assist in the removal of fatty and oily stains
	Cellulases	Used in biological fabric conditioners
Contact lens cleaners	Proteases	To remove proteins on contact lens to prevent infections
Rubber industry	Catalase	To generate oxygen from peroxide to convert latex into foam rubber
Photographic industry	Protease (ficin)	Dissolve gelatin off scrap film, allowing recovery of its silver content.
Molecular biology	Restriction enzymes, DNA ligase and polymerases	Used to manipulate DNA in genetic engineering, important in pharmacology, agriculture and medicine. Essential



Part of the DNA double helix

for restriction digestion and the polymerase chain reaction. Molecular biology is also important in forensic science.

Chapter 3

Heat Shock Protein

Heat shock proteins (**HSP**) are a class of functionally related proteins whose expression is increased when cells are exposed to elevated temperatures or other stress. This increase in expression is transcriptionally regulated. The dramatic upregulation of the heat shock proteins is a key part of the **heat shock response** and is induced primarily by heat shock factor (HSF). HSPs are found in virtually all living organisms, from bacteria to humans.

Heat-shock proteins are named according to their molecular weight. For example, Hsp60, Hsp70 and Hsp90 (the most widely-studied HSPs) refer to families of heat shock proteins on the order of 60, 70 and 90 kilodaltons in size, respectively. The small 8 kilodalton protein ubiquitin, which marks proteins for degradation, also has features of a heat shock protein.

Discovery

It is known that rapid heat hardening can be elicited by a brief exposure of cells to sub-lethal high temperature, which in turn provides protection from subsequent and more severe temperature. In 1962, Ritossa reported that heat and the metabolic uncoupler dinitrophenol induced a characteristic pattern of puffing in the chromosomes of Drosophila. This discovery eventually led to the identification of the heat-shock proteins (HSP) or stress proteins whose expression these puffs represented. Increased synthesis of selected proteins in Drosophila cells following stresses such as heat shock was first reported in 1974.

Beginning in the mid-1980s, investigators recognized that many HSPs function as molecular chaperones and thus play a critical role in protein folding, intracellular trafficking of proteins, and coping with proteins denatured by heat and other stresses. Accordingly, the study of stress proteins has undergone explosive growth.

Function

Upregulation in stress

Production of high levels of heat shock proteins can also be triggered by exposure to different kinds of environmental stress conditions, such as infection, inflammation, exercise, exposure of the cell to toxins (ethanol, arsenic, trace metals and ultraviolet light, among many others), starvation, hypoxia (oxygen deprivation), nitrogen deficiency (in plants), or water deprivation. Consequently, the heat shock proteins are also referred to as **stress proteins** and their upregulation is sometimes described more generally as part of the **stress response**.

The mechanism by which heat-shock (or other environmental stressors) activates the heat shock factor has not been determined. However, some studies suggest that an increase in damaged or abnormal proteins brings HSPs into action.

Some bacterial heat shock proteins are upregulated via a mechanism involving RNA thermometers such as the FourU thermometer, ROSE element and the Hsp90 cis-regulatory element.

Role as chaperone

Heat shock proteins function as intra-cellular chaperones for other proteins. They play an important role in protein-protein interactions such as folding and assisting in the establishment of proper protein conformation (shape) and prevention of unwanted protein aggregation. By helping to stabilize partially unfolded proteins, HSPs aid in transporting proteins across membranes within the cell.

Some members of the HSP family are expressed at low to moderate levels in *all* organisms because of their essential role in protein maintenance.

Housekeeping

Heat-shock proteins also occur under non-stressful conditions, simply "monitoring" the cell's proteins. Some examples of their role as "monitors" are that they carry old proteins to the cell's "recycling bin" (proteasome) and they help newly synthesised proteins fold properly.

These activities are part of a cell's own repair system, called the "cellular stress response" or the "heat-shock response".

Cardiovascular

Heat shock proteins appear to serve a significant cardiovascular role. Hsp90, hsp84, hsp70, hsp27, hsp20, and alpha B crystallin all have been reported as having roles in the cardiovasculature.

Hsp90 binds both endothelial nitric oxide synthase and soluble guanylate cyclase which in turn are involved in vascular relaxation.

A downstream kinase of the nitric oxide cell signalling pathway, protein kinase G, phosphorylates a small heat shock protein, hsp20. Hsp20 phosphorylation correlates well with smooth muscle relaxation and is one significant phosphoprotein involved in the process. Hsp20 appears significant in development of the smooth muscle phenotype during development. Hsp20 also serves a significant role in preventing platelet aggregation, cardiac myocyte function and prevention of apoptosis after ischemic injury, and skeletal muscle function and muscle insulin response.

Hsp27 is a major phosphoprotein during muscle contraction. Hsp27 functions in smooth muscle migration and appears to serve an integral role.

Immunity

Extracellular and membrane bound heat-shock proteins, especially Hsp70 are involved in binding antigens and presenting them to the immune system.

Clinical significance

Heat Shock Factor 1 (HSF1) is a transcription factor that is involved in the upregulation of Hsp70 protein expression. Recently it was discovered that HSF1 is a powerful multifaceted modifier of carcinogenesis. HSF1 knockout mice show significantly decreased incidence of skin tumor after topical application of DMBA (7,12-**d**i**m**ethyl**b**enz**a**nthracene), a mutagen.

Applications

Cancer vaccine adjuvant

Given their role in antigen presentation, HSPs are useful as immunologic adjuvants in boosting the response to a vaccine. Furthermore, some researchers speculate that HSPs may be involved in binding protein fragments from dead malignant cells and presenting them to the immune system. Therefore HSPs may be useful for increasing the effectiveness of cancer vaccines.

Anticancer therapeutics

Intracellular heat shock proteins are highly expressed in cancerous cells and are essential to the survival of these cell types. Hence small molecule inhibitors of Hsps, especially Hsp90 show promise as anticancer agents. The potent Hsp90 inhibitor 17-AAG is currently in clinical trials for the treatment of several types of cancer.

Agricultural

Researchers are also investigating the role of HSPs in conferring stress tolerance to hybridized plants, hoping to address drought and poor soil conditions for farming.

Classification

The principal heat-shock proteins that have chaperone activity belong to five conserved classes: HSP33, HSP60, HSP70, HSP90, HSP100, and the small heat-shock proteins (sHSPs).

Approximate molecular weight (kDa)	Prokaryotic proteins	Eukaryotic proteins	Function
10 kDa	GroES	Hsp10	
20-30 kDa	GrpE	The HspB group of Hsp. Eleven members in mammals including Hsp27 or HspB1	
40 kDa	DnaJ	Hsp40	Co-factor of Hsp70
60 kDa	GroEL, 60kDa antigen	Hsp60	Involved in protein folding after its post-translational import to the mitochondrion/chloroplast
70 kDa	DnaK	The HspA group of Hsp including Hsp71, Hsp70, Hsp72, Grp78 (BiP), Hsx70 found only in primates	Protein folding and unfolding, provides thermotolerance to cell on exposure to heat stress. Also prevents protein folding during post-translational import into the mitochondria/chloroplast.
90 kDa	HtpG, C62.5	The HspC group of Hsp including Hsp90, Grp94	Maintenance of steroid receptors and transcription factors
100 kDa	ClpB, ClpA, ClpX	Hsp104, Hsp110	Tolerance of extreme temperature

Although the most important members of each family are tabulated here, it should be noted that some species may express additional chaperones, co-chaperones, and heat shock proteins not listed. Additionally, many of these proteins may have multiple splice variants (Hsp90α and Hsp90β, for instance) or conflicts of nomenclature (Hsp72 is sometimes called Hsp70).

Chapter 4

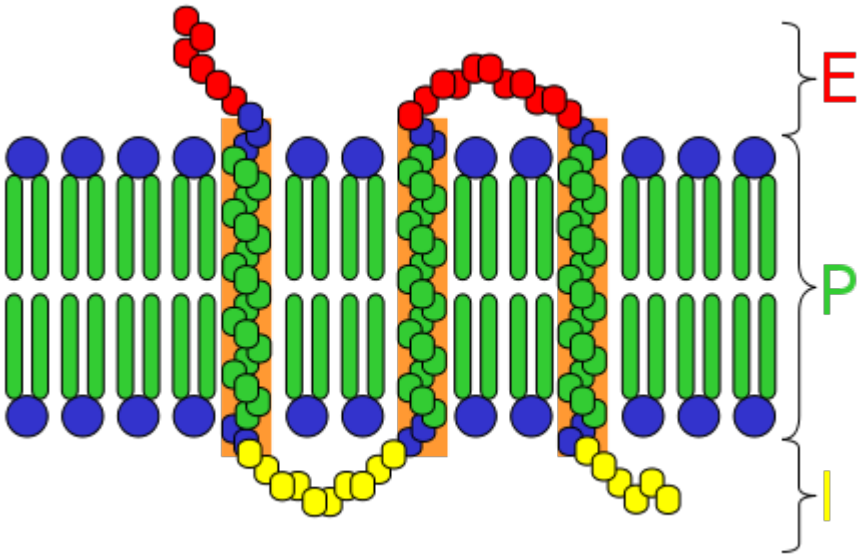
Receptor (Biochemistry)

In biochemistry, a **receptor** is a protein molecule, embedded in either the plasma membrane or the cytoplasm of a cell, to which one or more specific kinds of signaling molecules may attach. A molecule which binds (attaches) to a receptor is called a ligand, and may be a peptide (short protein) or other small molecule, such as a neurotransmitter, a hormone, a pharmaceutical drug, or a toxin. Each kind of receptor can bind only certain ligand shapes. Each cell typically has many receptors, of many different kinds. Simply put, a receptor functions as a keyhole that opens a neural path when the proper ligand is inserted.

Ligand binding stabilizes a certain receptor conformation (the three-dimensional shape of the receptor protein, with no change in sequence). This is often associated with gain of or loss of protein activity, ordinarily leading to some sort of cellular response. However, some ligands (e.g. antagonists) merely block receptors without inducing any response. Ligand-induced changes in receptors result in cellular changes which constitute the biological activity of the ligands. Many functions of the human body are regulated by these receptors responding uniquely to specific molecules like this.

Overview

The shapes and actions of receptors are studied by X-ray crystallography, dual polarisation interferometry, computer modelling, and structure-function studies, which have advanced the understanding of drug action at the binding sites of receptors. Structure activity relationships correlate induced conformational changes with biomolecular activity, and are studied using dynamic techniques such as circular dichroism and dual polarisation interferometry.



Transmembrane receptor:E=extracellular space; I=intracellular space; P=plasma membrane

Depending on their functions and ligands, several types of receptors may be identified:

- Some receptor proteins are peripheral membrane proteins.
- Many hormone and neurotransmitter receptors are transmembrane proteins: transmembrane receptors are embedded in the phospholipid bilayer of cell membranes, that allow the activation of signal transduction pathways in response to the activation by the binding molecule, or ligand.
    - Metabotropic receptors are coupled to G proteins and affect the cell indirectly through enzymes which control ion channels.
    - Ionotropic receptors (also known as ligand-gated ion channels) contain a central pore which opens in response to the binding of ligand.
- Another major class of receptors are intracellular proteins such as those for steroid and intracrine peptide hormone receptors. These receptors often can enter the cell nucleus and modulate gene expression in response to the activation by the ligand.

Membrane receptors are isolated from cell membranes by complex extraction procedures using solvents, detergents, and/or affinity purification.

Binding and activation

Ligand binding is an equilibrium process. Ligands bind to receptors and dissociate from them according to the law of mass action.

$$[\text{Ligand}] \cdot [\text{Receptor}] \overset{K_d}{\rightleftharpoons} [\text{Ligand-receptor complex}]$$

(the brackets stand for concentrations)

One measure of how well a molecule fits a receptor is the binding affinity, which is inversely related to the dissociation constant K_d. A good fit corresponds with high affinity and low K_d. The final biological response (e.g. second messenger cascade, muscle contraction), is only achieved after a significant number of receptors are activated.

The receptor-ligand affinity is greater than enzyme-substrate affinity. Whilst both interactions are specific and reversible, there is no chemical modification of the ligand as seen with the substrate upon binding to its enzyme.

If the receptor exists in two states, then the ligand binding must account for these two receptor states.

Constitutive activity

A receptor which is capable of producing its biological response in the absence of a bound ligand is said to display "constitutive activity". The constitutive activity of receptors may be blocked by inverse agonist binding. Mutations in receptors that result in increased constitutive activity underlie some inherited diseases, such as precocious puberty (due to mutations in luteinizing hormone receptors) and hyperthyroidism (due to mutations in thyroid-stimulating hormone receptors).

Theories of drug receptor interaction

Occupation theory

Drug effect is directly proportional to number of receptors occupied Drug effect ceases as drug-receptor complex dissociate

Ariens & Stephenson theory

introduced Terms of "affinity" & "efficacy" Affinity: ability of the drug to combine with receptor to create drug-receptor complex Efficacy: ability of the drug-receptor complex to initiate a response

Affinity “drug-receptor interaction” is governed by the law of mass action.

In this theory

Agonist: drug with high affinity & high intrinsic activity Partial agonist: drug with high affinity & low intrinsic activity Antagonist: drug with high affinity & low intrinsic activity

Rate theory

The activation of receptors is directly proportional to the total number of encounters of the drug with its receptors per unit time Pharmacological activity is directly proportional to the rate of dissociation & association not number of receptors occupied

Agonist:drug with fast association & fast dissociation Partial agonist:drug with intermediate association & intermediate dissociation Antagonist:drug with fast association & slow dissociation

Induced fit theory

As the drug approaches the receptor the receptors alters the conformation of its binding site to produce drug—receptor complex

Agonists versus antagonists

Not every ligand that binds to a receptor also activates the receptor. The following classes of ligands exist:

- *(Full) agonists* are able to activate the receptor and result in a maximal biological response. Most natural ligands are full agonists.
- *Partial agonists* do not activate receptors thoroughly, causing responses which are partial compared to those of full agonists.
- *Antagonists* bind to receptors but do not activate them. This results in receptor blockage, inhibiting the binding of other agonists.
- *Inverse agonists* reduce the activity of receptors by inhibiting their constitutive activity.

Peripheral membrane protein receptors

These receptors are relatively rare compared to the much more common types of receptors that cross the cell membrane. An example of a receptor that is a peripheral membrane protein is the elastin receptor.

Transmembrane receptors

G protein-coupled receptors

These receptors are also known as **seven transmembrane receptors** or **7TM** receptors, because they pass through the membrane seven times.

- Muscarinic acetylcholine receptor (Acetylcholine and Muscarine)
- Adenosine receptors (Adenosine)
- Adrenoceptors (also known as Adrenergic receptors, for *adren*aline, and other structurally related hormones and drugs)

- GABA receptors, Type-B (γ-Aminobutyric acid or GABA)
- Angiotensin receptors (Angiotensin)
- Cannabinoid receptors (Cannabinoids)
- Cholecystokinin receptors (Cholecystokinin)
- Dopamine receptors (Dopamine)
- Glucagon receptors (Glucagon)
- Metabotropic glutamate receptors (Glutamate)
- Histamine receptors (Histamine)
- Olfactory receptors (for the sense of smell)
- Opioid receptors (Opioids)
- Protease-activated receptors
- Rhodopsin (a photoreceptor protein)
- Secretin receptors (Secretin)
- Serotonin receptors, except Type-3 (Serotonin, also known as 5-Hydroxytryptamine or 5-HT)
- Somatostatin receptors (Somatostatin)
- Calcium-sensing receptor (Calcium)
- Chemokine receptors (Chemokines)
- *many more ...*

Receptor tyrosine kinases

These receptors detect ligands and propagate signals via the tyrosine kinase of their intracellular domains. This family of receptors includes;

- Erythropoietin receptor (Erythropoietin)
- Insulin receptor (Insulin)
- Eph receptors
- Insulin-like growth factor 1 receptor
- various other growth factor and cytokine receptors
- ....

Guanylyl cyclase receptors

- GC-A & GC-B: receptors for Atrial-natriuretic peptide (ANP) and other natriuretic peptides
- GC-C: Guanylin receptor

Ionotropic receptors

Ionotropic receptors are heteromeric or homomeric oligomers . They are receptors that respond to extracellular ligands and receptors that respond to intracellular ligands.

Extracellular ligands

Receptor	Ligand	Ion current
Nicotinic acetylcholine receptor	Acetylcholine, Nicotine	Na^+, K^+, Ca^{2+}
Glycine receptor (GlyR)	Glycine, Strychnine	Cl^- > HCO^-_3
GABA receptors: GABA-A, GABA-C	GABA	Cl^- > HCO^-_3
Glutamate receptors: NMDA receptor, AMPA receptor, and Kainate receptor	Glutamate	Na^+, K^+, Ca^{2+}
5-HT_3 receptor	Serotonin	Na^+, K^+
P2X receptors	ATP	Ca^{2+}, Na^+, Mg^{2+}

Intracellular ligands

Receptor	Ligand	Ion current
cyclic nucleotide-gated ion channels	cGMP (vision), cAMP and cGTP (olfaction)	Na^+, K^+
IP_3 receptor	IP_3	Ca^{2+}
Intracellular ATP receptors	ATP (closes channel)	K^+
Ryanodine receptor	Ca^{2+}	Ca^{2+}

The entire repertoire of human plasma membrane receptors is listed at the Human Plasma Membrane Receptome**.**

Intracellular receptors

Transcription factors

- nuclear receptor:
  - Steroid hormone receptor

Various

- Ionotropic receptors (IP_3 receptor above)
- sigma1 (neurosteroids)
- G protein-coupled receptors

Role in genetic disorders

Many genetic disorders involve hereditary defects in receptor genes. Often, it is hard to determine whether the receptor is nonfunctional or the hormone is produced at decreased level; this gives rise to the "pseudo-hypo-" group of endocrine disorders, where there appears to be a decreased hormonal level while in fact it is the receptor that is not responding sufficiently to the hormone.

Receptor regulation

Cells can increase (upregulate) or decrease (downregulate) the number of receptors to a given hormone or neurotransmitter to alter its sensitivity to this molecule. This is a locally acting feedback mechanism.

Receptor desensitization

Ligand-bound desensitation Vol. 135. No. 5 2130–2136</ref>

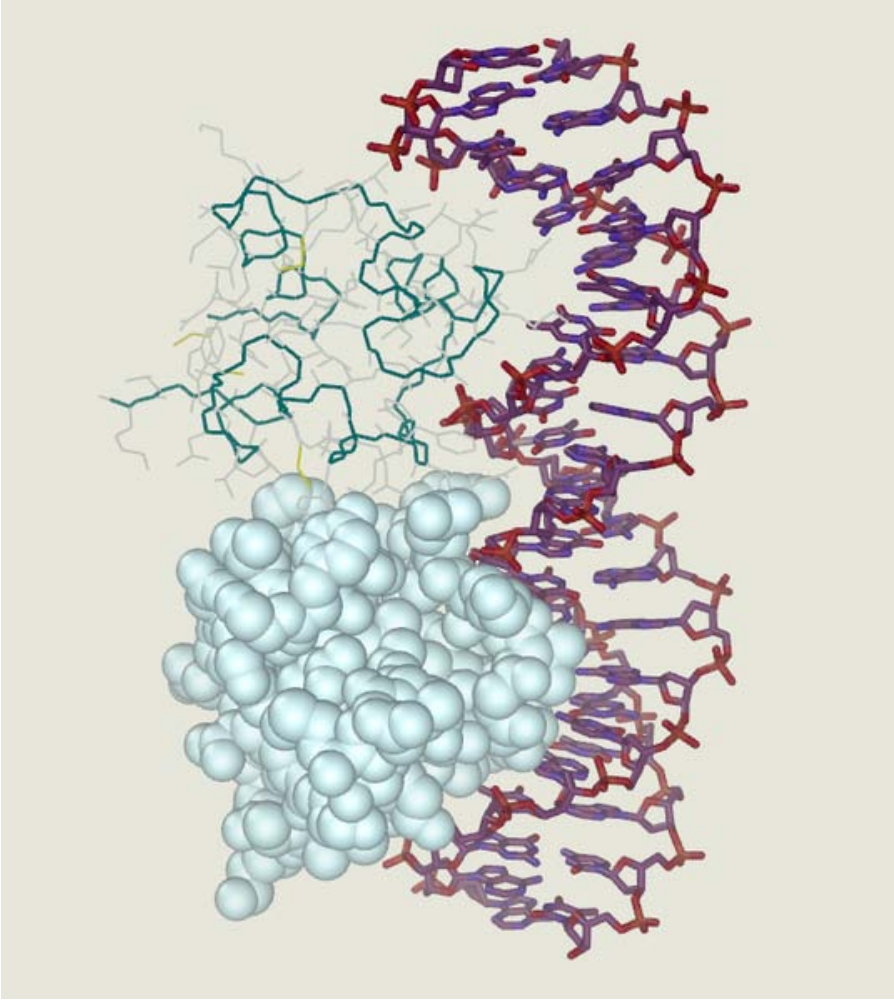
- Uncoupling of receptor effector molecules.
- Receptor sequestration (internalization).

In immune system

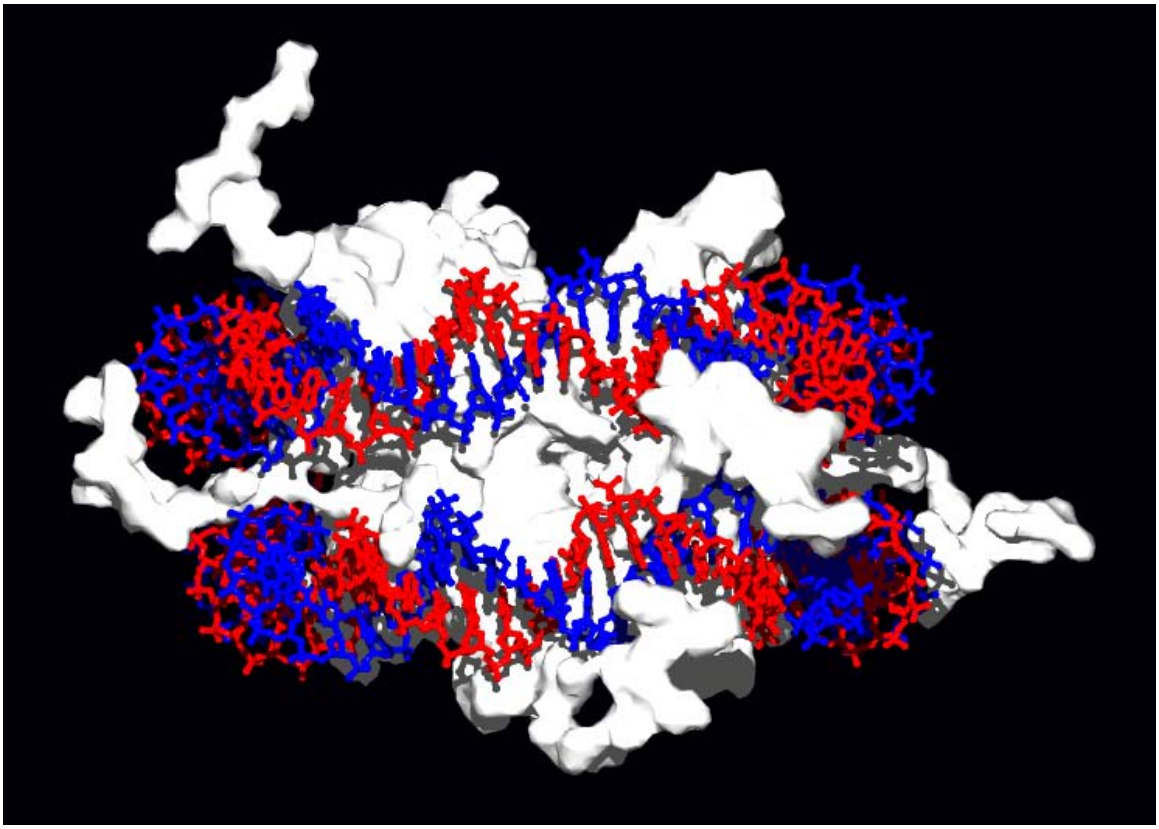
The main receptors in the immune system are pattern recognition receptors (PRRs), toll-like receptors (TLRs), killer activated and killer inhibitor receptors (KARs and KIRs), complement receptors, Fc receptors, B cell receptors and T cell receptors.

Chapter 5

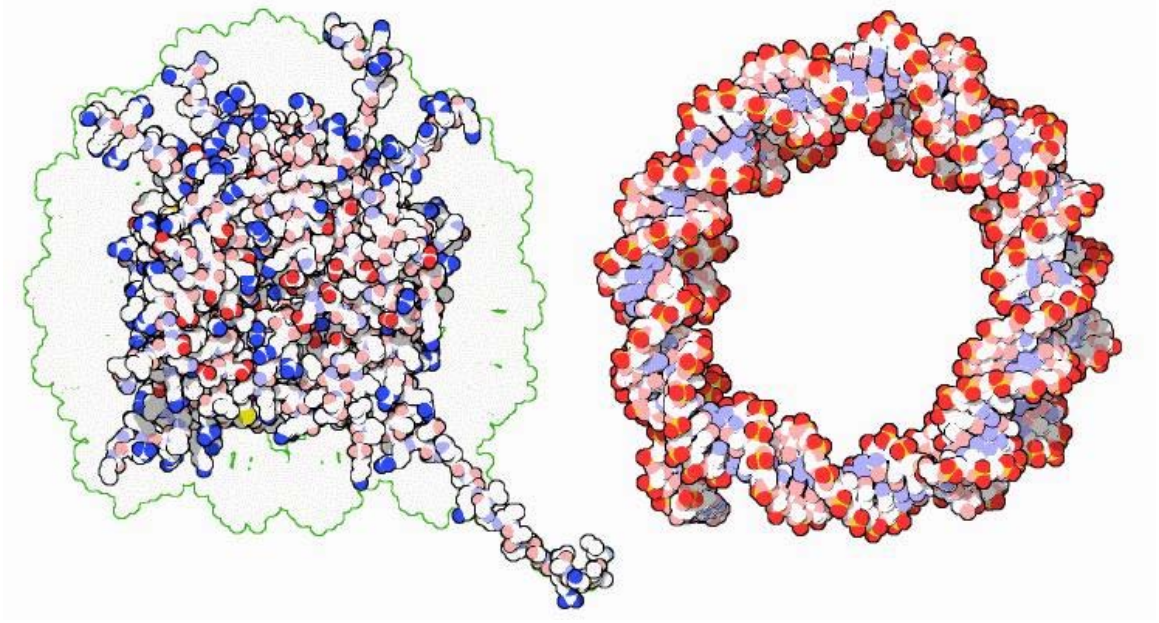
DNA-Binding Protein



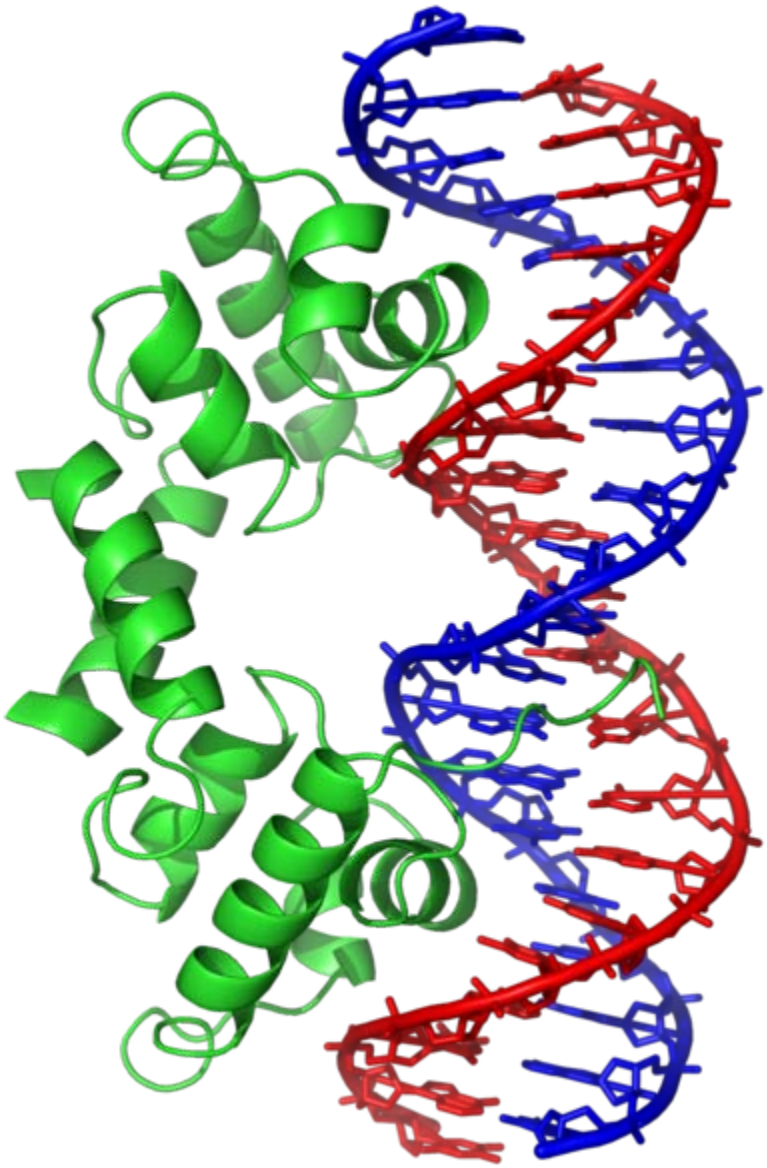
Cro protein complex with DNA



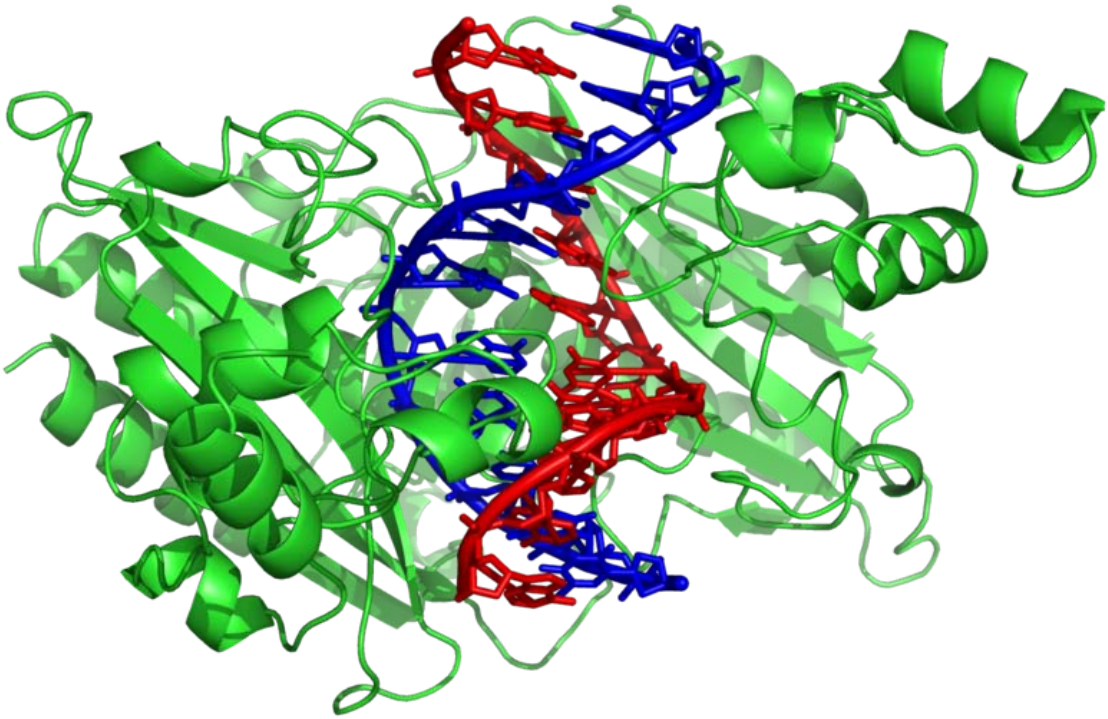
Interaction of DNA with histones (shown in white, top). These proteins' basic amino acids (below left, blue) bind to the acidic phosphate groups on DNA (below right, red).



Nucleosome (opposites attracts) (see above for color scheme)



The lambda repressor helix-turn-helix transcription factor bound to its DNA target



The restriction enzyme EcoRV (green) in a complex with its substrate DNA

DNA-binding proteins are proteins that are composed of DNA-binding domains and thus have a specific or general affinity for either single or double stranded DNA. Sequence-specific DNA-binding proteins generally interact with the major groove of B-DNA, because it exposes more functional groups that identify a base pair. However there are some known narrow-groove DNA-binding ligands such as Netropsin, Distamycin, Hoechst 33258, Pentamidine and others.

Examples

DNA-binding proteins include transcription factors which modulate the process of transcription, various polymerases, nucleases which cleave DNA molecules, and histones which are involved in chromosome packaging in the cell nucleus. DNA-binding proteins can incorporate such domains as the zinc finger, the helix-turn-helix, and the leucine zipper (among many others) that facilitate binding to nucleic acid.

Non-specific DNA-protein interactions

Structural proteins that bind DNA are well-understood examples of non-specific DNA-protein interactions. Within chromosomes, DNA is held in complexes with structural proteins. These proteins organize the DNA into a compact structure called chromatin. In eukaryotes this structure involves DNA binding to a complex of small basic proteins

called histones, while in prokaryotes multiple types of proteins are involved. The histones form a disk-shaped complex called a nucleosome, which contains two complete turns of double-stranded DNA wrapped around its surface. These non-specific interactions are formed through basic residues in the histones making ionic bonds to the acidic sugar-phosphate backbone of the DNA, and are therefore largely independent of the base sequence. Chemical modifications of these basic amino acid residues include methylation, phosphorylation and acetylation. These chemical changes alter the strength of the interaction between the DNA and the histones, making the DNA more or less accessible to transcription factors and changing the rate of transcription. Other non-specific DNA-binding proteins in chromatin include the high-mobility group proteins, which bind to bent or distorted DNA. These proteins are important in bending arrays of nucleosomes and arranging them into the larger structures that make up chromosomes.

DNA-binding proteins that specifically bind single-stranded DNA

A distinct group of DNA-binding proteins are the DNA-binding proteins that specifically bind single-stranded DNA. In humans, replication protein A is the best-understood member of this family and is used in processes where the double helix is separated, including DNA replication, recombination and DNA repair. These binding proteins seem to stabilize single-stranded DNA and protect it from forming stem-loops or being degraded by nucleases.

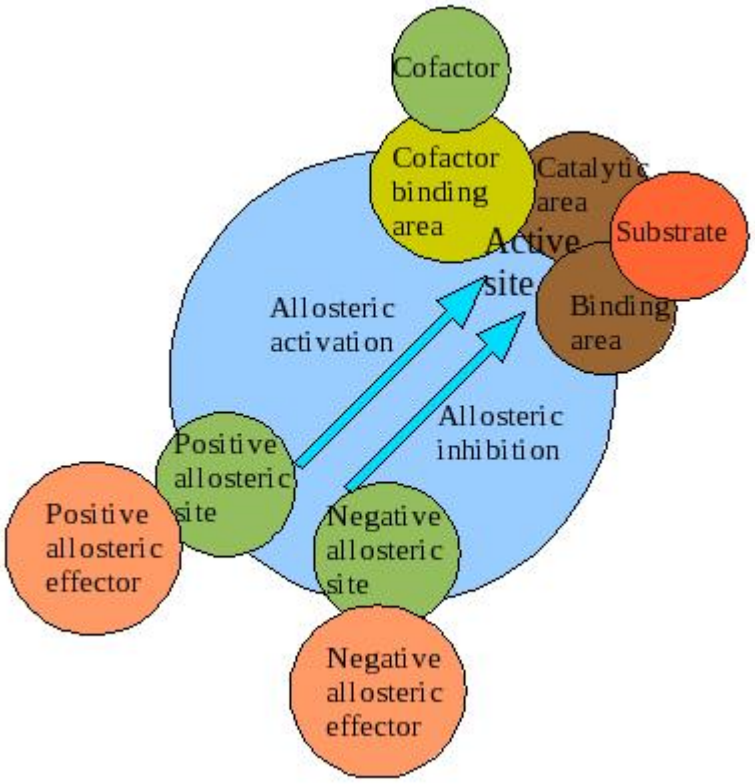
Binding to particular DNA sequences

In contrast, other proteins have evolved to bind to particular DNA sequences. The most intensively studied of these are the various transcription factors, which are proteins that regulate transcription. Each transcription factor binds to one particular set of DNA sequences and activates or inhibits the transcription of genes that have these sequences close to their promoters. The transcription factors do this in two ways. Firstly, they can bind the RNA polymerase responsible for transcription, either directly or through other mediator proteins; this locates the polymerase at the promoter and allows it to begin transcription. Alternatively, transcription factors can bind enzymes that modify the histones at the promoter; this will change the accessibility of the DNA template to the polymerase.

As these DNA targets can occur throughout an organism's genome, changes in the activity of one type of transcription factor can affect thousands of genes. Consequently, these proteins are often the targets of the signal transduction processes that control responses to environmental changes or cellular differentiation and development. The specificity of these transcription factors' interactions with DNA come from the proteins making multiple contacts to the edges of the DNA bases, allowing them to "read" the DNA sequence. Most of these base-interactions are made in the major groove, where the bases are most accessible. Mathematical descriptions of protein-DNA binding taking into account sequence-specificity, competitive and cooperative binding of proteins of different types are usually performed with the help of the lattice models.

Chapter 6

Allosteric Regulation



Allosteric regulation of an enzyme

In biochemistry, **allosteric regulation** is the regulation of an enzyme or other protein by binding an effector molecule at the protein's allosteric site (that is, a site other than the protein's active site). Effectors that enhance the protein's activity are referred to as *allosteric activators*, whereas those that decrease the protein's activity are called *allosteric inhibitors*. The term *allostery* comes from the Greek *allos* (ἄλλος), "other", and *stereos* (στερεός), "solid (object)", in reference to the fact that the regulatory site of an allosteric protein is physically distinct from its active site. Allosteric regulations are a

natural example of control loops, such as feedback from downstream products or feedforward from upstream substrates.

Models of allosteric regulation

Most allosteric effects can be explained by the *concerted* MWC model put forth by Monod, Wyman, and Changeux, or by the sequential model described by Koshland, Nemethy, and Filmer. Both postulate that enzyme subunits exist in one of two conformations, tensed (T) or relaxed (R), and that relaxed subunits bind substrate more readily than those in the tense state. The two models differ most in their assumptions about subunit interaction and the preexistence of both states.

Concerted model

The concerted model of allostery, also referred to as the symmetry model or MWC model, postulates that enzyme subunits are connected in such a way that a conformational change in one subunit is necessarily conferred to all other subunits. Thus, all subunits must exist in the same conformation. The model further holds that, in the absence of any ligand (substrate or otherwise), the equilibrium favours one of the conformational states, T or R. The equilibrium can be shifted to the R or T state through the binding of one ligand (the allosteric effector or ligand) to a site that is different from the active site (the allosteric site).

Sequential model

The sequential model of allosteric regulation holds that subunits are not connected in such a way that a conformational change in one induces a similar change in the others. Thus, all enzyme subunits do not necessitate the same conformation. Moreover, the sequential model dictates that molecules of substrate bind via an induced fit protocol. In general, when a subunit randomly collides with a molecule of substrate, the active site, in essence, forms a glove around its substrate. While such an induced fit converts a subunit from the tensed state to relaxed state, it does not propagate the conformational change to adjacent subunits. Instead, substrate-binding at one subunit only slightly alters the structure of other subunits so that their binding sites are more receptive to substrate. To summarize:

- subunits need not exist in the same conformation
- molecules of substrate bind via induced-fit protocol
- conformational changes are not propagated to all subunits
- substrate-binding causes increased substrate affinity in adjacent subunits.

Allosteric modulation

Positive modulation

Positive allosteric modulation (also known as *allosteric activation*) occurs when the binding of one ligand enhances the attraction between substrate molecules and other binding sites. An example is the binding of oxygen molecules to hemoglobin, where oxygen is effectively both the substrate and the effector. The allosteric, or "other", site is the active site of an adjoining protein subunit. The binding of oxygen to one subunit induces a conformational change in that subunit that interacts with the remaining active sites to enhance *their* oxygen affinity.

Negative modulation

Negative allosteric modulation (also known as *allosteric inhibition*) occurs when the binding of one ligand decreases the affinity for substrate at other active sites. For example, when 2,3-BPG binds to an allosteric site on hemoglobin, the affinity for oxygen of all subunits decreases.

Another example is strychnine, a convulsant poison, which acts as an allosteric inhibitor of the glycine receptor. Glycine is a major post-synaptic inhibitory neurotransmitter in mammalian spinal cord and brain stem. Strychnine acts at a separate binding site on the glycine receptor in an allosteric manner; i.e., its binding lowers the affinity of the glycine receptor for glycine. Thus, strychnine inhibits the action of an inhibitory transmitter, leading to convulsions.

Types of allosteric regulation

Homotropic

A homotropic allosteric modulator is a substrate for its target enzyme, as well as a regulatory molecule of the enzyme's activity. It is typically an activator of the enzyme.

Heterotropic

A heterotropic allosteric modulator is a regulatory molecule that is not also the enzyme's substrate. It may be either an activator or an inhibitor of the enzyme.

Some allosteric proteins can be regulated by both their substrates and other molecules. Such proteins are capable of both homotropic and heterotropic interactions.

Non-Regulatory Allostery

A non-regulatory allosteric site refers to any non-regulatory component of an enzyme (or any protein) that is not itself an amino acid. For instance, many enzymes require sodium

binding to ensure proper function. However, the sodium does not necessarily act as a regulatory subunit; the sodium is always present and there are no known biological processes to add/remove sodium to regulate enzyme activity. Non-regulatory allostery could comprise any other ions besides sodium (calcium, magnesium, zinc), as well as other chemicals and possibly vitamins.

Pharmacology

Allosteric modulation of a receptor results from the binding of allosteric modulators at a different site (a "regulatory site") from that of the endogenous ligand (an "active site") and enhances or inhibits the effects of the endogenous ligand. Under normal circumstances, it acts by causing a conformational change in a receptor molecule, which results in a change in the binding affinity of the ligand. In this way, an allosteric ligand modulates the receptor's activation by its primary (orthosteric) ligand, and can be thought to act like a dimmer switch in an electrical circuit, adjusting the intensity of the response.

For example, the $GABA_A$ receptor has two active sites that the neurotransmitter gamma-aminobutyric acid (GABA) binds, but also has benzodiazepine and general anaesthetic agent regulatory binding sites. These regulatory sites can each produce positive allosteric modulation, potentiating the activity of GABA. Diazepam is an agonist at the benzodiazepine regulatory site, and its antidote flumazenil is an antagonist.

More recent examples of drugs that allosterically modulate their targets include the calcium-mimicking cinacalcet and the HIV treatment maraviroc.

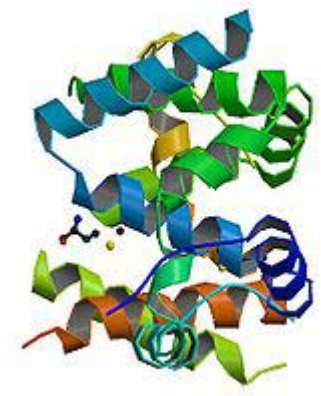
Allosteric sites as drug targets

Allosteric sites may represent a novel drug target. There is a number of advantages in using allosteric modulators as preferred therapeutic agents over classic orthosteric ligands. For example, G protein-coupled receptor (GPCR) allosteric binding sites have not faced the same evolutionary pressure as orthosteric sites to accommodate an endogenous ligand, so are more diverse. Therefore greater GPCR selectivity may be obtained by targeting allosteric sites. This is particularly useful for GPCRs where selective orthosteric therapy has been difficult because of sequence conservation of the orthosteric site across receptor subtypes.Also, these modulators have a decreased potential for toxic effects, since modulators with limited co-operativity will have a ceiling level to their effect, irrespective of the administered dose. Another type of pharmacological selectivity that is unique to allosteric modulators is based on co-operativity. An allosteric modulator may display neutral co-operativity with an orthosteric ligand at all subtypes of a given receptor except the subtype of interest, which is termed "absolute subtype selectivity". If an allosteric modulator does not possess appreciable efficacy, it can provide another powerful therapeutic advantage over orthosteric ligands, namely the ability to selectively tune up or down tissue responses only when the endogenous agonist is present.

Chapter 7

Cell Division Control Protein 4

Cdc4



Crystal structure of Cdc4

Identifiers	
Symbol	Cdc4
Alt. symbols	cell division control protein 4
Entrez	850539
UniProt	P07834

Cdc4 (cell division control protein 4) is a substrate recognition component of the SCF (SKP1-CUL1-F-box protein) ubiquitin ligase complex, which acts as a mediator of ubiquitin transfer to target proteins, leading to their subsequent degradation via the ubiquitin-proteasome pathway. Cdc4 targets primarily cell cycle regulators for proteolysis. It serves the function of an adaptor that brings target molecules to the core SCF complex. Cdc4 was originally identified in the model organism *Saccharomyces cerevisiae*. CDC4 gene function is required at G1/S and G2/M transitions during mitosis and at various stages during meiosis.

Homologues

The human homologue of the cdc4 gene is called FBXW7. The corresponding gene product is the F-box/WD repeat-containing protein 7.

F-box/WD repeat-containing protein 7

Identifiers	
Symbol	Fbw7
Alt. symbols	F-box and WD-40 domain-containing protein 7 F-box protein FBX30 Archipelago homolog (hAgo) SEL-10 hCdc4
UniProt	Q969H0

In the nematode *C. elegans*, the homologue to Cdc4 is F-box/WD repeat-containing protein sel-10.

F-box/WD repeat-containing protein sel-10

Identifiers	
Symbol	F-box/WD repeat-containing protein sel-10
Alt. symbols	Suppressor/enhancer of lin-12 protein 10 Egg laying defective protein 41
UniProt	Q93794

Some general features

Cdc4 has a molecular weight of 86'089Da, an isoelectric point of 7.14, and consists of 779 amino acids. It resides exclusively in the nucleus because of a single monopartite nuclear localisation sequence (NLS) comprising amino acids 82-85 in the N-terminal domain.

Structure

Cdc4 is one component of the E3 complex SCF (CDC4), which comprises CDC53, SKP1, RBX1, and CDC4. Its 779 amino acids (in *S. cerevisiae*) are arranged into one F-box domain (approximately 40 amino acids ("F-box" motif)) and 7 WD repeats.

Cdc4 is a WD-40 repeat F-box protein. Like all members of this family, it contains a conserved dimerization motif called D domain. In yeast Cdc4, the D domain protomers arrange in a superhelical homodimeric manner. SCF (Cdc4) dimerization hardly affects the affinity for target molecules, but significantly increases ubiquitin conjugation. Cdc4 adapts a suprafacial configuration: The substrate-binding sites lie in the same plane AS the catalytic sites, with a separation of 64Å within and 102Å between each SCF monomer. In Cdc4, the substrate binding domain is built on WD40 domains, which use repeats of 40 amino acids), each forming four anti-parallel beta-strands, to assemble the blades of a so-called beta-propeller. Beta-propellers are a quite frequent form of adaptable surface for interaction between different proteins. This substrate interaction region is located C-terminally. There are three isoforms of Cdc4 in mammals: α, ß, and γ. These are produced via alternative splicing of 3 unique 5' exons to 10 common 3' exons. This results in proteins that differ only at their N-termini.

Cdc4 protein interacts with Cdc34, an ubiquitin-conjugating enzyme, and Cdc53 *in vivo*. (There is a Cdc4p/Cdc53p-binding region on Cdc34p.) All three proteins are stable throughout the cell cycle.

Function

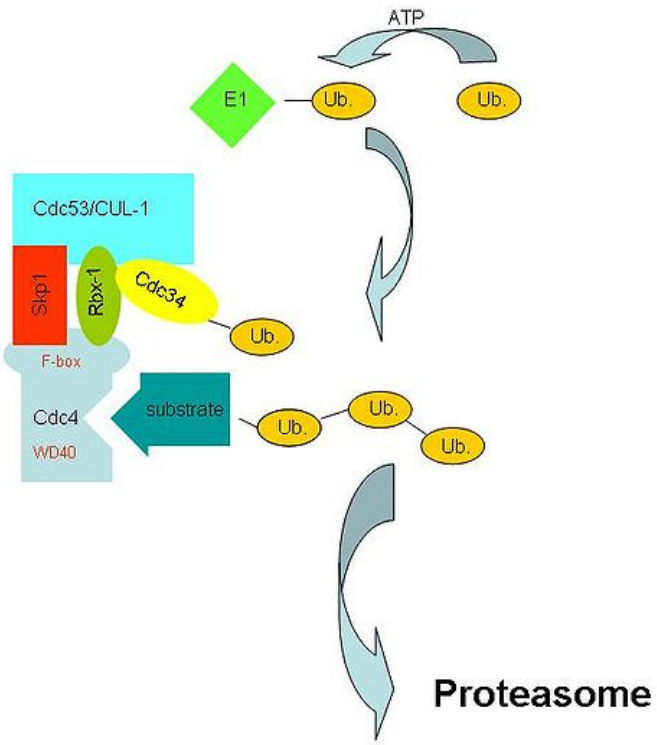


Fig. 1 The figure shows a sketch of how the SCF complex functions, and what the role of Cdc4 is within this structure (-> adaptor function).

Various cellular regulatory mechanisms heavily depend on ubiquitin-dependent degradation. The SCF (Cdc4) complex has a regulatory function in cell cycle progression, signal transduction, and transcription. In order for the cell cycle to proceed, several inhibitory proteins, as well as cyclins, have to be eliminated at given time points. Cdc4 assists there by recruiting target molecules via its C-terminal substrate interaction domain (WD40 repeat domain) to the ubiquitination machinery. This causes transfer of ubiquitin molecules to the target, hence marks it for degradation. Cdc4 recognizes and binds to phosphorylated target proteins.

Cdc4 can be essential, or nonessential, depending on the organism. For instance, it is essential in *S. cerevisiae*, while it is non-essential in *C. albicans*. It is essential for initiation of DNA replication and separation of spindle pole bodies, hence for the formation of the poles of the mitotic spindle. In budding yeast it is also involved in bud development, fusion of zygotic nuclei (karyogamy) after conjugation, and several aspects of sporulation. Roughly spoken, in the cell cycle Cdc4 function is required for G1/S and G2/M transition.

Some important interactions in which Cdc4 is involved are:

- ubiquitination of the phosphorylated form of the cell cycle kinase inhibitor (CKI) SIC1
- degradation of the CKI FAR1 in absence of pheromone; restriction of FAR1 degradation to the nucleus (since Cdc4 is exclusively nuclear)
- transcription activation of the HTA1-HTB1 locus
- degradation of the phosphorylated form of Cdc6

Onset of S-phase

Swi5 is a transcriptional activator of Sic1, which inhibits S-phase CDKs. Thus, Sic1 protein degradation is necessary to enter S-phase. SCF (Cdc4) complex's regulatory function concerning S-phase entry comprises not only degradation of Sic1, but also degradation of Swi5. In order for the substrate adapter unit Cdc4 to bind to Sic1, a minimum of any six of the nine cyclin-dependent kinase sites on Sic1 have to be phosphorylated. In other words: There is a threshold number of phosphorylation sites in order to achieve receptor-ligand binding. As recently stated, this "suggests that the ultrasensitivity in the Sic1-Cdc4 system may be driven at least in part by cumulative electrostatic interactions". In general, an ultrasensitive enzyme requires less than 81-fold increase in stimulus to drive it from 10% to 90% activity. "Ultrasensitivity" highlights that the upstroke of the stimulus/response curve is steeper than the one that is obtained for a hyperbolic Michaelis-Menten enzyme. Thus, ultrasensitivity allows a highly sensitive response: A graded input can be transformed into a sharply thresholded output. The development of B-type cyclin–cyclin-dependent kinase activity, as well as the onset of DNA replication, requires degradation of Sic1 in the late G1 phase of the cell cycle. The WD domain of Cdc4 binds to the phosphorylated form of Sic1. Each bond to a Sic1-

Phosphate is weak, but together the binding is strong enough to enable Sic1-degradation via the pathway described before. Hence, in this case ultrasensitivity allows precise definition ("fine tuning") of the time point in which destruction of Sic1 occurs, leading to initiation of the next step in the cell cycle (-> DNA replication).

G2/M transition

Up until now it is not satisfyingly understood how Cdc4 triggers G2-M transition. In general, the second degradation complex involved in cell cycle progression, APC, is responsible for proteolysis at that stage. However, experimental data suggests that Cdc4 function in G2/M transition may be linked to the degradation of Pds1 (anaphase inhibitor). And what is more, CDC4 and CDC20, an activator of APC, interact genetically.

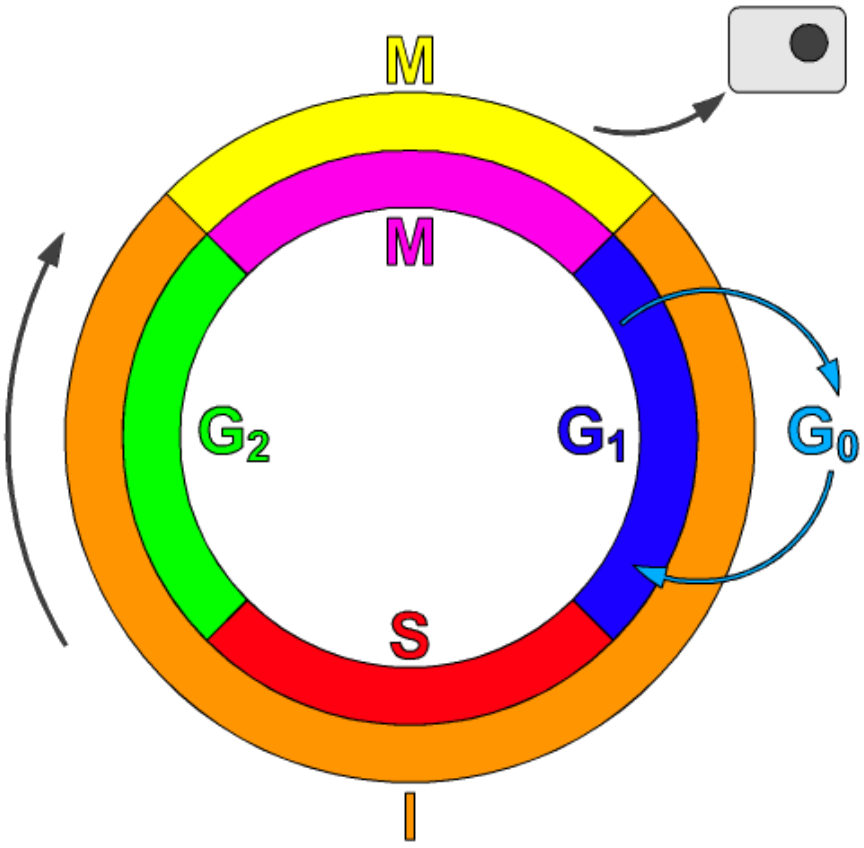
Cdc4 recruits several other substrates than Sic1 to the SCF core complex, including the Cln-Cdc28 inhibitor / cytoskeletal scaffold protein Far1, the transcription factor Gcn4, and the replication protein Cdc6. In addition to those functions mentioned above, Cdc4 is involved in some other degradation-dependent events in S. cerevisiae like for instance unfolded protein response.

Cdc4 and diseases

In mammals, amongst others c-Myc, Src3, Cyclin E, and the Notch intracellular domain are substrates of Cdc4. Due to its involvement in degradation of various cell cycle regulators, as well as several compounds of signaling pathways (e.g. Notch), Cdc4 is a highly sensible component of every organism in which it functions. The cdc4 gene is a haplo-insufficient tumor suppressor gene. Knock-out of this gene in mice leads to an embryonic lethal phenotype. CDC4 mutations occur in a number of cancer types. They are described best in colorectal tumors, and also have been found to be a mutational target in pancreatic cancer. Recently, it was found out that E3 has an additional function to its primary role in the degradation of certain cell cycle regulators: It is also involved in formation of the neural crest. Hence, Cdc4 is a protein “with separable but complementary functions in control of cell proliferation and differentiation” . This evokes the assumption -beyond regulating cell cycle progression- Cdc4 as a tumor suppressor protein may extend its ability to directly regulate tissue differentiation. However, its concrete role in diseases is still to be elucidated.

Chapter 8

Cyclin-Dependent Kinase



Schematic of the cell cycle. outer ring: I=Interphase, M=Mitosis; inner ring: M=Mitosis; G1=Gap phase 1; S=Synthesis; G2=Gap phase 2. The duration of mitosis in relation to the other phases has been exaggerated in this diagram

Cyclin-dependent kinases (CDKs) are a family of protein kinases first discovered for their role in regulating the cell cycle. They are also involved in regulating transcription, mRNA processing, and the differentiation of nerve cells. They are present in all known eukaryotes, and their regulatory function in the cell cycle has been evolutionarily

conserved. In fact, yeast cells can proliferate normally when their CDK gene has been replaced with the homologous human gene. CDKs are relatively small proteins, with molecular weights ranging from 34 to 40 kDa, and contain little more than the kinase domain. By definition, a CDK binds a regulatory protein called a cyclin. Without cyclin, CDK has little kinase activity; only the cyclin-CDK complex is an active kinase. CDKs phosphorylate their substrates on serines and threonines, so they are serine-threonine kinases. The consensus sequence for the phosphorylation site in the amino acid sequence of a CDK substrate is [S/T*]PX[K/R], where S/T* is the phosphorylated serine or threonine, P is proline, X is any amino acid, K is lysine, and R is arginine

Types

Table 1: Known CDKs, their cyclin partners, and their functions in the human and consequences of deletion in mice.

CDK	Cyclin partner	Function	Deletion Phenotype in Mice
Cdk1	Cyclin B	M phase	None. ~E2.5.
Cdk2	Cyclin E	G1/S transition	Reduced size, imparted neural progenitor cell proliferation. Viable, but both males & females sterile.
Cdk2	Cyclin A	S phase, G2 phase	
Cdk3	Cyclin C	G1 phase ?	No defects. Viable, fertile.
Cdk4	Cyclin D	G1 phase	Reduced size, insulin deficient diabetes. Viable, but both male & female infertile.
Cdk5	p35	Transcription	Severe neurological defects. Died immediately after birth.
Cdk6	Cyclin D	G1 phase	
Cdk7	Cyclin H	CDK-activating kinase, transcription	
Cdk8	Cyclin C	Transcription	
Cdk11	Cyclin L		Mitotic defects. E3.5.
	Cyclin F		Defects in extraembryonic tissues. E10.5.
	Cyclin G		

CDKs and Cyclins in the Cell Cycle

Most of the known cyclin-CDK complexes regulate the progression through the cell cycle. Animal cells contain at least nine CDKs, four of which, Cdk1, 2, 3, and 4, are directly involved in cell cycle regulation. In mammalian cells, CDK1, with its partners cyclin A2 and B1, alone can drive the cell cycle. . Another one, Cdk5, is involved

indirectly as the CDK-activating kinase. Cyclin-CDK complexes phosphorylate substrates appropriate for the particular cell cycle phase. Cyclin-CDK complexes in earlier cell-cycle phase help activate cyclin-CDK complexes in later phases.

Table 2: Cyclins and CDKs by Cell-Cycle Phase

Phase	Cyclin	CDK
G0	C	Cdk3
G1	D, E	Cdk4, Cdk2, Cdk6
S	A, E	Cdk2
G2	A	Cdk2, Cdk1
M	B	Cdk1

Table 3: Cyclin-dependent kinases that control the cell cycle in model organisms.

Species	Name	Original name	Size (amino acids)	Function
Saccharomyces cerevisiae	Cdk1	Cdc28	298	All cell-cycle stages
Schizosaccharomyces pombe	Cdk1	Cdc2	297	All cell-cycle stages
Drosophila melanogaster	Cdk1	Cdc2	297	M
	Cdk2	Cdc2c	314	G1/S, S, possibly M
	Cdk4	Cdk4/6	317	G1, promotes growth
Xenopus laevis	Cdk1	Cdc2	301	M
	Cdk2		297	S, possibly M
Homo sapiens	Cdk1	Cdc2	297	M
	Cdk2		298	G1, S, possibly M
	Cdk4		301	G1
	Cdk6		326	G1

A list of CDKs with their regulator protein, cyclin or other.

- CDK1; cyclin A, cyclin B
- CDK2; cyclin A, cyclin E
- *CDK3*; cyclin C
- CDK4; cyclin D1, cyclin D2, cyclin D3
- CDK5; CDK5R1, CDK5R2.
- CDK6; cyclin D1, cyclin D2, cyclin D3

- CDK7; cyclin H
- *CDK8*; cyclin C
- CDK9; cyclin T1, cyclin T2a, cyclin T2b, cyclin K
- *CDK10*
- CDK11 (*CDC2L2*) ; cyclin L
- CDK12 (*CRKRS*) ; cyclin L
- CDK13 (*CDC2L5*) ; cyclin L

Regulation of CDK activity

CDK levels remains relatively constant throughout the cell cycle and most regulation is post-translational. Most knowledge of CDK structure and function is based on CDKs of S. pombe (Cdc2), S. cerevisiae (CDC28), and vertebrates (CDC2 and CDK2). The four major mechanisms of CDK regulation are cyclin binding, CAK phosphorylation, regulatory inhibitory phosphorylation, and binding of CDK inhibitory subunits (CKIs).

Cyclin binding

The active site, or ATP-binding site, of all kinases is a cleft between a small amino-terminal lobe and a larger carboxy-terminal lobe. The structure of human Cdk2 revealed that CDKs have a modified ATP-binding site that can be regulated by cyclin binding. Phosphorylation by CDK-activating kinase (CAK) at Thr 161 on the T-loop increases the complex activity. Without cyclin, a flexible loop called the activation loop or T-loop blocks the cleft, and the position of several key amino acid residues is not optimal for ATP-binding. With cyclin, two alpha helices change position to permit ATP binding. One of them, the L12 helix that comes just before the T-loop in the primary sequence, becomes a beta strand and helps rearrange the T-loop, so it no longer blocks the active site. The other alpha helix called the PSTAIRE helix rearranges and helps changes the position of the key amino acid residues in the active site.

There is considerable specificity in which cyclin binds with CDK. Furthermore, cyclin binding determines the specificity of the cyclin-CDK complex for particular substrates. Cyclins can directly bind the substrate or localize the Cdk to a subcellular area where the substrate is found. Substrate specificity of S cyclins is imparted by the hydrophobic batch (centered on the MRAIL sequence), which has affinity for substrate proteins that contain a hydrophobic RXL (or Cy) motif. Cyclin B1 and B2 can localize Cdk1 to the nucleus and the Golgi, respectively, through a localization sequence outside the Cdk-binding region.

Phosphorylation

Full kinase activity requires an activating phosphorylation on a threonine adjacent to the active site. The identity of the CDK-activating kinase (CAK) that performs this phosphorylation varies across the model organisms. The timing of this phosphorylation

varies as well. In mammalian cells, the activating phosphorylation occurs after cyclin binding. In yeast cells, it occurs before cyclin binding. CAK activity is not regulated by known cell-cycle pathways and cyclin binding is the limiting step for CDK activation.

Unlike activating phosphorylation, CDK inhibitory phosphorylation is vital for regulation of the cell cycle. Various kinases and phosphatases regulate their phosphorylation state. One of the kinases that place the tyrosine phosphate is Wee1, a kinase conserved in all eukaryotes. Fission yeast also contains a second kinase Mik1 that can phosphorylate the tyrosine. Vertebrates contain a different second kinase called Myt1 that is related to Wee1 but can phosphorylate both the threonine and the tyrosine. Phosphatases from the Cdc25 family dephosphorylate both the threonine and the tyrosine.

CDK Inhibitors

A cyclin-dependent kinase inhibitor (CKI) is a protein that interacts with a cyclin-CDK complex to block kinase activity, usually during G1 or in response to signals from the environment or from damaged DNA. In animal cells, there are two major CKI families: the INK4 family and the Cip/Kip family. The INK4 family proteins are strictly inhibitory and bind CDK monomers. Crystal structures of CDK6-INK4 complexes show that INK4 binding twists the CDK to distort cyclin binding and kinase activity. The Cip/Kip family proteins bind both the cyclin and the CDK of a complex and can be inhibitory or activating. Cip/Kip proteins activate cyclin D and CDK4 or CDK6 complexes by enhancing complex formation.

In yeast and Drosophila, CKIs are strong inhibitors of S- and M-CDK, but do not inhibit G1/S-CDKs. During G1, high levels of CKIs prevent cell cycle events from occurring out of order, but do not prevent transition through the Start checkpoint, which is initiated through G1/S-CDKs. Once the cell cycle is initiated, phosphorylation by early G1/S-CDKs leads to destruction of CKIs, relieving inhibition on later cell cycle transitions. In mammalian cells, the CKI regulation works differently. Mammalian protein p27 (Dacapo in Drosophila) inhibits G1/S- and S-CDKs, but does not inhibits S- and M-CDKs.

Suk1 or Cks

The CDKs directly involved in the regulation of the cell cycle associate with small, 9- to 13-kiloDalton proteins called Suk1 or Cks. These proteins are required for CDK function, but their precise role is unknown. Cks1 binds the carboxy lobe of the Cdk, and recognizes phosphorylated residues. It may help the cyclin-CDK complex with substrates that have multiple phosphorylation sites by increasing affinity for the substrate.

Non-cyclin CDK Activators

Viral Cyclins

Viruses can encode proteins with sequence homology to cyclins. One much-studied example is K-cyclin (or v-cyclin) from Kaposi sarcoma herpes virus, which activates CDK6. Viral cyclin-CDK complexes have different substrate specificities and regulation sensitivities.

CDK5 Activators

The proteins p35 and p39 activate CDK5. Although they lack cyclin sequence homology, crystals structures show that p35 folds in a similar way as the cyclins. However, activation of CDK5 does not require activation loop phosphorylation.

RINGO/Speedy

Proteins with no homology to the cyclin family can be direct activators of CDKs. One family of such activators is the RINGO/Speedy family, which was originally discovered in Xenopus. All five members discovered so far directly activate Cdk1 and Cdk2, but the RINGO/Speedy-CDK2 complex recognizes different substrates than cyclin A-CDK2 complex.

History

Leland H. Hartwell, R. Timothy Hunt, and Paul M. Nurse received the 2001 Nobel Prize in Physiology or Medicine for their complete description of cyclin and cyclin-dependent kinase mechanisms, which are central to the regulation of the cell cycle.

Medical significance

CDKs are considered a potential target for anti-cancer medication. If it is possible to selectively interrupt the cell cycle regulation in cancer cells by interfering with CDK action, the cell will die. At present, some CDK inhibitors such as Seliciclib are undergoing clinical trials. Although it was originally developed as a potential anti-cancer drug, in recent laboratory tests Seliciclib has also proven to induce apoptosis in neutrophil granulocytes, which mediate inflammation. This means that novel drugs for treatment of chronic inflammation diseases such as arthritis and cystic fibrosis could be developed.

Flavopiridol (Alvocidib) is the first CDK inhibitor to be tested in clinical trials after being identified in an anti-cancer agent screen in 1992. It competes for the ATP site of the CDKs.

More research is required, however, because disruption of the CDK-mediated pathway has potentially serious consequences; while CDK inhibitors seem promising, it has to be

determined how side-effects can be limited so that only target cells are affected. As such diseases are currently treated with glucocorticoids, which have often serious side-effects, even a minor success would mean an improvement.

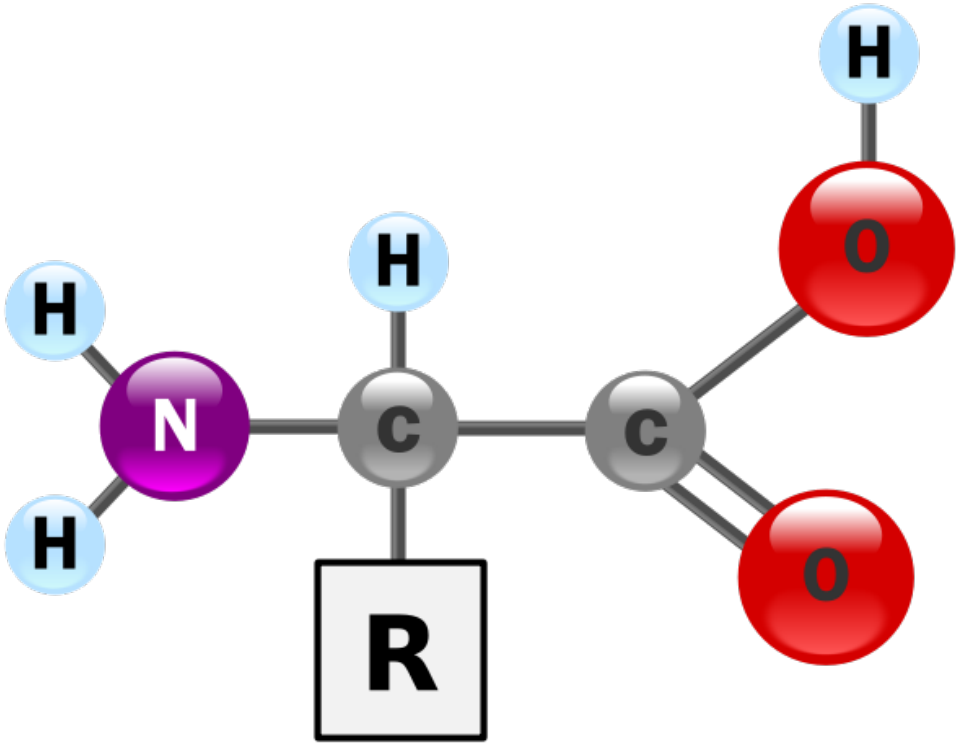
Complications of developing a CDK drug include the fact that many CDKs are not involved in the cell cycle such as transcription, viral infection, neural physiology, and glucose homeostatsis. .

Table 4: Cyclin-dependent kinase inhibitor drugs

Drug	CDKs Inhibited
Flavopiridol (Alvocidib)	1, 2, 4, 6, 9
Olomoucine	1, 2, 5
Roscovitine	1, 2, 5
Purvalanol	1, 2, 5
Paullones	1, 2, 5
Butryolactone	1, 2, 5
Thio/oxoflavopiridols	1
Oxindoles	2
Aminothiazoles	4
Benzocarbazoles	4
Pyrimidines	4
Seliciclib	

Chapter 9

Protein (Nutrient)



Amino acids are the building blocks of protein.

Proteins are polymer chains made of amino acids linked together by peptide bonds. Amino acids can be divided into either essential amino acids or non-essential amino acids. Proteins and carbohydrates contain 4 kcal per gram as opposed to lipids which contain 9 kcal per gram. The essential amino acids, which must be obtained from food sources, are leucine, isoleucine, valine, lysine, threonine, tryptophan, methionine, phenylalanine and histidine. On the other hand, non-essential amino acids can be made by the body from other amino acids. The non-essential amino acids are arginine, alanine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, proline, serine, and tyrosine.

In **nutrition**, proteins are broken down in the stomach during digestion by enzymes known as proteases into smaller polypeptides to provide amino acids for the body, including the essential amino acids that the organism cannot be biosynthesized by the body itself. Thus, protein from one's diet should provide both essential and non-essential amino acids for protein synthesis.

Most animal sources and certain vegetable sources have the complete complement of all the essential amino acids in adequate proportions. However, it is not necessary to consume a single food source that contains all the essential amino acids, as long as all the essential amino acids are eventually present in the diet.

Lentils and rice combined is a plant source of protein

Protein functions in body

Protein is a nutrient needed by the human body for growth and maintenance. Aside from water, protein is the most abundant molecule in the body. Protein is found in all cells of the body and is the major structural component of all cells in the body, especially muscle. This also includes body organs, hair and skin. Proteins also are utilized in membranes, such as glycoproteins. When broken down into amino acids, they are used as precursors to nucleic acid and vitamins. Hormones and enzymes are also formed from amino acids in which they help regulate metabolism, support the immune system and other body functions. Finally, protein is needed to form blood cells.

Protein function in exercise

Proteins are one of the key nutrients for success in terms of sports. They play a major role in the response to exercise. Amino acids, the building blocks of proteins, are used for building new tissue, including muscle, as well as repairing damaged tissues. Proteins, however, only provide a small source of fuel for the exercising muscles when carbohydrates and lipid resources are low.

Sources

Protein milkshakes, made from protein powder (center) and milk (left), are a common bodybuilding supplement.

There are many different sources of protein ranging from whole protein foods (such as milk, meat, fish, egg, and vegetables) to a variety of protein powders (such as casein, whey, soy). Protein powders are processed and manufactured sources of protein. Protein powders may provide an additional source of protein for exercising muscles. The type of protein is important in terms of its influence on protein metabolic response and possibly on the muscle's exercise performance. The different physical and/or chemical properties within the various types of protein may affect the rate of protein digestion. As a result, the amino acid availability and the accumulation of tissue protein is altered because of the various protein metabolic responses.

Food	Amount of protein (grams)
Spirulina 1 cup	64
Tempeh 1 cup	41
Dried Parsley 1 cup	31
Lentils, cooked 1 cup	18
Black Beans, cooked 1 cup	15
Tofu, firm 4 oz	11
Quinoa, cooked 1 cup	9
Peanut Butter 2 tbsp	8
Almonds 1/4 cup	8
Sun-dried Tomato 1 cup	8
Brown Rice, cooked 1 cup	5
Broccoli, cooked 1 cup	4
Potato 1 med.	4
Lambsquarters 1 cup	4

This table compares the different plant foods with the amount of protein in grams per serving

Protein quality

Different proteins have different levels of biological availability (BA) to the human body. Many methods have been introduced to measure protein utilization and retention rates in humans. They include biological value, net protein utilization, and PDCAAS (Protein Digestibility Corrected Amino Acids Score) which was developed by the FDA as an improvement over the Protein Efficiency Ratio (PER) method. These methods examine which proteins are most efficiently used by the body.

Eggs have been determined to have the standard biological value of 100 (though some sources may have higher biological values), which means that most of the absorbed nitrogen from egg protein can be retained and used by the body. Corn has a BA of 70 while peanuts have a relatively low BA of 40.

Digestion

Digestion typically begins in the stomach when pepsinogen is converted to pepsin by the action of hydrochloric acid, and continued by trypsin and chymotrypsin in the intestine. The amino acids and their derivatives into which dietary protein is degraded are then absorbed by the gastrointestinal tract. The absorption rates of individual amino acids are highly dependent on the protein source; for example, the digestibilities of many amino acids in humans, the difference between soy and milk proteins and between individual

milk proteins, beta-lactoglobulin and casein. For milk proteins, about 50% of the ingested protein is absorbed between the stomach and the jejunum and 90% is absorbed by the time the digested food reaches the ileum. Biological value (BV) is a measure of the proportion of absorbed protein from a food which becomes incorporated into the proteins of the organism's body.

Dietary requirements

Considerable debate has taken place regarding issues surrounding protein intake requirements. How much protein needed in a person's daily diet is determined in large part by overall energy intake, by the body's need for nitrogen and essential amino acids, body weight and composition, rate of growth in the individual, physical activity level, individual's energy and carbohydrate intake, as well as the presence of illness or injury. Physical activity and exertion as well as enhanced muscular mass increase the need for protein. Requirements are also greater during childhood for growth and development, during pregnancy or when breast-feeding in order to nourish a baby, or when the body needs to recover from malnutrition or trauma or after an operation.

If enough energy is not taken in through diet, as in the process of starvation, the body will use protein from the muscle mass to meet its energy needs, leading to muscle wasting over time. If the individual does not consume adequate protein in nutrition, then muscle will also waste as more vital cellular processes (e.g. respiration enzymes, blood cells) recycle muscle protein for their own requirements.

According to US & Canadian Dietary Reference Intake guidelines, women aged 19–70 need to consume 46 grams of protein per day, while men aged 19–70 need to consume 56 grams of protein per day to avoid a deficiency. The American and Canadian guidelines recommend a daily protein dietary allowance, measured as intake per kilogram body weight, is 0.8 g/kg. However, this recommendation is based on structural requirements, but disregards use of protein for energy metabolism. This requirement is for a normal sedentary person.

Several studies have concluded that active people and athletes may require elevated protein intake (compared to 0.8 g/kg) due to increase in muscle mass and sweat losses, as well as need for body repair and energy source. Suggested amounts vary between 1.6 g/kg and 1.8 g/kg, while a proposed *maximum* daily protein intake would be approximately 25% of energy requirements i.e. approximately 2 to 2.5 g/kg. However, many questions still remain to be resolved.

Aerobic exercise protein needs

Endurance athletes differ from strength-building athletes in that endurance athletes do not build muscle mass from training. Research suggests that individuals performing endurance activity require more protein intake than sedentary individuals so that muscles broken down during endurance workouts can be repaired. Although the protein requirement for athletes still remains controversial, research does show that endurance

athletes can benefit from increasing protein intake because the type of exercise endurance athletes participate in still alters the protein metabolism pathway. The overall protein requirement increases because of amino acid oxidation in endurance-trained athletes. Endurance athletes who exercise over a long period (2–5 hours per training session) use protein as a source of 5–10% of their total energy expended. Therefore, a slight increase in protein intake may be beneficial to endurance athletes by replacing the protein lost in energy expenditure and protein lost in repairing muscles. Some scientists suggest that endurance athletes may increase daily protein intake to a maximum of 1.2–1.4 g per kg body weight.

Anaerobic exercise protein needs

Research also indicates that individuals performing strength-training activity require more protein intake than sedentary individuals. Strength-training athletes may increase their daily protein intake to a maximum of 1.4–1.8 g per kg body weight to enhance muscle protein synthesis or to make up for the loss of amino acid oxidation during exercise. This protein intake can easily be achieved because most athletes consume very high energy intake. Often anaerobic exercise athletes assume that this level of protein intake is very high and tend to over consume. Research shows that many athletes do already consume intakes above the recommended intake even without the use of protein supplements.

Excess consumption

When a high dietary protein intake is consumed, there is an increase in urea excretion, which suggests that amino acid oxidation is increased. High levels of protein intake increases the activity of branched-chain ketoacid dehydrogenase. As a result, oxidation is facilitated and the amino group of the amino acid is excreted to the liver. This process suggests that excess protein consumption results in protein oxidation and that the protein is excreted. The body is unable to store excess protein. Protein is digested into amino acids which enter the bloodstream. Excess amino acids are converted to other usable molecules by the liver in a process called deamination. Deamination converts nitrogen from the amino acid into ammonia which is converted by the liver into urea in the urea cycle. Excretion of urea is performed by the kidneys. These organs can normally cope with any extra workload but if a kidney disease occurs, a decrease in protein will often be prescribed. Furthermore, as noted protein provides the body with 4 Calories per gram, when there is excess protein intake the body will utilize as much of it for energy. After that stage, the body will produce fat from the excess protein, turning it into fat cells. On the other hand, if people do not eat enough calories, body protein and protein from the food will be utilized for energy. This is not ideal as the main function of protein is to maintain the body. Finally, as food protein intake is high or low, the body tries to keep protein levels at an equilibrium. This concept is known as the “labile protein reserve” that serves as a short-term protein store to be used for emergencies or daily variations in protein intake. It is not utilized as longer-term storage for future needs.

Many researchers have also found that excessive intake of protein increases calcium excretion in urine. It has been thought that this occurs to maintain the pH unbalance from the oxidation of sulfur amino acids. Also whether if bone resorption contributes to bone loss and osteoporosis is inconclusive. However, it is also found that a regular intake of calcium would be able to stabilize this loss.

Another issue arising from overconsumption of protein is a higher risk of kidney stone formation from calcium in the renal circulatory system. It has been found that high animal protein intake in healthy individuals increases probability of forming kidney stones by 250 percent.

Testing in foods

The classic assays for protein concentration in food are the Kjeldahl method and the Dumas method. These tests determine the total nitrogen in a sample. The only major component of most food which contains nitrogen is protein (fat, carbohydrate and dietary fibre do not contain nitrogen). If the amount of nitrogen is multiplied by a factor depending on the kinds of protein expected in the food the total protein can be determined. This value is known as the "crude protein" content. On food labels the protein is given by the nitrogen multiplied by 6.25, because the average nitrogen content of proteins is about 16%. The Kjeldahl test is typically used because it is the method the AOAC International has adopted and is therefore used by many food standards agencies around the world, though the Dumas method is also approved by some standards organizations.

Accidental contamination and intentional adulteration of protein meals with non-protein nitrogen sources that inflate crude protein content measurements have been known to occur in the food industry for decades. To ensure food quality, purchasers of protein meals routinely conduct quality control tests designed to detect the most common non-protein nitrogen contaminants, such as urea and ammonium nitrate.

In at least one segment of the food industry, the dairy industry, some countries (at least the U.S., Australia, France and Hungary), have adopted "true protein" measurement, as opposed to crude protein measurement, as the standard for payment and testing: "True protein is a measure of only the proteins in milk, whereas crude protein is a measure of all sources of nitrogen and includes nonprotein nitrogen, such as urea, which has no food value to humans. ... Current milk-testing equipment measures peptide bonds, a direct measure of true protein." Measuring peptide bonds in grains has also been put into practice in several countries including Canada, the UK, Australia, Russia and Argentina where near-infrared reflectance (NIR) technology, a type of infrared spectroscopy is used. The Food and Agriculture Organization of the United Nations (FAO) recommends that only amino acid analysis be used to determine protein in, *inter alia*, foods used as the sole source of nourishment, such as infant formula, but also provides: "When data on amino acids analyses are not available, determination of protein based on total N content by Kjeldahl (AOAC, 2000) or similar method ... is considered acceptable."

The limitations of the Kjeldahl method were at the heart of the Chinese protein export contamination in 2007 and the 2008 Chinese Milk Scandal in which the industrial chemical melamine was added to the milk or glutens to increase the measured "protein".

Food allergies

Specific proteins found in certain food items are often the cause of allergies and allergic reactions. This is because the structure of each form of protein is slightly different; some may trigger a response from the immune system while others remain harmless. Many people are allergic to casein, the protein in milk; gluten, the protein in wheat and other grains; the particular proteins found in peanuts; or those in shellfish or other seafoods. Food allergies should not be confused with food intolerance.

Deficiency in developing countries.

Protein deficiency is a serious cause of ill health and death in developing countries. Protein deficiency plays a part in the disease kwashiorkor. Famine, overpopulation and other factors can increase rates of malnutrition and protein deficiency. Protein deficiency can lead to reduced intelligence or mental retardation.

In countries that suffer from widespread protein deficiency, food is generally full of plant fibers, which makes adequate energy and protein consumption very difficult. Protein deficiency is generally caused by lack of total food energy, making it an issue of not getting food in total. Symptoms of kwashiorkor include apathy, diarrhea, inactivity, failure to grow, flaky skin, fatty liver, and edema of the belly and legs. This edema is explained by the normal functioning of proteins in fluid balance and lipoprotein transport.

Moringa trees are known to overcome protein deficiency in developing countries as the leaves and other parts of the tree contain comparably to soy bean high amount of crude proteins and amino acids.

Dr. Latham, director of the Program in International Nutrition at Cornell University claims that malnutrition is a frequent cause of death and disease in third world countries. Protein-energy malnutrition (PEM) affects 500 million people and kills 10 million annually. In severe cases white blood cell numbers decline and the ability of leukocytes to fight infection decreases.

Future directions in research

More research is underway to better understand whether or not there is a standardized protein amount needed per day. Because of some difficulties in scientific techniques used to measure the increase in daily protein needs from a sedentary person to an athlete, the increase in protein needed is therefore not universal. Finally, the effects of increased protein intake, both animal and plant, should be furthered tested with respect to increasing osteoporosis risk.

Chapter 10

Antifreeze Protein

Antifreeze proteins (AFPs) or **ice structuring proteins (ISPs)** refer to a class of polypeptides produced by certain vertebrates, plants, fungi and bacteria that permit their survival in subzero environments. AFPs bind to small ice crystals to inhibit growth and recrystallization of ice that would otherwise be fatal. There is also increasing evidence that AFPs interact with mammalian cell membranes to protect them from cold damage. This work suggests the involvement of AFPs in cold acclimatization.

Non-colligative properties

Unlike the widely used automotive antifreeze, ethylene glycol, AFPs do not lower freezing point in proportion to concentration. Rather, they work in a noncolligative manner. This allows them to act as an antifreeze at concentrations 1/300th to 1/500th of those of other dissolved solutes. This minimizes their effect on osmotic pressure. The unusual capabilities of AFPs are attributed to their binding ability at specific ice crystal surfaces.

Thermal hysteresis

AFPs create a difference between the melting point and freezing point known as thermal hysteresis. The addition of AFPs at the interface between solid ice and liquid water inhibits the thermodynamically favored growth of the ice crystal. Ice growth is kinetically inhibited by the AFPs covering the water-accessible surfaces of ice.

Thermal hysteresis is easily measured in the lab with a nanolitre osmometer. Different organisms have different values of thermal hysteresis. The maximum level of thermal hysteresis shown by fish AFP is approximately -1.5°C (2.7°F). However, insect antifreeze proteins are 10–30 times more active than any known fish protein. This is probably because insects encounter lower temperatures on land than the –1°C or –2°C that fish face in freezing waters. During the extreme winter months, the spruce budworm can battle temperatures approaching –30°C and resist freezing, though the Alaskan beetle *Upis ceramboides* can survive in a temperature of –60°C by using an antifreeze molecule that is not composed of proteins.

The rate of cooling can influence the thermal hysteresis value of AFPs. Rapid cooling can substantially decrease the nonequilibrium freezing point, and hence the thermal hysteresis value. This means organisms may not be able to adapt to their subzero environment if the temperature drops abruptly.

Freeze tolerance versus *freeze avoidance*

Species containing AFPs may be classified as:

Freeze avoidant: These species are able to prevent their body fluids from freezing altogether. Generally, the AFP function may be overcome at extremely cold temperatures, leading to rapid ice growth and death.

Freeze tolerant: These species are able to survive body fluid freezing. Some freeze tolerant species are thought to use AFPs as cryoprotectants to prevent the damages of freezing, but not freezing altogether. The exact mechanism is still unknown. However, it is thought AFPs may inhibit recrystallization and stabilize cell membranes to prevent damage by ice. They may work in conjunction with protein ice nucleators (PINs) to control the rate of ice propagation following freezing.

Diversity

There are many known nonhomologous types of AFP.

Fish AFPs

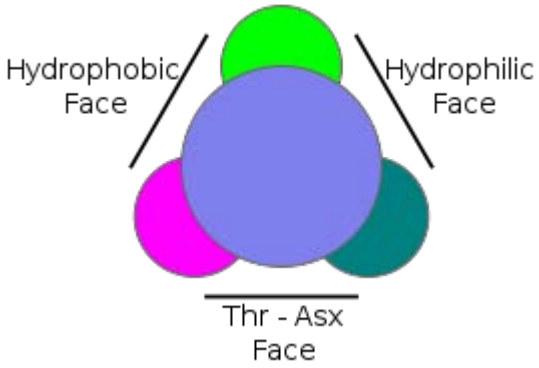


Figure 1. The three faces of Type I AFP

Antifreeze glycoproteins or AFGPs are found in Antarctic notothenioids and northern cod. They are 2.6-3.3 kD.

Type I AFP is found in winter flounder, longhorn sculpin and shorthorn sculpin. It is the best documented AFP because it was the first to have its three dimensional structure determined. Type I AFP consists of a single, long, amphipathic alpha helix, about 3.3-4.5

kD in size. There are three faces to the 3D structure: the hydrophobic, hydrophilic, and Thr-Asx face.

Type I-hyp AFP (where hyp stands for hyperactive) are found in several righteye flounders. It is approximately 32 kD (two 17 kD dimeric molecules). The protein was isolated from the blood plasma of winter flounder. It is considerably better at depressing freezing temperature than most fish AFPs.

Type II AFPs are found in sea raven, smelt and herring. They are cysteine-rich globular proteins containing five disulfide bonds.

Type III AFPs are found in Antarctic eelpout. They exhibit similar overall hydrophobicity at ice binding surfaces to type I AFPs. They are approximately 6kD in size.

Type IV AFPs are found in longhorn sculpins. They are alpha helical proteins rich in glutamate and glutamine. This protein is approximately 12KDa in size and consists of a 4-helix bundle. Its only posttranslational modification is a pyroglutamate residue, a cyclized glutamine residue at its N-terminal. Scientists at the University of Guelph in Canada are currently examining the role of this pyroglutame residue in the antifreeze activity of type IV AFP from the longhorn sculpin.

Plant AFPs

The classification of AFPs became more complicated when antifreeze proteins from plant were discovered. Plant AFPs are rather different from the other AFPs in the following aspects:

1. They have much weaker thermal hysteresis activity when compared to other AFPs.
2. Their physiological function is likely in inhibiting the recrystallization of ice rather than in the preventing ice formation.
3. Most of them are evolved pathogenesis-related proteins, sometimes retaining antifungal properties.

Insect AFPs

There are two types of insect antifreeze proteins, *Tenebrio* and *Dendroides* AFPs which are both in different insect families. They are similar to one another, both being hyperactive (i.e. greater thermal hysteresis value) and consist of varying numbers of 12- or 13-mer repeats of approximately 8.3 to 12.5 kD. Throughout the length of the protein, at least every sixth residue is a cysteine.

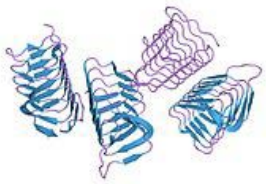
Tenebrio or Type V AFPs are found in beetles, whereas *Dendroides* or *Choristoneura fumiferana* AFPs are found in some Lepidoptera.

Insect antifreeze protein repeat

Structure of the *Tenebrio molitor* beta-helical antifreeze protein

Identifiers	
Symbol	AFP
Pfam	PF02420
InterPro	IPR003460
SCOP	1ezg
Available protein structures:	

***Choristoneura fumiferana* antifreeze protein (CfAFP)**



Structure of *Choristoneura fumiferana* (spruce budworm) beta-helical antifreeze protein

Identifiers	
Symbol	CfAFP
Pfam	PF05264
InterPro	IPR007928
SCOP	1m8n
Available protein structures:	

Sea ice organisms AFPs

AFPs were also found in microorganisms living in sea ice. The diatoms *Fragilariopsis cylindrus* and *F. curta* play a key role in polar sea ice communities, dominating the assemblages of both platelet layer and within pack ice. AFPs are widespread in these species, and the presence of AFP genes as a multigene family indicates the importance of this group for the genus *Fragilariopsis*. AFPs identified in *F. cylindrus* belong to an AFP family which is represented in different taxa and can be found in other organisms related to sea ice (*Colwellia* spp., *Navicula glaciei*, *Chaetoceros neogracile* and *Stephos longipes*) and Antarctic inland ice bacteria (Flavobacteriaceae,) as well as in cold-tolerant fungi (*Typhula ishikariensis*, *Lentinula edodes* and *Flammulina populicola*.)

Evolution

The remarkable diversity and distribution of AFPs suggest the different types evolved recently in response to sea level glaciation occurring 1-2 million years ago in the Northern hemisphere and 10-30 million years ago in Antarctica. This independent development of similar adaptations is referred to as convergent evolution. There are two reasons why many types of AFPs are able to carry out the same function despite their diversity:

1. Although ice is uniformly composed of oxygen and hydrogen, it has many different surfaces exposed for binding. Different types of AFPs may interact with different surfaces.
2. Although the five types of AFPs differ in their primary sequence of amino acids, when each folds into a functioning protein, they may share similarities in their three dimensional or tertiary structure that facilitates the same interactions with ice.

Mechanisms of action

AFPs are thought to inhibit growth by an adsorption–inhibition mechanism. They adsorb to nonbasal planes of ice, inhibiting thermodynamically favored ice growth. The presence of a flat, rigid surface in AFPs seems to facilitate its interaction with ice via Van der Waals force surface complementarity.

Binding to ice

Normally, ice crystals grown in solution only exhibit the basal (0001) and prism faces (1010), and appear as round and flat discs. However, it appears the presence of AFPs exposes other faces. It now appears the ice surface 2021 is the preferred binding surface, at least for AFP type I. Through studies on type I AFP, ice and AFP were initially thought to interact through hydrogen bonding (Raymond and DeVries, 1977). However, when parts of the protein thought to facilitate this hydrogen bonding were mutated, the hypothesized decrease in antifreeze activity was not observed. Recent data suggest hydrophobic interactions could be the main contributor. It is difficult to discern the exact mechanism of binding because of the complex water-ice interface. Currently, attempts to uncover the precise mechanism are being made through use of molecular modelling programs (molecular dynamics or the Monte Carlo method).

Binding mechanism and antifreeze function

According to the structure and function study on the antifreeze protein from the fish winter flounder, the antifreeze mechanism of the AFP molecule was shown to be due to the binding to an ice nucleation structure in a zipper-like fashion through hydrogen bonding of the hydroxyl groups of its four Thr residues to the oxygens along the $[01\bar{1}2]$ direction in ice lattice, subsequently stopping or retarding the growth of ice pyramidal planes so as to depress the freeze point.

The above mechanism can be used to elucidate the structure-function relationship of other antifreeze proteins with the following two common features:

1. recurrence of a Thr residue (or any other polar amino acid residue whose side-chain can form a hydrogen bond with water) in an 11-amino-acid period along the sequence concerned, and
2. a high percentage of an Ala residue component therein.

History

In the 1950s, Canadian scientist Scholander set out to explain how Arctic fish can survive in water colder than the freezing point of their blood. His experiments led him to believe there was “antifreeze” in the blood of Arctic fish. Then in the late 1960s, animal biologist Arthur DeVries was able to isolate the antifreeze protein through his investigation of Antarctic fish. AFGPs were the first AFPs to be discovered. At the time, they were called "glycoproteins as biological antifreeze agents". These proteins were later called antifreeze glycoproteins (AFGPs) or antifreeze glycopeptides to distinguish them from newly discovered nonglycoprotein biological antifreeze agents (AFPs). DeVries worked with Robert Feeney (1970) to characterize the chemical and physical properties of antifreeze proteins. In 1992, Griffith *et al.* documented their discovery of AFP in winter rye leaves. Around the same time, Urrutia, Duman and Knight (1992) documented thermal hysteresis protein in angiosperms. The next year, Duman and Olsen noted AFPs had also been discovered in over 23 species of angiosperms, including ones eaten by humans. As well, they reported their presence in fungi and bacteria.

Name change

Recent attempts have been made to relabel antifreeze proteins as ice structuring proteins to more accurately represent their function and to dispose of any assumed negative relation between AFPs and automotive antifreeze, ethylene glycol. These two things are completely separate entities, and show loose similarity only in their function.

Commercial applications

Commercially, there appear to be countless applications for antifreeze proteins. Numerous fields would be able to benefit from the protection of tissue damage by freezing. Businesses are currently investigating the use of these proteins in:

- increasing freeze tolerance of crop plants and extending the harvest season in cooler climates
- improving farm fish production in cooler climates
- lengthening shelf life of frozen foods
- improving cryosurgery
- enhancing preservation of tissues for transplant or transfusion in medicine
- therapy for hypothermia

Recent news

One recent, successful business endeavor has been the introduction of AFPs into ice cream and yogurt products. This ingredient, labelled ice-structuring protein, has been approved by the Food and Drug Administration. The proteins are isolated from fish and replicated, on a larger scale, in yeast.

There is concern from organizations opposed to genetically modified organisms (GMOs), arguing modified antifreeze proteins may cause inflammation. However, as stated, ISPs have been approved for human consumption following diligent tests. Intake of AFPs in diet is likely substantial in most northerly and temperate regions already. Given the known historic consumption of AFPs, it is safe to conclude their functional properties do not impart any toxicologic or allergenic effects in humans.

As well, the transgenic process of ISP production is widely used in society already. This is how mass amounts of insulin are made to treat people with type I diabetes each year. The process does not impact the product; it merely makes production more efficient and prevents the death of many fish who would, otherwise, be killed for the extraction of such protein.

Currently, Unilever incorporates AFPs into some of its American products, including some popsicles and a new line of Breyers Light Double Churned ice cream bars. In ice cream, AFPs allow the production of very creamy, dense, reduced fat ice cream with fewer additives. They control ice crystal growth brought on by thawing on the loading dock or kitchen table which drastically reduces texture quality.

In November 2009, the Proceedings of the National Academy of Sciences published the discovery of a molecule in an Alaskan beetle that behaves like AFPs, but is composed of saccharides and fatty acids.

A 2010 study demonstrated the stability of superheated water ice crystals in an AFP solution, showing while the proteins can inhibit freezing, they can also inhibit melting.

Chapter 11

Protein Purification

Protein purification is a series of processes intended to isolate a single type of protein from a complex mixture. Protein purification is vital for the characterization of the function, structure and interactions of the protein of interest. The starting material is usually a biological tissue or a microbial culture. The various steps in the purification process may free the protein from a matrix that confines it, separate the protein and non-protein parts of the mixture, and finally separate the desired protein from all other proteins. Separation of one protein from all others is typically the most laborious aspect of protein purification. Separation steps may exploit differences in (for example) protein size, physico-chemical properties, binding affinity and biological activity.

Purpose

Purification may be *preparative* or *analytical*. Preparative purifications aim to produce a relatively large quantity of purified proteins for subsequent use. Examples include the preparation of commercial products such as enzymes (e.g. lactase), nutritional proteins (e.g. soy protein isolate), and certain biopharmaceuticals (e.g. insulin). Analytical purification produces a relatively small amount of a protein for a variety of research or analytical purposes, including identification, quantification, and studies of the protein's structure, post-translational modifications and function. Among the first purified proteins were urease and Concanavalin A.

Strategies

Recombinant bacteria can be grown in a flask containing growth media.

Choice of a starting material is key to the design of a purification process. In a plant or animal, a particular protein usually isn't distributed homogeneously throughout the body; different organs or tissues have higher or lower concentrations of the protein. Use of only the tissues or organs with the highest concentration decreases the volumes needed to produce a given amount of purified protein. If the protein is present in low abundance, or if it has a high value, scientists may use recombinant DNA technology to develop cells that will produce large quantities of the desired protein (this is known as an expression system). Recombinant expression allows the protein to be tagged, e.g. by a His-tag, to facilitate purification, which means that the purification can be done in fewer steps. In

addition, recombinant expression usually starts with a higher fraction of the desired protein than is present in a natural source.

An analytical purification generally utilizes three properties to separate proteins. First, proteins may be purified according to their isoelectric points by running them through a pH graded gel or an ion exchange column. Second, proteins can be separated according to their size or molecular weight via size exclusion chromatography or by SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) analysis. Proteins are often purified by using 2D-PAGE and are then analysed by peptide mass fingerprinting to establish the protein identity. This is very useful for scientific purposes and the detection limits for protein are nowadays very low and nanogram amounts of protein are sufficient for their analysis. Thirdly, proteins may be separated by polarity/hydrophobicity via high performance liquid chromatography or reversed-phase chromatography.

Evaluating purification yield

The most general method to monitor the purification process is by running a SDS-PAGE of the different steps. This method only gives a rough measure of the amounts of different proteins in the mixture, and it is not able to distinguish between proteins with similar apparent molecular weight.

If the protein has a distinguishing spectroscopic feature or an enzymatic activity, this property can be used to detect and quantify the specific protein, and thus to select the fractions of the separation, that contains the protein. If antibodies against the protein are available then western blotting and ELISA can specifically detect and quantify the amount of desired protein. Some proteins function as receptors and can be detected during purification steps by a ligand binding assay, often using a radioactive ligand.

In order to evaluate the process of multistep purification, the amount of the specific protein has to be compared to the amount of total protein. The latter can be determined by the Bradford total protein assay or by absorbance of light at 280 nm, however some reagents used during the purification process may interfere with the quantification. For example, imidazole (commonly used for purification of polyhistidine-tagged recombinant proteins) is an amino acid analogue and at low concentrations will interfere with the bicinchoninic acid (BCA) assay for total protein quantification. Impurities in low-grade imidazole will also absorb at 280 nm, resulting in an inaccurate reading of protein concentration from UV absorbance.

Another method to be considered is Surface Plasmon Resonance (SPR). SPR can detect binding of label free molecules on the surface of a chip. If the desired protein is an antibody, binding can be translated directly to the activity of the protein. One can express the active concentration of the protein as the percent of the total protein. SPR can be a powerful method for quickly determining protein activity and overall yield. It is a powerful technology that requires an instrument to perform.

Methods of protein purification

The methods used in protein purification can roughly be divided into analytical and preparative methods. The distinction is not exact, but the deciding factor is the amount of protein that can practically be purified with that method. Analytical methods aim to detect and identify a protein in a mixture, whereas preparative methods aim to produce large quantities of the protein for other purposes, such as structural biology or industrial use. In general, the preparative methods can be used in analytical applications, but not the other way around.

Extraction

Depending on the source, the protein has to be brought into solution by breaking the tissue or cells containing it. There are several methods to achieve this: Repeated freezing and thawing, sonication, homogenization by high pressure, filtration (either via cellulose-based depth filters or cross-flow filtration), or permeabilization by organic solvents. The method of choice depends on how fragile the protein is and how sturdy the cells are. After this extraction process soluble proteins will be in the solvent, and can be separated from cell membranes, DNA etc. by centrifugation. The extraction process also extracts proteases, which will start digesting the proteins in the solution. If the protein is sensitive to proteolysis, it is usually desirable to proceed quickly, and keep the extract cooled, to slow down proteolysis.

Precipitation and differential solubilization

In bulk protein purification, a common first step to isolate proteins is precipitation with ammonium sulfate $(NH_4)_2SO_4$. This is performed by adding increasing amounts of ammonium sulfate and collecting the different fractions of precipitate protein. One advantage of this method is that it can be performed inexpensively with very large volumes.

The first proteins to be purified are water-soluble proteins. Purification of integral membrane proteins requires disruption of the cell membrane in order to isolate any one particular protein from others that are in the same membrane compartment. Sometimes a particular membrane fraction can be isolated first, such as isolating mitochondria from cells before purifying a protein located in a mitochondrial membrane. A detergent such as sodium dodecyl sulfate (SDS) can be used to dissolve cell membranes and keep membrane proteins in solution during purification; however, because SDS causes denaturation, milder detergents such as Triton X-100 or CHAPS can be used to retain the protein's native conformation during complete purification.

Ultracentrifugation

Centrifugation is a process that uses centrifugal force to separate mixtures of particles of varying masses or densities suspended in a liquid. When a vessel (typically a tube or bottle) containing a mixture of proteins or other particulate matter, such as bacterial cells,

is rotated at high speeds, the angular momentum yields an outward force to each particle that is proportional to its mass. The tendency of a given particle to move through the liquid because of this force is offset by the resistance the liquid exerts on the particle. The net effect of "spinning" the sample in a centrifuge is that massive, small, and dense particles move outward faster than less massive particles or particles with more "drag" in the liquid. When suspensions of particles are "spun" in a centrifuge, a "pellet" may form at the bottom of the vessel that is enriched for the most massive particles with low drag in the liquid. Non-compacted particles still remaining mostly in the liquid are called the "supernatant" and can be removed from the vessel to separate the supernatant from the pellet. The rate of centrifugation is specified by the angular acceleration applied to the sample, typically measured in comparison to the *g*. If samples are centrifuged long enough, the particles in the vessel will reach equilibrium wherein the particles accumulate specifically at a point in the vessel where their buoyant density is balanced with centrifugal force. Such an "equilibrium" centrifugation can allow extensive purification of a given particle.

Sucrose gradient centrifugation — a linear concentration gradient of sugar (typically sucrose, glycerol, or a silica based density gradient media, like Percoll) is generated in a tube such that the highest concentration is on the bottom and lowest on top. Percoll is a trademark owned by GE Healthcare companies. A protein sample is then layered on top of the gradient and spun at high speeds in an ultracentrifuge. This causes heavy macromolecules to migrate towards the bottom of the tube faster than lighter material. During centrifugation in the absence of sucrose, as particles move farther and farther from the center of rotation, they experience more and more centrifugal force (the further they move, the faster they move). The problem with this is that the useful separation range of within the vessel is restricted to a small observable window. Spinning a sample twice as long doesn't mean the particle of interest will go twice as far, in fact, it will go significantly further. However, when the proteins are moving through a sucrose gradient, they encounter liquid of increasing density and viscosity. A properly designed sucrose gradient will counteract the increasing centrifugal force so the particles move in close proportion to the time they have been in the centrifugal field. Samples separated by these gradients are referred to as "rate zonal" centrifugations. After separating the protein/particles, the gradient is then fractionated and collected.

Chromatographic methods

Chromatographic equipment. Here set up for a size exclusion chromatography. The buffer is pumped through the column (right) by a computer controlled device.

Usually a protein purification protocol contains one or more chromatographic steps. The basic procedure in chromatography is to flow the solution containing the protein through a column packed with various materials. Different proteins interact differently with the column material, and can thus be separated by the time required to pass the column, or the conditions required to elute the protein from the column. Usually proteins are detected as they are coming off the column by their absorbance at 280 nm. Many different chromatographic methods exist:

Size exclusion chromatography

Chromatography can be used to separate protein in solution or denaturing conditions by using porous gels. This technique is known as size exclusion chromatography. The principle is that smaller molecules have to traverse a larger volume in a porous matrix. Consequentially, proteins of a certain range in size will require a variable volume of eluent (solvent) before being collected at the other end of the column of gel.

In the context of protein purification, the eluant is usually pooled in different test tubes. All test tubes containing no measurable trace of the protein to purify are discarded. The remaining solution is thus made of the protein to purify and any other similarly-sized proteins.

Separation based on charge or hydrophobicity

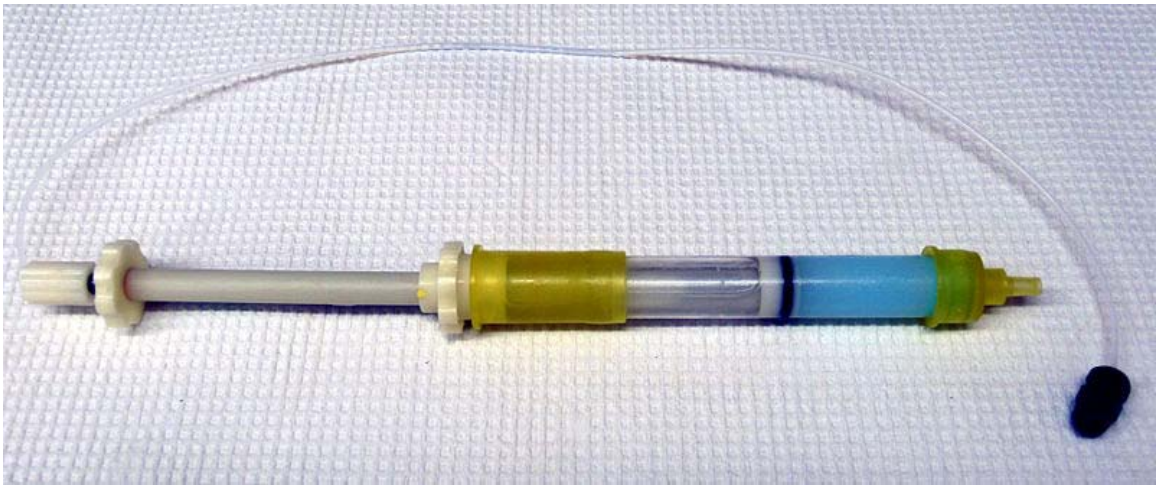
Hydrophobic Interaction Chromatography

Ion exchange chromatography

Ion exchange chromatography separates compounds according to the nature and degree of their ionic charge. The column to be used is selected according to its type and strength of charge. Anion exchange resins have a positive charge and are used to retain and separate negatively charged compounds, while cation exchange resins have a negative charge and are used to separate positively charged molecules.

Before the separation begins a buffer is pumped through the column to equilibrate the opposing charged ions. Upon injection of the sample, solute molecules will exchange with the buffer ions as each competes for the binding sites on the resin. The length of retention for each solute depends upon the strength of its charge. The most weakly charged compounds will elute first, followed by those with successively stronger charges. Because of the nature of the separating mechanism, pH, buffer type, buffer concentration, and temperature all play important roles in controlling the separation.

Ion exchange chromatography is a very powerful tool for use in protein purification and is frequently used in both analytical and preparative separations.



Nickel-affinity column. The resin is blue since it has bound nickel.

Affinity chromatography

Affinity Chromatography is a separation technique based upon molecular conformation, which frequently utilizes application specific resins. These resins have ligands attached to their surfaces which are specific for the compounds to be separated. Most frequently, these ligands function in a fashion similar to that of antibody-antigen interactions. This "lock and key" fit between the ligand and its target compound makes it highly specific, frequently generating a single peak, while all else in the sample is unretained.

Many membrane proteins are glycoproteins and can be purified by lectin affinity chromatography. Detergent-solubilized proteins can be allowed to bind to a chromatography resin that has been modified to have a covalently attached lectin. Proteins that do not bind to the lectin are washed away and then specifically bound glycoproteins can be eluted by adding a high concentration of a sugar that competes with the bound glycoproteins at the lectin binding site. Some lectins have high affinity binding to oligosaccharides of glycoproteins that is hard to compete with sugars, and bound glycoproteins need to be released by denaturing the lectin.

Metal binding

A common technique involves engineering a sequence of 6 to 8 histidines into the N- or C-terminal of the protein. The polyhistidine binds strongly to divalent metal ions such as nickel and cobalt. The protein can be passed through a column containing immobilized nickel ions, which binds the polyhistidine tag. All untagged proteins pass through the column. The protein can be eluted with imidazole, which competes with the polyhistidine tag for binding to the column, or by a decrease in pH (typically to 4.5), which decreases the affinity of the tag for the resin. While this procedure is generally used for the purification of recombinant proteins with an engineered affinity tag (such as a 6xHis tag or Clontech's HAT tag), it can also be used for natural proteins with an inherent affinity for divalent cations.

Immunoaffinity chromatography

A HPLC. From left to right: A pumping device generating a gradient of two different solvents, a steel enforced column and an apparatus for measuring the absorbance.

Immunoaffinity chromatography uses the specific binding of an antibody to the target protein to selectively purify the protein. The procedure involves immobilizing an antibody to a column material, which then selectively binds the protein, while everything else flows through. The protein can be eluted by changing the pH or the salinity. Because this method does not involve engineering in a tag, it can be used for proteins from natural sources.

Purification of a tagged protein

Adding a tag to the protein such as RuBPS gives the protein a binding affinity it would not otherwise have. Usually the recombinant protein is the only protein in the mixture with this affinity, which aids in separation. The most common tag is the Histidine-tag (His-tag), that has affinity towards nickel or cobalt ions. Thus by immobilizing nickel or cobalt ions on a resin, an affinity support that specifically binds to histidine-tagged proteins can be created. Since the protein is the only component with a His-tag, all other proteins will pass through the column, and leave the His-tagged protein bound to the resin. The protein is released from the column in a process called elution, which in this

case involves adding imidazole, to compete with the His-tags for nickel binding, as it has a ring structure similar to histidine. The protein of interest is now the major protein component in the eluted mixture, and can easily be separated from any minor unwanted contaminants by a second step of purification, such as size exclusion chromatography or RP-HPLC.

Another way to tag proteins is to engineer an antigen peptide tag onto the protein, and then purify the protein on a column or by incubating with a loose resin that is coated with an immobilized antibody. This particular procedure is known as immunoprecipitation. Immunoprecipitation is quite capable of generating an extremely specific interaction which usually results in binding only the desired protein. The purified tagged proteins can then easily be separated from the other proteins in solution and later eluted back into clean solution.

When the tags are not needed anymore, they can be cleaved off by a protease. This often involves engineering a protease cleavage site between the tag and the protein.

HPLC

High performance liquid chromatography or high pressure liquid chromatography is a form of chromatography applying high pressure to drive the solutes through the column faster. This means that the diffusion is limited and the resolution is improved. The most common form is "reversed phase" hplc, where the column material is hydrophobic. The proteins are eluted by a gradient of increasing amounts of an organic solvent, such as acetonitrile. The proteins elute according to their hydrophobicity. After purification by HPLC the protein is in a solution that only contains volatile compounds, and can easily be lyophilized. HPLC purification frequently results in denaturation of the purified proteins and is thus not applicable to proteins that do not spontaneously refold.

Concentration of the purified protein

A selectively permeable membrane can be mounted in a centrifuge tube. The buffer is forced through the membrane by centrifugation, leaving the protein in the upper chamber.

At the end of a protein purification, the protein often has to be concentrated. Different methods exist.

Lyophilization

If the solution doesn't contain any other soluble component than the protein in question the protein can be lyophilized (dried). This is commonly done after an HPLC run. This simply removes all volatile component leaving the proteins behind.

Ultrafiltration

Ultrafiltration concentrates a protein solution using selective permeable membranes. The function of the membrane is to let the water and small molecules pass through while retaining the protein. The solution is forced against the membrane by mechanical pump or gas pressure or centrifugation.

Analytical

Denaturing-Condition Electrophoresis

Gel electrophoresis is a common laboratory technique that can be used both as preparative and analytical method. The principle of electrophoresis relies on the movement of a charged ion in an electric field. In practice, the proteins are denatured in a solution containing a detergent (SDS). In these conditions, the proteins are unfolded and coated with negatively charged detergent molecules. The proteins in SDS-PAGE are separated on the sole basis of their size.

In analytical methods, the protein migrate as bands based on size. Each band can be detected using stains such as Coomassie blue dye or silver stain. Preparative methods to purify large amounts of protein, require the extraction of the protein from the electrophoretic gel. This extraction may involve excision of the gel containing a band, or eluting the band directly off the gel as it runs off the end of the gel.

In the context of a purification strategy, denaturing condition electrophoresis provides an improved resolution over size exclusion chromatography, but does not scale to large quantity of proteins in a sample as well as the late chromatography columns.

Non-Denaturing-Condition Electrophoresis

An important non-denaturing electrophoretic procedure for isolating bioactive metalloproteins in complex protein mixtures is termed 'quantitative native continuous polyacrylamide gel electrophoresis (QPNC-PAGE).