Protein Structure and Function

Carsten Müssig

eLearning in Studienbereichen – MNF – Universität Potsdam

Content

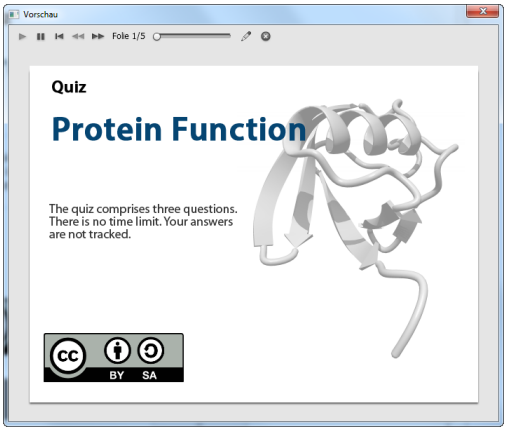
Letzte Überarbeitung: **Oktober 2014**.

I. Protein structure

Proteins are assembled from a set of 20 amino acids. Amino acids are linked by peptide bonds to form polypeptide chains. The secondary structure is defined as the local conformation of the peptide backbone. The tertiary structure of a protein is its three-dimensional arrangement. Proteins containing more than one polypeptide chain exhibit a fourth level of structural organization.

Quiz

Start this chapter with the quiz to test for your mastery. You may reattempt the quiz later to check your progress.



Quiz 1. Protein structure. Click to play - HTML5 or Flash (opens new window in your standard web browser).

Amino acids

Thousands of different proteins make up a very large fraction of the "machinery" of a cell. No matter what biological process we consider, a group of special proteins is required.

Proteins are assembled from a set of **20 amino acids**. The amino acid units that make up a protein molecule are joined in a precise sequence when the protein is made on a ribosome. The chain is then folded, sometimes is then cut in specific places.

A metal ion, a coenzyme, a single methyl group, or another chemical group may be attached to form the biologically active protein.

The functional properties of proteins are determined by their three-dimensional structures. The three-dimensional structure of a protein is dictated entirely by the sequence of amino acids.

The ability of proteins to self-assemble into complex structures is responsible for their dominant role in biochemistry.

All amino acids are known as **α-amino acids** because, with the exception of proline, they have a primary amino group and a carboxylic acid group substituent on the same carbon atom (Figure 1).

Most amino acids also contain different **functional groups** as part of the side chain. These functional groups include alcohols, thiols, thioethers, carboxylic acids, carboxamides, and basic groups. The multiple side chains of amino acids can undergo chemical reactions. When combined in various sequences, this array of functional groups accounts for the broad spectrum of protein function.

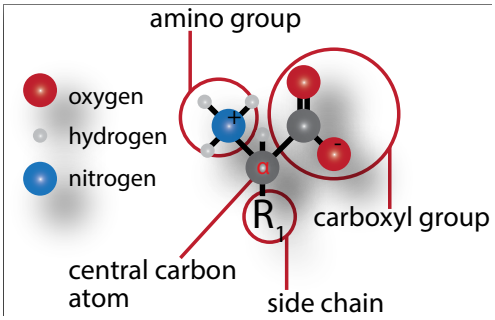


Figure 1. A α-amino acid consists of a central carbon atom, called the α carbon, linked to an amino group, a carboxylic acid group, a hydrogen atom, and a distinctive R group. The R group is often referred to as the side chain. The zwitterionic form occurs at physiological pH values.

For more details on amino acids see:

Khanacademy (2014) Amino acid structure

Wikibooks (2013) Principles of Biochemistry/Amino acids and proteins

Amino acid classification

The 20 amino acids vary in size, shape, hydrogen-bonding capacity, hydrophobic character, and chemical reactivity.

The most common and perhaps the most useful way of classifying the 20 amino acids is according to the **polarities of their side chains** (R groups).

This is because proteins fold to their native conformation largely in response to the tendency to remove and to solvate their hydrophilic side chains.

According to this classification scheme, there are three major types of amino acids:

1. Those with nonpolar hydrophobic R groups.
2. Those with uncharged polar R groups.
3. Those with charged polar R groups.

Nonpolar hydrophobic side chains

The **nonpolar amino acid** side chains have a variety of shapes and sizes. Eight amino acids are classified as having nonpolar side chains according to this scheme.

Sometimes also **glycine** is regarded as a nonpolar uncharged amino acid.

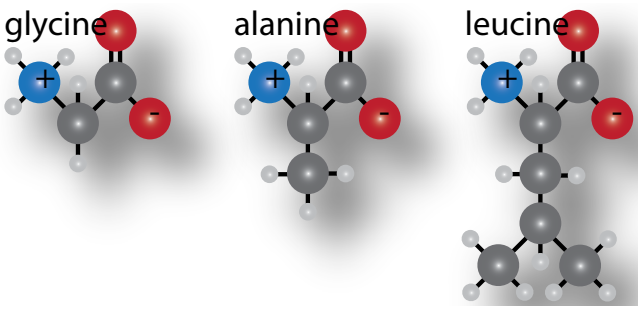
- Glycine represents a link in the peptide chain, because it provides a minimum of steric hindrance to rotation and to placement of adjacent groups.
- Glycine is the unique amino acid that is not optically active.

Alanine, valine, leucine, and isoleucine have aliphatic hydrocarbon side chains ranging in size from a methyl group for alanine to isomeric butyl groups for leucine and isoleucine.

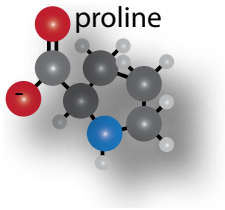
Aliphatic compounds are organic compounds in which carbon atoms are joined together in straight or branched chains or in rings, that can either be saturated or unsaturated, but not aromatic.

Alanine is rather simple, and relatively small. The methyl group of alanin is *non-reactive* and is thus almost never directly involved in protein function.

The hydrocarbon side chains of **valine**, **leucine**, and **isoleucine** are larger.

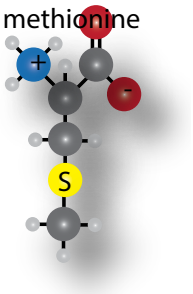


- The larger aliphatic side chains are **hydrophobic** – that is, they tend to **cluster together rather than contact water**.
- The three-dimensional structures of water-soluble proteins are stabilized by this tendency of hydrophobic groups to come together, the hydrophobic effect (see Protein Folding).



Proline is a cyclic secondary amino acid, it is unique among the standard 20 amino acids.

- The aliphatic side chain of proline is bonded to both the nitrogen and the α-carbon atoms.
- Proline **markedly influences protein architecture** because its ring structure makes it more conformationally restricted than the other amino acids.



Methionine has a **thiol ether** sidechain, which is **rather reactive**. Methionine is classified as **nonpolar**.

- Together with cysteine (the side chain of which is classified as uncharged and polar) methionine is one of two sulphur-containing amino acids.
- Methionine is encoded by a single codon (**AUG**) in the genetic code.
- The codon AUG is also significant, in that it carries the "**START**" message for a ribosome to begin protein translation from mRNA. As a consequence, methi-

onine is incorporated into the N-terminal position of proteins.

Three amino acids have **simple aromatic side chains**, phenylalanine, tryptophan, and tyrosine.

Phenylalanine, as its name indicates, contains a **phenyl ring** attached in place of one of the hydrogens of alanine.

- Phenylalanine is **purely hydrophobic**.

Tryptophan has an **indole ring** joined to a methylene (-CH_2-) group.

- Phenylalanine and tryptophan are bulky as well as nonpolar.

Uncharged polar side chains

Six amino acids are commonly classified as having uncharged polar side chains.

Serine and **threonine** bear **hydroxylic R** groups of different sizes.

- The aliphatic side chains contain hydroxyl groups, these amino acids could be termed **alcohols**.
- The hydroxyl groups of serine and threonine make them much more hydrophilic and reactive than alanine and valine.
- The side chains can undergo O-linked glycosylation.

Cysteine is structurally similar to serine but contains a sulfhydryl, or thiol (**-SH**), group in place of the hydroxyl (-OH) group.

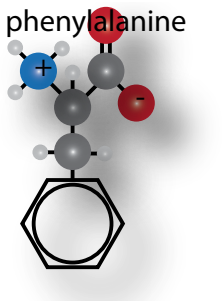
- The sulfhydryl group is much more reactive.
- Pairs of sulhydryl groups may come together to form **disulfide bonds**, which are particularly important in stabilizing some proteins.

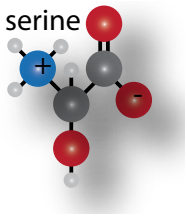
Asparagine and **glutamine** have **amide-bearing side chains** of different sizes.

- Asparagine and glutamine are uncharged derivatives of aspartate and glutamate.
- The amide group (-$CONH_2$) is relatively inert chemically.
- It is polar and participates in hydrogen bonding.

Tyrosine (4-hydroxyphenylalanine) has a **phenolic group**.

- The phenolic group together with the aromatic groups of phenylalanine and tryptophane, accounts for most of the UV absorbance and fluorescence exhibited by proteins.
- The hydroxyl group adds polarity to the side group.

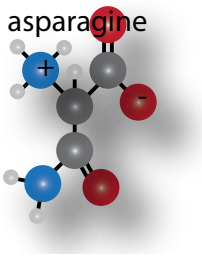




Charged polar side chains

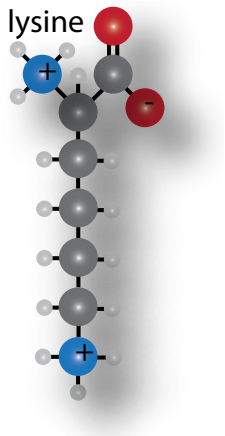
Five amino acids have charged polar side chains, which may be positively or negatively charged.

The **basic amino acids** are **positively charged** at physiological pH values because they accept a positively charged proton; they are lysine, arginine, and histidine. Lysine and arginine have relatively long side chains that terminate with groups that are positively charged at neutral pH.



Lysine has a flexible side arm with a potentially reactive amino group at the end.

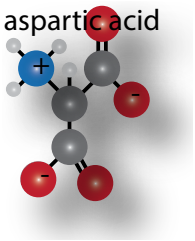
- The amino-group of the side chain of lysine is subject to posttranslational modification (including methylation, giving methyl-, dimethyl-, and trimethyllysine, and acetylation).
- This is part of the histone code, an important epigenetic mechanism to control gene expression.



The amino-group of the side chain of **arginine** is subject to methylation.

Histidine is the third basic amino acid.

- The imidazole group in the histidin side chain is part of the active site of many enzymes.
- Like other basic groups in proteins they also may bind metal ions.



The acidic amino acids, **aspartic acids** and **glutamic acid**, are negatively charged above pH 3.

- The carboxyl groups of these side chains are dissociated at neutral pH and provide **anionic groups** on the surface of proteins.
- In their ionized state, they are often referred to as **aspartate** and **glutamate**.

Primary structure

Proteins are linear polymers formed by linking the α-carboxyl group of an amino acid to the α-amino group of another amino acid with a **peptide bond** (also called an amide bond).

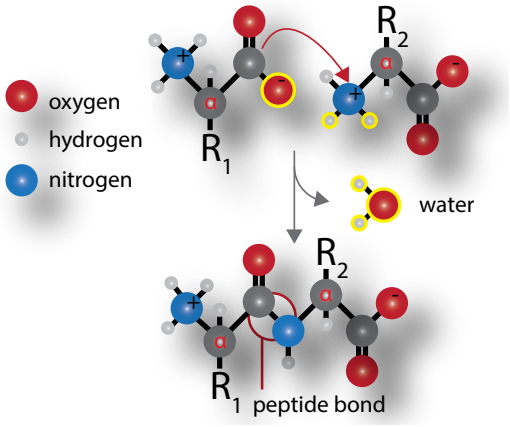


Figure 2. Amino acids are linked by peptide bonds. A covalent peptide bond forms when the carbon of the carboxyl group shares electrons with the nitrogen from the amino group of a second amino acid. A molecule of water is generated.

The formation of a dipeptide from two amino acids is accompanied by the loss of a water molecule (Figure 2). The equilibrium of this reaction lies on the side of hydrolysis rather than synthesis. Hence, the biosynthesis of peptide bonds requires an input of free energy. Nonetheless, peptide bonds are stable kinetically.

A series of amino acids joined by peptide bonds form a **polypeptide chain**, and each amino acid unit in a polypeptide is called a residue (Figure 3). **Peptides** made of small numbers of amino acids are called oligopeptides or simply peptides. Peptides and polypeptide chains have polarity because the ends are different, with a α-amino group at one end and a α-carboxyl group at the other. The amino end is taken to be the beginning of a polypeptide chain.

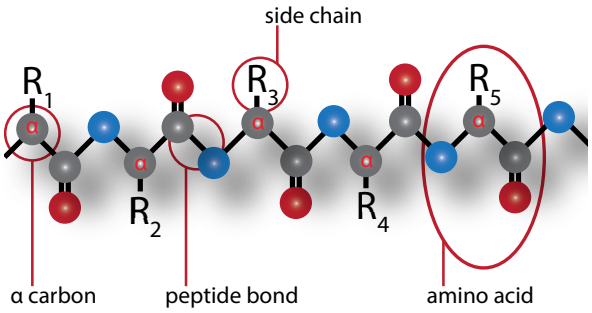


Figure 3. The primary structure is the sequence of a chain of amino acids. The amino group of one amino acid is linked to the carbon atom of the carboxyl group of the next amino acid. The α-carbon atom of each amino acid is the central atom that is linked to the amino group the side chain, and the carboxyl group. The only exception is proline.

Protein mass unit

We can refer to the mass of a protein, which is expressed in units of Daltons. One Dalton is equal to one unified atomic mass unit. One unified mass unit is approximately the mass of a single proton or a single neutron.

It is defined as one twelfth of the mass on an carbon-12 atom. One hydrogen atom has a mass of approximately 1 Da. The mean molecular weight of an **amino acid** residue is about **110 Da**.

Unit	Explanation
1 Da	alternate name for the atomic mass unit $\approx 1.661 \times 10^{-27}$ kg
1 kDa	1,000 Da
100 pmol of 1 kDa protein	100 ng
100 pmol of 100 kDa protein	10 µg

The peptide group

Each polypeptide consists of a backbone that supports the different amino acid

side chains. The polypeptide backbone is made from the repeating sequence of the core atoms of the amino acids. The side chains give each amino acid its unique properties.

Long polypeptide chains are flexible. Many of the covalent bonds allow free rotation of the atoms they join. Thus, proteins can in principle fold in many ways. However, several properties are responsible for the fact that polypeptide chains are conformationally restricted and, ultimately, fold in a very specific way (i.e. achieve the characteristic secondary and tertiary structure).

These factors are noncovalent bonds (hydrogen bonds, electrostatic attractions, van der Waals attractions) and the hydrophobic effect.

In addition, the conformation of a polypeptide chain is restricted because the peptide bond is essentially **planar**. For a pair of amino acids linked by a peptide bond, six atoms lie in the same plane: the α-carbon atom and CO group from the first amino acid and the NH group and α-carbon atom from the second amino acid. Thus, the peptide bond has a rigid, planar structure (Figure 4).

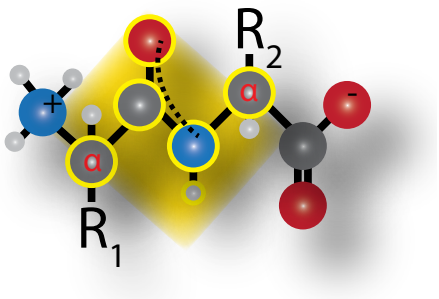


Figure 4. Peptide bonds are planar. Six atoms (C_α, C, O, N, H, and C_α) lie in a plane.

The planar structure is a consequence of **resonance interaction** that gives the peptide bond a partial double-bond character (Figure 5).

The C-N bond has partial double-bond character and, therefore, is shorter than that of a normal single bond. The C=O bond is lengthened.

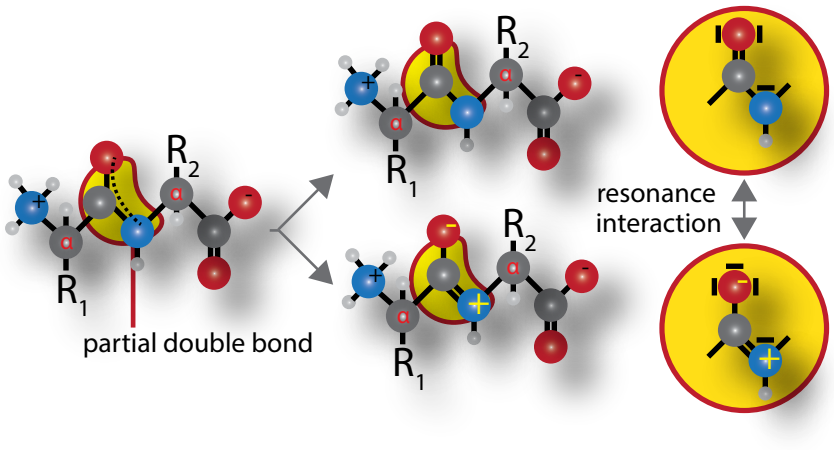
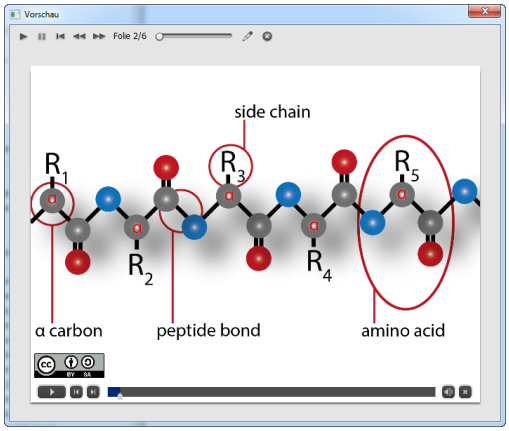


Figure 5. Peptide bond resonance structures. The peptide bond has double-bond character. This prevents rotation rotation about this bond and constrains the conformation of the polypeptide backbone.

The partial double-bond character of the peptide bond results in the inability of the bond to rotate, and this constrains the conformation of the peptide backbone and accounts for the bond's planarity (Video 1).



Video 1. Peptide Bond. Click to play - HTML5 or Flash.

cis and *trans* conformation

Two conformations are possible for a planar peptide bond, the ***cis*** and ***trans*** conformation (Figure 6).

In the *cis* configuration, the two α-carbon atoms are on the same side of the peptide bond. Peptide groups, with few exceptions, assume the *trans* conformation: that in which successive C_α atoms

are on opposite sides of the peptide bond. The *cis* conformation is less stable than the *trans* conformation. This preference for *trans* over *cis* can be explained by the fact that steric clashes between groups attached to the α-carbon atoms hinder formation of the *cis* form. By far the most common *cis* peptide bonds are X-Pro linkages.

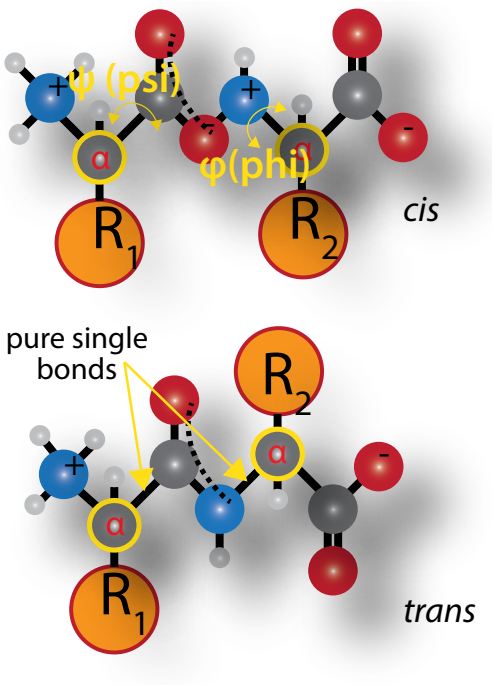


Figure 6. *Trans* and *cis* peptide bonds. The bond between the α-carbon and the nitrogen atom and the bon between the α-carbon and carbonyl carbon atom have pure single bond character. Both bonds allow rotation. The angle of roation is termed psi (ψ) and phi (φ), respectively. Ramachandran diagrams show possible values of phi and psi in a polypeptide backbone.

Conformation of polypeptides

The ability of proteins to fold into well-defined structures is remarkable thermodynamically. The favourable entropy associated with the large number of conformations in the unfolded form opposes folding and must be overcome by interactions favouring the folded form.

In contrast with the peptide bond, the bonds between the amino group and the α-carbon atom (indicated by **phi**) and between the α-carbon atom and the carbonyl group (indicated by **psi**) are pure single bonds.

The two adjacent rigid peptide units may rotate about these bonds, taking on various orientations. This freedom of rotation about two bonds of each amino acid allows proteins to fold in many different ways.

Since phi and psi can vary for each residue in a protein, there are a large number of possible conformations. However, the angle of rotation is restricted due to steric collisions. Certain atoms are brought into collision (two atoms cannot be in the same place at the same time).

Steric exclusion, the fact that two atoms cannot be in the same place at the same time, can be a powerful organizing principle. Sterically forbidden conformations are those in which any nonbonding interatomic distance is less than its corresponding van der Waals distance.

Such information is summarized in a conformation map or **Ramachandran diagram**. The Ramachandran diagram gives allowed values for the rotation angles about the C_α-N and the C_α-C bond.

The rigidity of the peptide unit and the restricted set of allowed rotations of the other bonds of the backbone limit the number of structures accessible to the unfolded form sufficiently to allow protein folding to occur.

For further details on protein conformation and stability see:

Khanacademy (2014) Conformational stability: protein folding and denaturation

Khanacademy (2014) Four levels of protein structure

Protein Data Bank (2014) Looking at structures: methods for determining atomic structures

Scitable by Nature Education (2014) Protein structure

Secondary structure

The secondary structure is defined as the **local conformation of the peptide backbone**. The local conformations are the basic structural motifs of proteins. Proteins fold into regularly repeating structures.

Two periodic structures are called the α helix and the β sheet. Other structures such as the β turn and omega loop were identified. These turn or loop structures contribute with α-helices and β sheets to form the final protein structure. The distribution of helices, β sheets, and turns along a protein chain is referred to as its secondary structure.

The Alpha helix

In 1951, Linus Pauling and Robert Corey proposed two periodic structures called the α helix and the β sheet. The α helix is a rodlike structure. A tightly coiled backbone forms the inner part of the rod and the side chains extend outward in a helical array.

The α helix is stabilized by hydrogen bonds between the NH and CO groups of the main chain.

- All main chain CO and NH groups are hydrogen bonded.
- Each residue is related to the next by a rotation of 100 degrees, which gives 3.6 amino acids per turn of helix.
- The hydrogen bonds of an α helix are arranged such that the peptide C=O bonds of the *n*th residue points along the helix towards the peptide N-H group of the (*n* + 4)th residue.
- In addition, the core of the α helix is tightly packed; that is, its atoms are in van der Waals contact across the helix, thereby maximising their association energies.
- The R groups all project outwards from the helix so as to avoid steric interference with the polypeptide backbone and with each other.
- The α-helical content of proteins ranges widely, from nearly none to almost 100%.

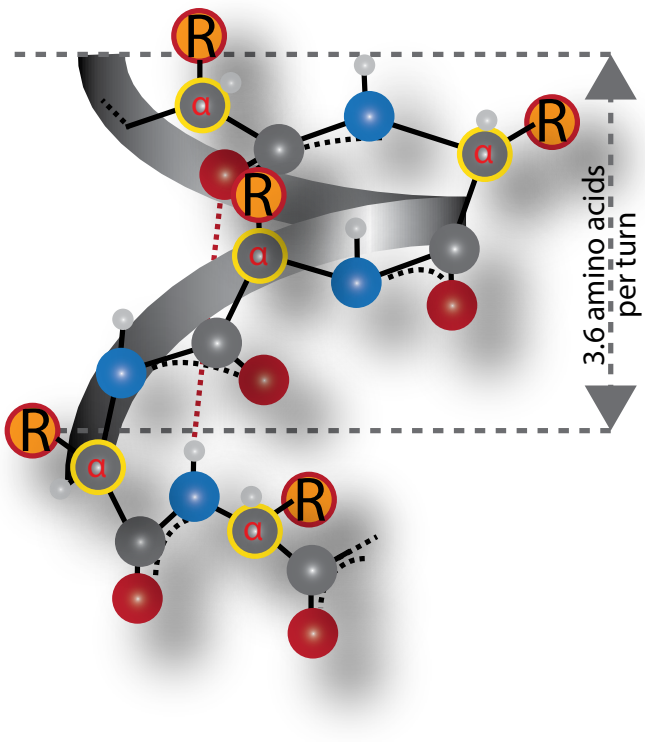


Figure 7. Structure of the α helix. The CO group of residue *n* forms a hydrogen bond with the NH group of residue *n* + 4. The number of residues per turn of the helix is 3.6.

The α helices can be particularly stable structures in hydrophobic environments such as the plasma membrane. Examples include the seven-transmembrane-helix (**7TM**) receptors, which are responsible for transmitting information initiated by signals as diverse as photons, hormones, and neurotransmitters.

Two or three α helices can wrap around one another to form a stable structure called **coiled-coil**.

This structure forms when α helices have nonpolar side chains on one side, so that they can twist around each other with these side chains facing inward. Examples include α-keratin (a component of the outer layer of the skin) and myosin (involved in muscle contraction).

Beta sheets

Pauling and Corey discovered another periodic structural motif, which they named the β sheet (β because it was the second structure that they elucidated, the α helix having been the first).

The β sheet is an important structural element in many proteins. The β sheet differs markedly from the rodlike α helix. The β sheet is composed of two or more polypeptide chains called **β strand**.

The β strands in a β sheet are almost fully extended rather than being tightly coiled as in the α helix. In β sheets, hydrogen bonding occurs between neighboring polypeptide chains rather than within one as in α helices.

- A β sheet is formed by linking two or more β strands by hydrogen bonds.
- Adjacent chains in a β sheet can run in opposite (**antiparallel β sheet**) (Figure 8), or in the same direction (**parallel β sheet**).

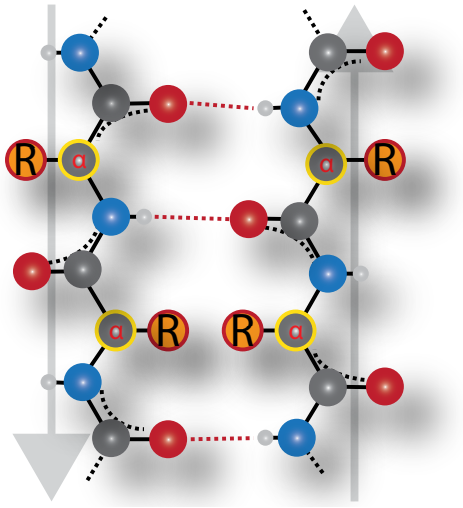


Figure 8. Structure of an antiparallel β sheet. Adjacent β strands run in opposite directions. Hydrogen bonds between NH and CO groups stabilize the structure.

- In the antiparallel β sheet the neighboring hydrogen bonded polypeptide chains run in opposite directions. The NH group and the CO group of each amino acid are respectively hydrogen-bonded to the CO group and the NH group of a partner on the adjacent chain.
- In the parallel β sheet the hydrogen bonded chains extend in the same direction. For each amino acid, the NH group is hydrogen bonded to the CO group of one amino acid on the adjacent strand, whereas the CO group is hydrogen bonded to the NH group on the amino acid two residues farther along the chain.
- The β sheets of proteins can be purely antiparallel, purely parallel, or mixed.

Nonrepetitive structures

Regular secondary structures – helices and β sheets – comprise around half of the average globular protein. To form a globular protein, a polypeptide chain must repeatedly fold back on itself.

The turns or bends by which this is accomplished can be regarded as a third major secondary structural element.

Reverse turns

Globular proteins consist largely of runs of secondary structures joined by stretches of polypeptide that abruptly change direction. Many of these reversals are accomplished by a common structural element called the reverse turn (also known as the **β turn** or **hairpin bend**).

- These turns often connect successive strands of antiparallel β sheets and almost always occur at protein surfaces.
- In many reverse turns, the CO group of residue *n* of a polypeptide is hydrogen bonded to the NH group of residue *n* + 3. This interaction stabilizes abrupt changes in direction of the polypeptide chain.

Omega loops

The omega loops (Ω loops) are more elaborate structures. It is named after its shape, which resembles the Greek capital letter Omega. Almost all proteins contain one or more loops of six to 16 residues that are not components of helices or β sheets. Ω loops are located on the protein surface.

Tertiary structure

The tertiary structure of a protein is its three-dimensional arrangement; that is, the folding of its secondary structural elements, together with the spatial dispositions of its side chains.

The secondary structure elements pack together in part to bury the hydrophobic side chains. The hydrophobic core is charactericed by a high density of atoms. However, packing is not perfect. There are small cavities and small channels that provide room for atoms to move, allowing conformational changes.

The first three-dimensional X-ray structure of a protein, that of whale **myoglobin**, was elucidated in the late 1950s (Figure 9).

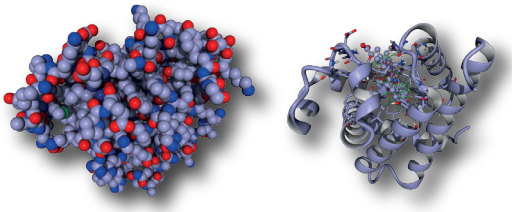


Figure 9. Whale myoglobin. Myoglobin stores oxygen in cells. The space-filling model (left) provides a contour map of the protein and shows which amino acids are exposed to the surface. The ribbon model (right) represents an easy way to visualize sedondary structures such as α helices and β sheets (drawn from 1MBN.pdb using YASARA; Watson, 1969).

For more details on the tertiary structure of proteins see:

J.M. Berg, J.L. Tymoczko, L. Stryer (2002) Examination of Three-Dimensional Structure Enhances Our Understanding of Evolutionary Relationships. In: Biochemistry, 5. edition. W.H. Freeman.

A. Kropinski (2012) Tertiary structure predictions. In: The Bio-Web.

Quaternary structure

Proteins containing more than one polypeptide chain exhibit a fourth level of structural organization. Each polypeptide chain in such a protein is called a **subunit**.

The term 'quaternary structure' refers to the spatial arrangement of subunits and the nature of their interactions. The simplest sort of quaternary structure is a **dimer**, consisting of two identical subunits.

More complicated quaternary structures also are common. More than one type of subunit can be present, often in variable numbers.

An example is human haemoglobin, the oxygen-carrying protein in blood, consists of two subunits of one type (designated α) and two subunits of another type (designated β). Thus, the haemoglobin molecule exists as a $\alpha_2\beta_2$ tetramer (Figure 10).

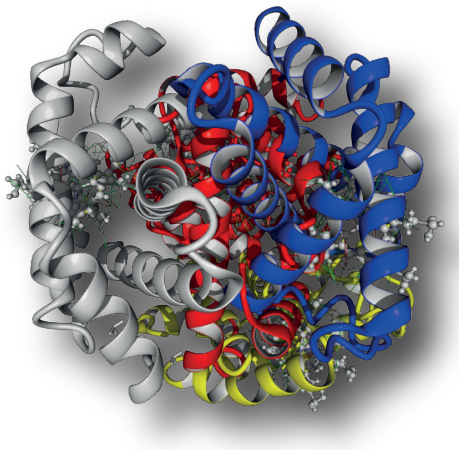


Figure 10. Human haemoglobin. Haemoglobin transports oxygen in the blood. The ribbon model represents two alpha chains (141 amino acids length, red and blue) and two beta chains (146 amino acids, yellow and grey; drawn from 4HHB.pdb using YASARA; Fermi et al., 1984).

Protein Folding

Proteins fold into a single stable conformation. The final folded structure, or conformation, is the shape in which the free energy (*G*) is minimized.

Protein folding and unfolding is an "all or none" process that results from a cooperative transition. The essence of protein folding is the tendency to retain partly correct intermediates.

Protein folding is a highly cooperative process

Proteins fold into a single stable conformation. Each protein has a particular three-dimensional structure, which is determined by the order of amino acids in its chain.

Stability is defined as a net loss of free energy, a function of the combined effects of entropy and enthalpy. The final folded structure, or conformation, is the shape in which the free energy (*G*) is minimized.

The amino acid sequences have evolved to guarantee that the polypeptide will adopt a stable conformation. The change of single amino acids (missense mutations) can sometimes disrupt the structure of a protein.

Protein folding and unfolding is an "**all or none**" process that results from a cooperative transition. Conditions that lead to the disruptions of any part of a protein structure are likely to unravel the protein completely. Structures that are partly intact and partly disrupted are not thermodynamically stable and exist only transiently.

Cooperative folding ensures that partly folded structures that might interfere with processes within cells do not accumulate. Proteins do not fold by testing every possible conformation. The essence of protein folding is the tendency to retain partly correct intermediates.

Interactions that lead to cooperative folding can stabilize intermediates as structure builds up. Local regions tend to adopt their favoured structures and can interact with on other, leading to increasing stabilization (**nucleation-condensation model**).

Protein domains and sequence motifs

Larger proteins (> 20 kDa) usually fold into two or more independent structural parts, termed **domains**. For example, the homeodomain binds DNA and is a structural feature of several transcription factors.

A domain is a compact region of 50 to 200 amino acids that often is stable enough to exist on its own in aqueous solution. Domains are made up of secondary structure elements that form a tertiary structure. Domains are stabilized by hydrophobic cores. Hydrophobic groups are concentrated in the center of the domain, and hydrophilic groups are exposed to water.

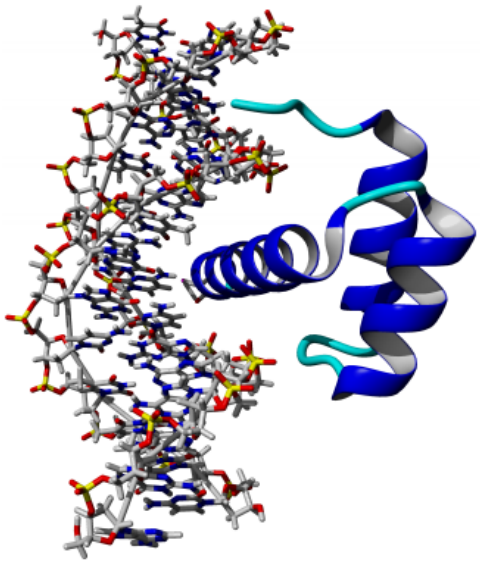


Figure 11. Homeodomain-DNA complex. Homeodomains are eukaryotic DNA-binding motifs. For example, the Antennapedia protein (≈ 42 kDa) in Drosophila is a transcription factors that controls the formation of legs during development. The homeodomain is the DNA-binding domain of the Antp protein. It is 62 amino acid in length and shown here in complex with DNA (drawn from 9ANT.pdb using YASARA; Fraenkel and Pabo, 1998).

Larger protein structures are modular. Two or more domains are fused together. Each domain has its specific function, and the function of the entire protein is determined by the sum of the properties of the domains. Thus, the structure and function of protein families are defined on the basis of the domains they contain.

The modular nature of proteins allows for sequence insertions and deletions. Protein evolution is partly based on shuffling of domains to produce new proteins.

Sequence motifs represent a smaller structure with functional significance.

Fom example, zinc finger motifs are found in transcription factors and bind to DNA. The zinc finger motif comprises characteristically Cys and His residues which are connected by a loop.

Many other sequence motifs are discontinuous and difficult to detect.

For more details on protein domains and sequence motifs see:

S. Hunter, P. Jones, A. Mitchell et al. (2011) InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res. 40: D306-D312

I. Letunic, T. Doerks and P. Bork (2012) SMART7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 40: D302-D305

C.J.A. Sigrist, E. de Castro, L. Cerutti, B.A. Cuche, N. Hulo, A. Bridge, L. Bougueleret and I. Xenarios (2013) New and continuing developments at PROSITE. Nucleic Acids Res. 41: D344-D347

Denaturation

Surprisingly, native proteins are only **marginally stable entities** under physiological conditions. The free energy required to denature them is ~0.4 kJ mol^{-1} of amino acid residues so that 100-residue proteins are typically stable by only around 40 kJ mol^{-1} (the energy required to break a typical hydrogen bond is ~20 kJ mol^{-1}).

The low conformational stabilities of native proteins make them easily susceptible to **denaturation** by altering the balance of the weak nonbonding forces that maintain the native conformation.

Proteins can be denatured by high temperatures, denaturants such as urea, and detergents such as SDS.

Weak chemical bonds

Electrostatic forces do not greatly stabilize proteins. Free ions in aqueous solution are highly solvated. The free energy of solvation of two separated ions is about equal to the free energy for formation of their unsolvated ion pairs.

Hydrogen bonding has a major influence on the structures of proteins, but **internal hydrogen bonding** does not significantly stabilize the structure of a native protein relative to its unfolded state.

Hydrogen bonds are much more directional than are van der Waals forces. The internal hydrogen bonding groups of a protein are arranged such that nearly all possible hydrogen bonds are formed. An unfolded protein, however, makes all its hydrogen bonds with the water molecules of the aqueous solvent. The free energy of stabilization that internal hydrogen bonds confer upon a native protein is equal to the difference in the free energy of hydrogen bonding between the native protein and the unfolded protein.

Van der Waals forces significantly stabilize protein structures.

The **hydrophobic effect** also is an important determinant of protein structures, because native proteins form a sort of intramolecular micelle in which the nonpolar side chains are largely out of contact with the aqueous solvent.

The transfer of a hydrocarbon from an aqueous medium to a nonpolar medium is entropically driven. The same is true of the transfer of a nonpolar protein group from an aqueous environment to the protein's nonpolar interior. The water molecules that were interacting with the polar groups of a polypeptide chain are freed to rejoin the structure of liquid water.

Disulfide bonds

Disulfide bonds form as a protein folds to its native conformation. Disulfide bonds in proteins are formed between the **thiol groups** of cysteine residues. Disulfide bonds stabilize protein structures.

Almost all proteins with disulfide bonds are secreted to the extracellular matrix. The oxidising extracellular conditions allow stable disulfide bonds. The energy required to break a disulfide bond is much larger than the energy required to break even a whole set of noncovalent bonds.

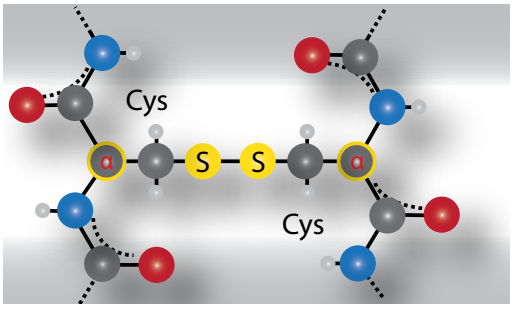


Figure 12. Proteins are cross-linked by disulfide bonds. Disulfide bonds form between cysteine residues. Disulfide bonds can be cleaved by agents such as β-mercaptoethanol that reduce the disulfide into sulhydryls (cysteines). Protein disulfide isomerases catalyze the formation of disulfide bonds.

Folding accessory proteins

Most unfolded proteins renature *in vitro* with low efficiency. *In vivo*, however, polypeptides efficiently fold to their native conformations as they are being synthesized, a process that normally requires only a few minutes.

This is because all cells contain three types of accessory proteins that assist polypeptides in folding to their native conformations.

Protein disulfide isomerase

PDI is an enzyme in the endoplasmic reticulum in eukaryotes or periplasmic space of prokaryotes that catalyzes the formation and breakage of disulfide bonds between cysteine residues within proteins as they fold.

PDI catalyzes **disulfide interchange reactions**, thereby facilitating the shuffling of the disulfide bonds in proteins until they achieve their native pairing.

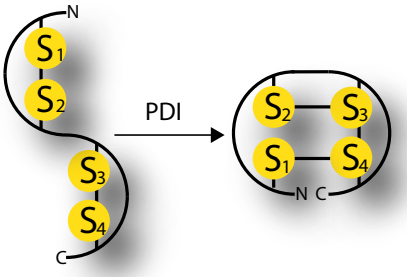


Figure 13. Rearrangement of disulfide bonds by protein disulfide isomerase (PDI).

Peptidyl prolyl cis-trans isomerases (PPIs)

Most peptide bonds are in the *trans* conformation. The reason is that steric interference causes the *cis* conformation to be less stable than the *trans* conformation (see page 6).

This energy difference is less in peptide bonds followed by a proline residue. Indeed, about 10 % of proline residues have a ***cis*** X-Pro peptide bond. This compares to less than 1% for peptide bonds overall.

The ***cis-trans*** isomerization has been shown to be the rate limiting step for the folding of many proteins such as collagen, carbonic anhydrase, ribonuclease and others. The isomerization is catalyzed by proteins known as prolyl isomerases, a large family of proteins.

Peptidyl prolyl *cis-trans* isomerases (PPIs) catalyze the otherwise slow **interconversion of X-Pro peptide bonds** between their *cis* and *trans* conformations, thereby accelerating the folding of Pro-con-

taining polypeptides.

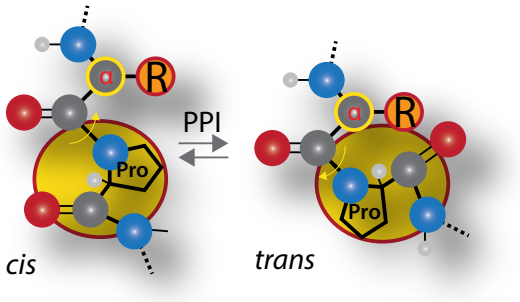


Figure 14. Peptidyl-prolyl isomerases catalyze the rotation of peptidyl-prolyl bonds in polypeptides.

Two structurally unrelated families of peptidyl prolyl cis-trans isomerases (PPIs) have been characterized: The **cyclophilins** and the family for which FK506 binding protein (**FKBP**) is prototypic.

Molecular chaperones

Unfolded proteins contain numerous solvent-exposed hydrophobic regions and therefore have a great tendency to form both intramolecular and intermolecular aggregates.

Molecular chaperons are proteins that function to prevent or reverse such improper associations. Molecular chaperones bind and stabilize proteins at intermediate stages of folding, assembly, translocation across membranes and degradation.

The molecular chaperones comprise several unrelated classes of proteins. Most **heat shock proteins** (HSPs) are molecular chaperones. HSP functions are required at normal temperatures, but the level of expression is reduced. HSPs have been classified by molecular weight (for example, Hsp70 for 70 kDa HSP). HSPs stabilize proteins prior to complete folding, during transport across membranes and for proteolysis. Several HSPs have ATPase activity. The mechanisms by which molecular chaperones carry out their functions are poorly understood.

Further Readings

References

B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter (2009) Essential Cell Biology, 3rd edition. Garland Science

J.M. Berg, J.L. Tymoczko, L. Stryer (2012) Biochemistry, 7th edition. W.H. Freeman

H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne (2000) The Protein Data Bank. Nucleic Acids Res. 28: 235-242

G. Fermi, M.F. Perutz, B. Shaanan and R. Fourme (1984) The crystal structure of human deoxyhaemoglobin and 1.74 A resolution. J. Mol. Biol. 175: 159-174

E. Fraenkel and C.O. Pabo (1998) Comparison of X-ray and NMR structures for the Antennapedia homeodomain-DNA complex. Nat. Struct. Biol. 5: 692-697

E. Krieger and G. Vriend (2014) YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics doi:10.1093/bioinformatics/btu426

D. Voet and J.G. Voet (2011) Biochemistry, 4th edition. Wiley

H.C. Watson (1969) The stereochemistry of the protein myoglobin. Prog. Stereochem. 4: 299

OER and OCW

J. King (2007) Protein folding problem. MIT OCW

StarBiochem (2014) A 3-D protein viewer that visually represents protein structural information based on the four different levels of protein structure. OEIT and MIT

OpenLearn (2014) Proteins. The Open University

II. How Proteins are Controlled

Most proteins do not work continuously in a cell. Instead their levels and activities are controlled in a coordinated fashion so that the cell can maintain itself in an optimal state (Figure 1).

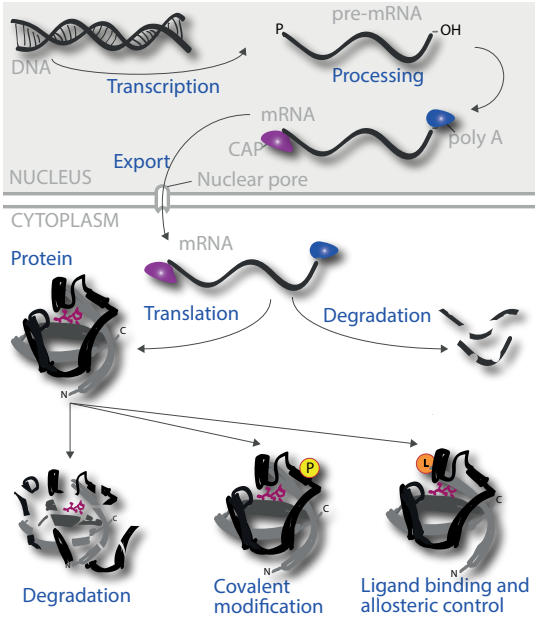
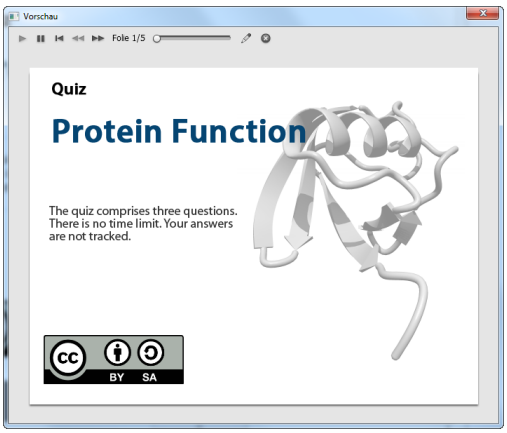


Figure 1. Levels and activity of proteins are controlled at several levels. Additional control modes are discussed in the text.

Quiz

Start this chapter with the quiz to test for your mastery. You may reattempt the quiz later to check your progress.



Quiz 1. Protein function. Click to play - HTML5 of Flash (opens new window in your standard web browser).

Control of Protein Levels

Biosynthesis

The synthesis of mRNA and synthesis of protein is controlled at several levels. This determines the **maximal protein level**.

Protein biosynthesis in a eukaryotic cell depends on numerous factors, including:

- mature mRNA
- transport of the mature mRNA into the cytoplasm
- proteins that bind to and stabilize the mRNA
- factors that mediate the formation of a functional ribosome
- active ribosomes

Thus, the biosynthesis of proteins is controlled at multiple levels.

Degradation

After a protein is released from the ribosome, it becomes subject to a number of controls by the cell. The breakdown of proteins into their amino acids is a way of regulating the amount of a particular protein.

Proteins vary in their **life span**. Structural proteins may last for years, whereas regulatory proteins may last only for seconds.

Cells possess specialized pathways to enzymatically break proteins down into the amino acids. This process is called **proteolysis** and mediated by various proteolytic pathways and enzymes called **proteases**.

- Proteases hydrolyze peptide bonds between amino acids.
- Proteolytic pathways degrade proteins whose lifetimes must be short and also eliminate proteins that are damaged or misfolded.
- Eliminating misfolded proteins is critical for cells. Alzheimer's, Creutzfeldt-Jacob, and other diseases are caused by

the aggregation of misfolded proteins. Proteins are usually tagged for selective destruction in proteasomes by attachment of **ubiquitin**.


ubiquitin

Ubiquitin is a small, highly conserved protein (see ribbon model drawn from 1UBQ.pdb; Vijay-Kumar et al., 1987). Yeast and human ubiquitin differ at only 3 of 76 residues. The carboxyl-terminal glycine residue of ubiquitin becomes covalently attached to the ε-amino groups of several lysine residues on a protein destined to be degraded.

The ATP-dependent joining of ubiquitin to a condemned protein is done by three enzymes (**E1**, **E2**, and **E3**) (Figure 2).

- Initially the terminal carboxyl group of ubiquitin is joined in a thioester bond to a cysteine residue on Ubiquitin-Activating Enzyme (E1). This reaction requires ATP.
- The ubiquitin is then transferred to a sulfhydryl group on an ubiquitin-conjugating enzyme (E2).
- An ubiquitin-protein ligase (E3) then promotes transfer of ubiquitin from E2 to a Lys residue of a protein recognized by that E3.
- More ubiquitins are added to form a chain of ubiquitins. A chain of four or more ubiquitins targets proteins for degradation in proteasomes (attachment of a single ubiquitin to a protein has other regulatory effects).

Most proteins are degraded by proteolytic complexes called **proteasomes** which are present both in the cytoplasm and in the nucleus.

Proteasomes mediate selective protein degradation. The 26S proteasome is a 2.5 MDa protein complex that comprises multiple polypeptides. Substrates are degraded in a processive manner. The ubiquitin is recycled. The released amino acids are left intact and used for biosynthesis or degraded.

E3 ubiquitin ligases

Interestingly, several thousand proteins are involved in the control of protein degradation in a eukaryotic cell. Eukaryotic cells have only a small number of distinct E1 enzymes, but many distinct E2 and E3 enzymes.

Substrate specificity is governed by the substrate-recruiting E3 ligases, which group into three classes on the basis of the E2-interacting region, either a Homology to E6-Associated Carboxy-Terminus (HECT), U-box or Really Interesting New Gene (RING) domain.

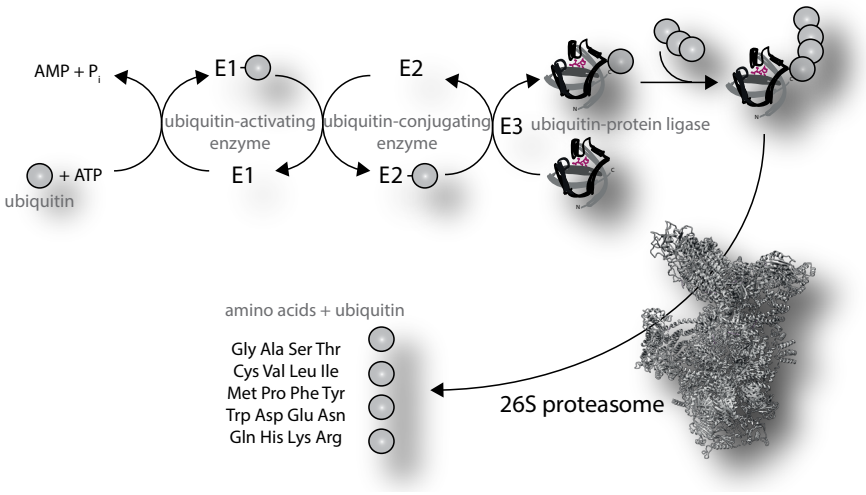


Figure 2. Proteins are tagged for selective destruction in proteolytic complexes called proteasomes by attachment of ubiquitin. The ribbon model of the yeast 26S proteasome was drawn from 4CR2.pdb using YASARA (Unverdorben et al., 2014).

The three distinct families of E3 proteins consist of hundreds of members. The diversity of target proteins that must be tagged for destruction requires a large number of E3 proteins as readers.

- The **RING finger** is a zinc-binding motif that binds to the E2 Ub-conjugating enzyme. RING finger proteins can function as single subunit E3s or participate as part of multisubunit E3 complexes.
- **U-box** E3 Ub ligases comprise a much smaller subfamily. The U-box motif was revealed to be a modified RING-finger domain. While structurally similar to the RING motif, the U-box does not use zinc

ions to stabilize its secondary structure.

- The smallest E3 subfamily is the **HECT domain** proteins. Unlike RING and U-box E3s that indirectly mediate Ub ligation by docking the E2 enzyme and substrate protein for Ub transfer from the E2 to the substrate, HECT E3s accept activated Ub from the E2 and then transfer it directly to the substrate protein.

Control of protein activity

Most proteins (e.g. enzymes, receptors, and transcription factors) can adopt two or more slightly different conformations, and by a shift from one to another, their activity or ability to interact with other factors is regulated.

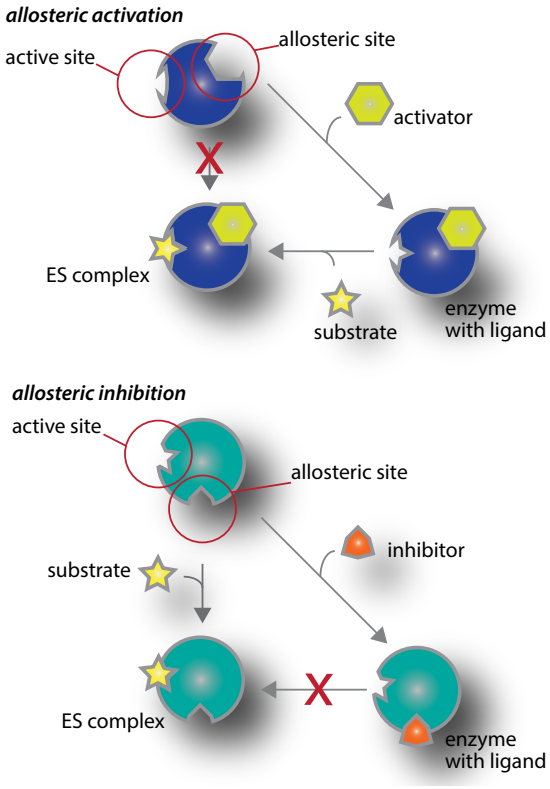


Figure 3. Allosteric activation and inhibition. Top: With an activator present, the protein adopts the active conformation. Bottom: With an inhibitor present, the protein adobts the inactive conformation.

Allosteric control

The catalytic activities of many enzymes vary in response to the concentrations of compounds other than their substrates. The catalytic activity of enzymes may be directly regulated through **conformational alterations**.

Conformational changes can be induced by the binding of specific molecules and ions (Figure 3). Similarly, the activity of receptors, transcription factors, and motor proteins usually is regulated by a shift from one conformation to another.

Any substance that is bound by a protein (e.g. an ion, a metabolite, another protein or a nucleic acid) is referred to as a **ligand** for that protein (Latin *ligare*, to bind).

Proteins bind to ligands by weak chemical interactions (i.e., hydrogen bonds, electrostatic attractions, van der Waals attractions, and hydrophobic effect). The region of a protein that associates with a ligand is known as **binding site**.

The binding site of **allosteric regulators** (regulatory small compounds) is different from the **active site** that binds the substrate. Both sites "communicate" via conformational changes. Binding of a ligand (e.g. a metabolite) stabilizes a specific conformation which is associated with a specific level of activity.

For example, binding of an allosteric inhibitor at one site on the protein causes the protein to shift to a conformation in which the active site has a lower substrate-binding affinity. Thus, enzymatic activity may be regulated by allosteric alteration of substrate-binding affinity.

Covalent modification

The activity of a protein (i.e. its conformation) can also be controlled by covalent modifications (Table 1).

A common reversible covalent modification is **phosphorylation** of tyrosine, serine, and threonine residues.

ATP serves as the phosphoryl donor. The

phosphorylation is catalysed by **protein kinases**, the removal of phosphoryl groups by hydrolysis is catalysed by **protein phosphatases**.

Further examples of covalent modifications include **acetylation** and **methylation** of lysine residues, and **farnelysation** and **palmitoylation** of cysteine residues (thioether linkages) (Figure 4).

Addition of a **single ubiquitin** molecule controls protein activity. In contrast, addition of several (at least four) ubiquitin molecules tags proteins for degradation by the 26S proteasome (see page 15).

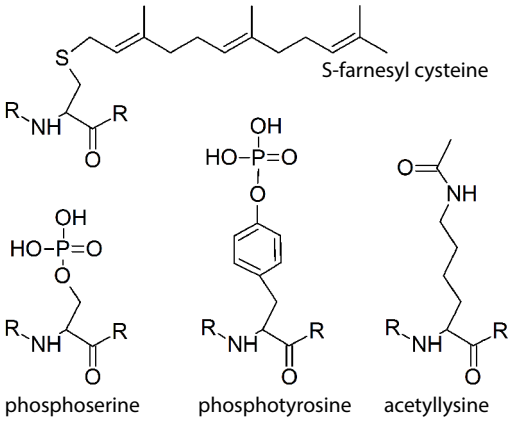


Figure 4. Covalent modifications of amino acids. For further modifications see Table 1 and web resources below.

Table 1. Posttranslational modifications.

Modification	Example
Phosphorylation	Glycogen phosphorylase
Acetylation	Histones
Methylation	Histones
Myristoylation, Palmitoylation (fatty acyl group)	Src, Ras
ADP ribolysation	RNA polymerase
Farnesylation (isoprenoid group)	Ras
Ubiquitination	Cyclin
O-linked-N-acetylglusamine	CaMKIV

For further details on post-translational modifications of proteins see:

G.A. Khoury, R.C. Baliban and C.A. Floudas (2011) Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci. Rep. 1: 90 (doi:10.1038/srep00090)

Thermo Scientific - Pierce Protein Biology Products (2014) Overview of Post-Translational modifications (PTMs): http://www.piercenet.com/method/overview-post-translational-modification

Proteolytic activation

Proenzymes are inactive precursor and are not catalytically active. Proenzymes (or **zymogens**) are irreversibly activated by hydrolysis of one or a few peptide bonds (Figure 5).

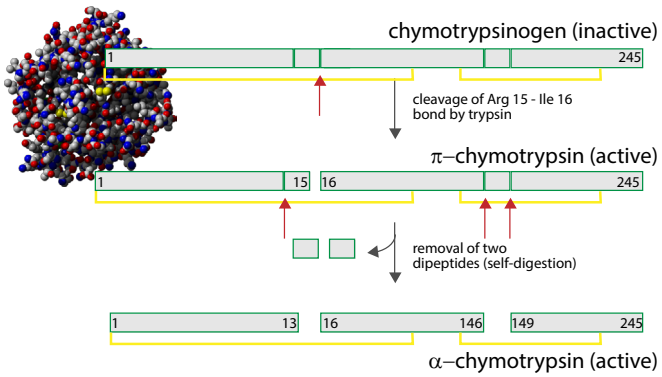


Figure 5. Proteolytic activation of chymotrypsinogen. The conformation of the inactive chymotrypsinogen (see 3D-model) is similar to the conformation of the active chymotrypsin, but differs in the position of specific amino acids that are essential for enzymatic activity. Disulfide bonds are highlighted in yellow (model drawn from 1EX3.pdb using YASARA; Pjura et al., 2000).

Localisation

Subcellular localization is a key functional characteristic of proteins.

To function in a specific physiological function (e.g., metabolic pathway, signalling cascade, structural associate etc.), proteins must be localized in the same cellular compartment.

Thus, the protein has be to transported to the correct cellular compartment, either in soluble form or attached to a membrane of transport vesicles.

For example, a DNA-binding transcription factor can only control gene expression in the nucleus. If a transcription factor is initially localized in the cytoplasm, binding of ligands or covalent modifications (e.g., phosphorylation) cause movement into the nucleus, where it binds to specific DNA sequences and controls gene activity.

The **subcellular localization** of proteins is predicted by homology with characterized proteins. Some proteins have sequence signals that determine their transport to organelles or the extracellular space. In addition, 3D structures and surface properties of proteins may provide targeting information.

Transmembrane helices of membrane proteins are predicted by means of hydrophobicity scales, prediction of helices, and amino acid preferences for membrane proteins. However, predictions often have low accuracy.

For a summary of protein subcellular prediction tools see:

J. Meinken and J. Min (2012) Subcellular Localization Prediction Tools

OMICtools (2014) Protein subcellular localization analysis (omic data)

Further Readings

References

B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter (2009) Essential Cell Biology, 3rd edition. Garland Science

J.M. Berg, J.L. Tymoczko, L. Stryer (2012) Biochemistry, 7th edition. W.H. Freeman

H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne (2000) The Protein Data Bank. Nucleic Acids Res. 28: 235-242

E. Krieger and G. Vriend (2014) YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics doi:10.1093/bioinformatics/btu426

P.E. Pjura, A.M. Lenhoff, S.A. Leonard and A.G. Gittis (2000) Protein crystalllization by design: chymotrypsinogen without precipitants. J. Mol. Biol. 300: 235-239

P. Unverdorben, F. Beck, P. Sledz, A. Schweitzer, G. Pfeifer, J.M. Plitzko, W. Baumeister, and F. Förster (2014) Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. Proc. Natl. Acad. Sci. USA 111: 5544-5549

S. Vijay, C.E. Bugg and W.J. Cook (1987) Structure of ubiquitin refined at 1.8 A resolution. J. Mol. Biol. 194: 531-544

D. Voet and J.G. Voet (2011) Biochemistry, 4th edition. Wiley

OER and OCW

E. Lander, R. Weinberg, T. Jacks, H. Sive, G. Walker, S. Chisholm, M. Mischke (2011) Fundamentals of Biology. MIT OCW

OpenLearn (2014) Proteins. The Open University