Orders of protein structure

Orders of protein structure: primary, secondary, tertiary, and quaternary. Alpha helix and beta pleated sheet.

Introduction

Have you ever wondered why egg whites go from clear to opaque when you fry an egg? If so, this section is for you!

Egg whites contain large amounts of proteins called albumins, and the albumins normally have a specific 3D shape, thanks to bonds formed between different amino acids in the protein. Heating causes these bonds to break and exposes hydrophobic (water-hating) amino acids usually kept on the inside of the protein[1, 2]. The hydrophobic amino acids, trying to get away from the water surrounding them in the egg white, will stick to one another, forming a protein network that gives the yolk structure and makes it white and opaque. Ta-da! Thank you, protein denaturation, for another delicious breakfast.

As we mentioned in the last article on proteins and amino acids, the shape of a protein is very important to its function. To understand how a protein gets its final shape or conformation, we need to understand the four levels of protein structure: primary, secondary, tertiary, and quaternary.

Primary structure

The simplest level of protein structure, **primary structure**, is simply the sequence of amino acids in a polypeptide chain. For example, the hormone insulin has two polypeptide chains, A and B, shown in diagram below. (The insulin molecule shown here is cow insulin, although its structure is similar to that of human insulin.) Each chain has its own set of amino acids, assembled in a particular order. For instance, the sequence of the A chain starts with glycine at the N-terminus and ends with asparagine at the C-terminus, and is different from the sequence of the B chain.

You may notice that the insulin chains are linked together by sulphur-containing bonds between cysteine's. Don't let yourself get confused: although these bonds are an important part of insulin's overall structure, they are not part of its primary structure! We'll talk more about these bonds, called disulphide bonds, in later sections.
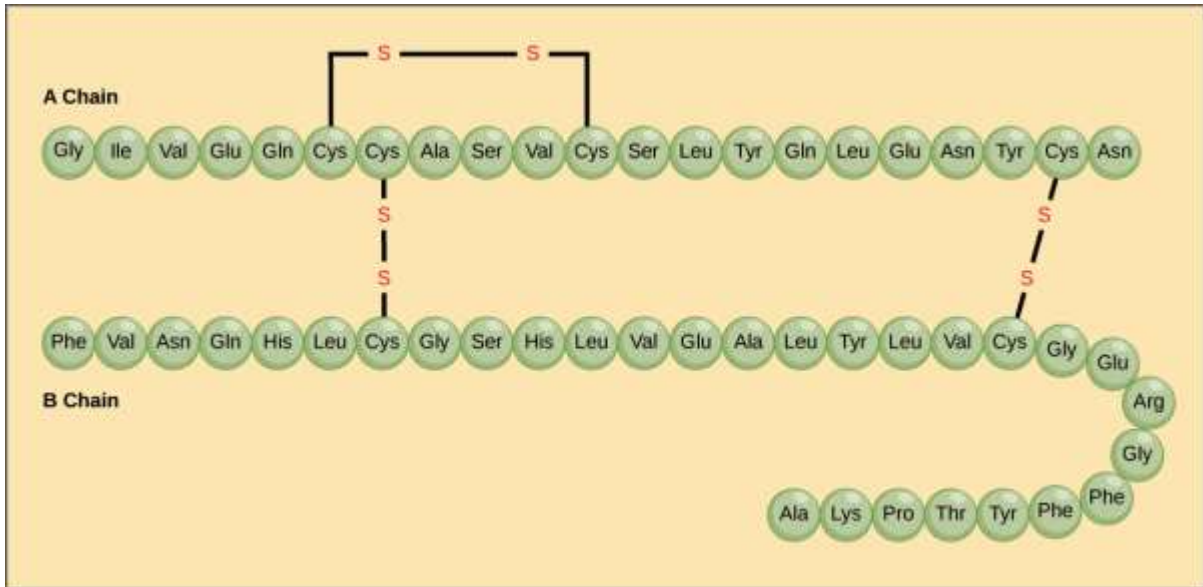
Image of insulin. Insulin consists of an A chain and a B chain. They are connected to one another by disulphide bonds (sulphur - sulphur bonds between cysteine's). The A chain also contains an internal disulphide bond. The amino acids that make up each chain of insulin are represented as connected circles, each with the three-letter abbreviation of the amino acid's name.

image credit: OpenStax Biology.

The sequence of a protein is determined by the DNA of the gene that encodes the protein (or that encodes a portion of the protein, for multi-subunit proteins). A change in the gene's DNA sequence may lead to a change in the amino acid sequence of the protein. Even changing just one amino acid in a protein's sequence can affect the protein's overall structure and function. For instance, a single amino acid change is associated with sickle cell anaemia, an inherited disease that affects red blood cells. In sickle cell anaemia, one of the polypeptide chains that make up haemoglobin, the protein that carries oxygen in the blood, has a slight sequence change. The glutamate that is normally the seventh amino acid of the haemoglobin β chain (one of two types of protein chains that make up haemoglobin) is replaced by a valine. This substitution is shown for a fragment of the β chain in the diagram below.
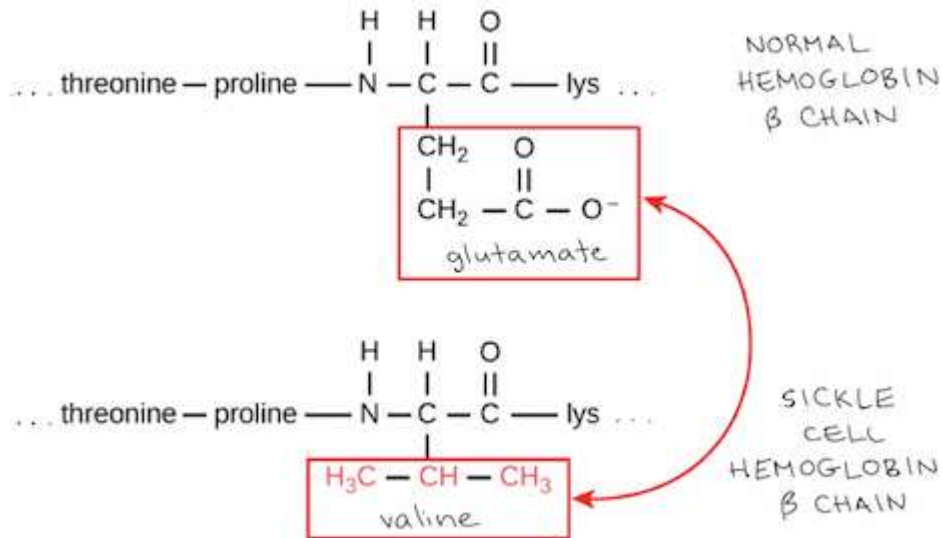

Image of normal and sickle cell mutant haemoglobin chains, showing substitution of valine for glutamate in the sickle cell version.

Image modified from OpenStax Biology.

What is most remarkable to consider is that a haemoglobin molecule is made up of two α chains and two β chains, each consisting of about 150 amino acids, for a total of about 600 amino acids in the whole protein. The difference between a normal haemoglobin molecule and a sickle cell molecule is just 2 amino acids out of the approximately 600.

A person whose body makes only sickle cell haemoglobin will suffer symptoms of sickle cell anaemia. These occur because the glutamate-to-valine amino acid change makes the haemoglobin molecules assemble into long fibres. The fibres distort disc-shaped red blood cells into crescent shapes. Examples of "sickled" cells can be seen mixed with normal, disc-like cells in the blood sample below.
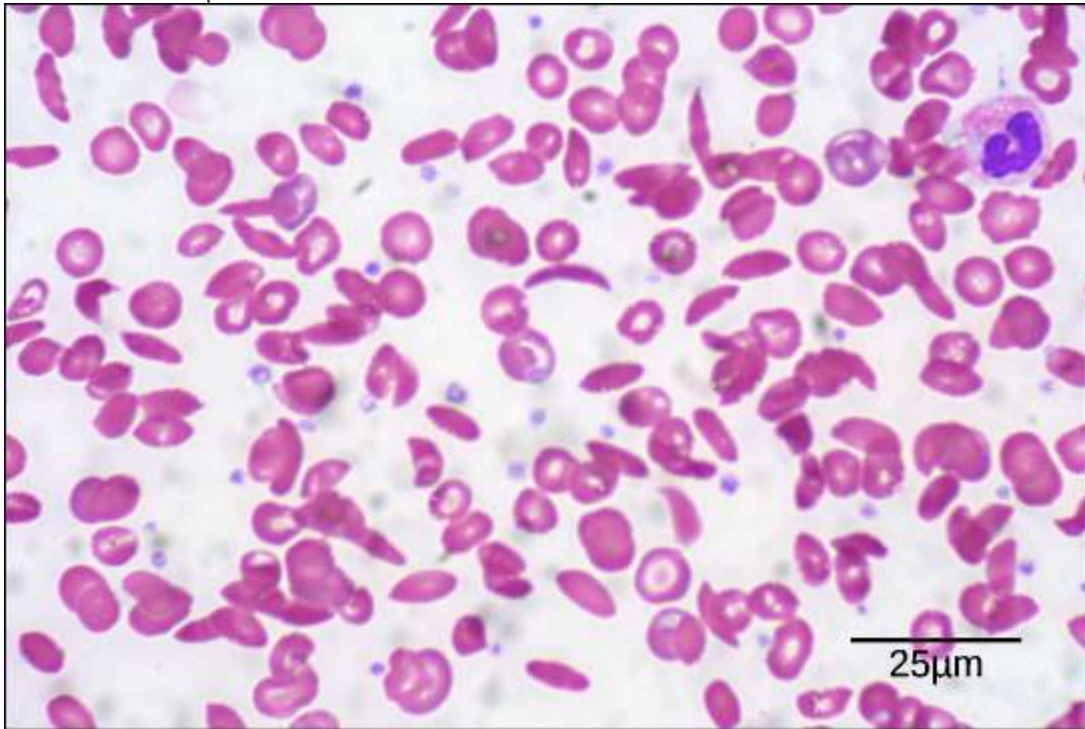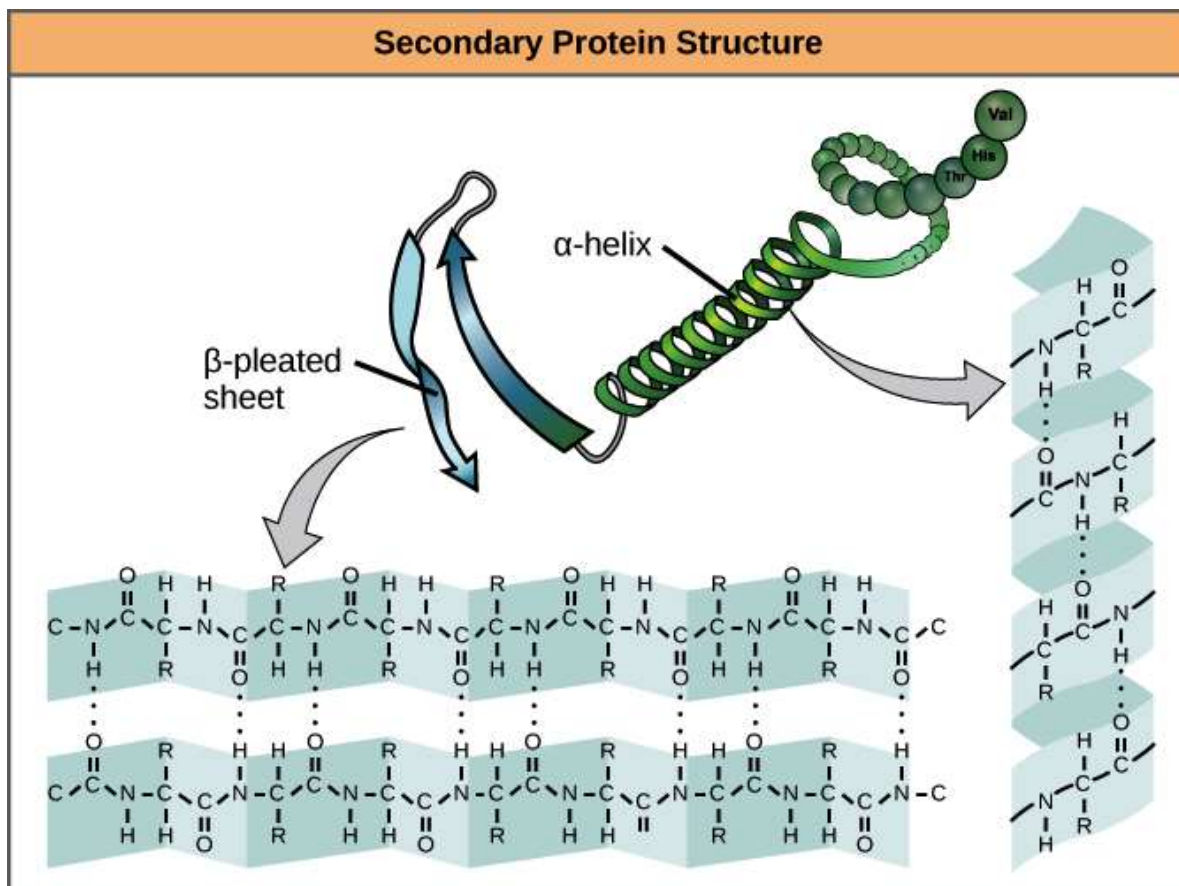

Image credit: OpenStax Biology modification of work by Ed Uthman; scale-bar data from Matt Russell.

The sickled cells get stuck as they try to pass through blood vessels. The stuck cells impair blood flow and can cause serious health problems for people with sickle cell anaemia, including breathlessness, dizziness, headaches, and abdominal pain.

Secondary structure

The next level of protein structure, **secondary structure**, refers to local folded structures that form within a polypeptide due to interactions between atoms of the backbone. (The backbone just refers to the polypeptide chain apart from the R groups – so all we mean here is that secondary structure does not involve R group atoms.) The most common types of secondary structures are the α helix and the β pleated sheet. Both structures are held in shape by hydrogen bonds, which form between the carbonyl O of one amino acid and the amino H of another.

Secondary Protein Structure

α-helix

β-pleated sheet

Images showing hydrogen bonding patterns in beta pleated sheets and alpha helices.

Image credit: OpenStax Biology.

In an **α helix**, the carbonyl (C=O) of one amino acid is hydrogen bonded to the amino H (N-H) of an amino acid that is four down the chain. (E.g., the carbonyl of amino acid 1 would form a hydrogen bond to the N-H of amino acid 5.) This pattern of bonding pulls the polypeptide chain into a helical structure that resembles a curled ribbon, with each turn of the helix containing 3.6 amino acids. The R groups of the amino acids stick outward from the α helix, where they are free to interact[3].

In a **β pleated sheet**, two or more segments of a polypeptide chain line up next to each other, forming a sheet-like structure held together by hydrogen bonds. The hydrogen bonds form between carbonyl and amino groups of backbone, while the R groups extend above and below the plane of the sheet[3]. The strands of a β pleated sheet may be **parallel**, pointing in the same direction (meaning that their N- and C-termini match up), or **antiparallel**, pointing in opposite directions (meaning that the N-terminus of one strand is positioned next to the C-terminus of the other).

Certain amino acids are more or less likely to be found in α-helices or β pleated sheets. For instance, the amino acid proline is sometimes called a "helix breaker" because its unusual R group (which bonds to the amino group to form a ring) creates a bend in the chain and is not compatible with helix formation[4]. Proline is typically found in bends, unstructured regions between secondary structures. Similarly, amino acids such as tryptophan, tyrosine, and phenylalanine, which have large ring structures in their R groups, are often found in β pleated sheets, perhaps because the β pleated sheet structure provides plenty of space for the side chains[4].

Many proteins contain both α helices and β pleated sheets, though some contain just one type of secondary structure (or do not form either type).

Tertiary structure

The overall three-dimensional structure of a polypeptide is called its **tertiary structure**. The tertiary structure is primarily due to interactions between the R groups of the amino acids that make up the protein.

R group interactions that contribute to tertiary structure include hydrogen bonding, ionic bonding, dipole-dipole interactions, and London dispersion forces – basically, the whole gamut of non-covalent bonds. For example, R groups with like charges repel one another, while those with opposite charges can form an ionic bond. Similarly, polar R groups can form hydrogen bonds and other dipole-dipole interactions. Also important to tertiary structure are **hydrophobic interactions**, in which amino acids with nonpolar, hydrophobic R groups cluster together on the inside of the protein, leaving hydrophilic amino acids on the outside to interact with surrounding water molecules.

Finally, there's one special type of covalent bond that can contribute to tertiary structure: the disulphide bond. **Disulphide bonds**, covalent linkages between the sulphur-containing side chains of cysteine's, are much stronger than the other types of bonds that contribute to tertiary structure. They act like molecular "safety pins," keeping parts of the polypeptide firmly attached to one another.
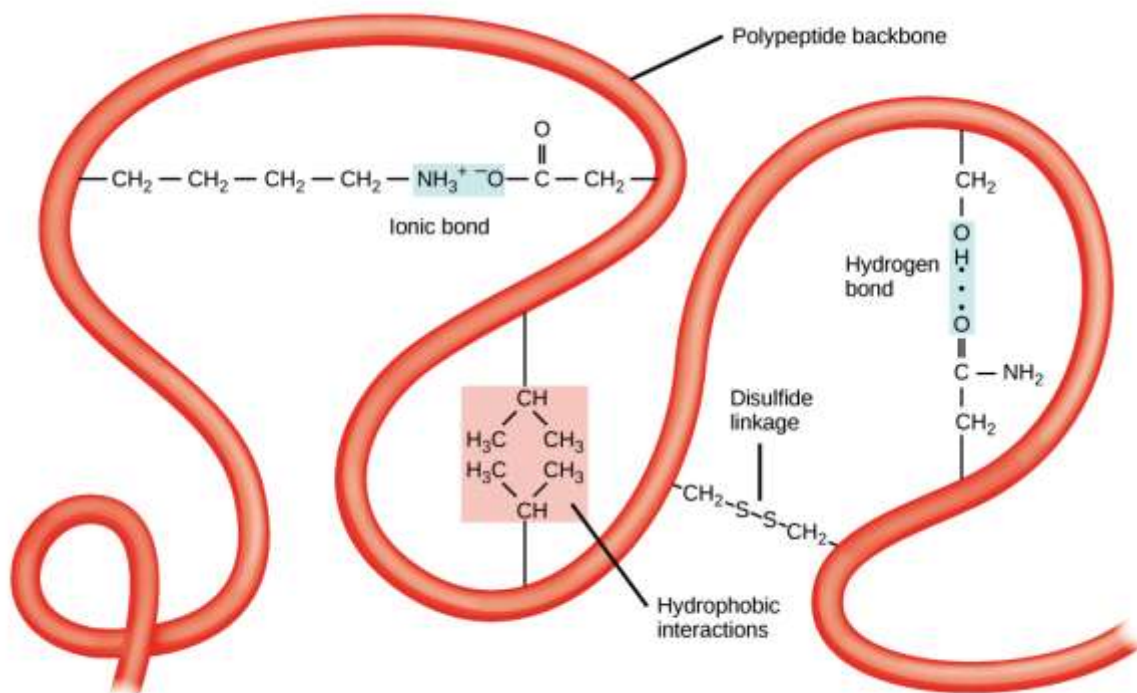

Image of a hypothetical polypeptide chain, depicting different types of side chain interactions that can contribute to tertiary structure. These include hydrophobic interactions, ionic bonds, hydrogen bonds, and disulphide bridge formation.
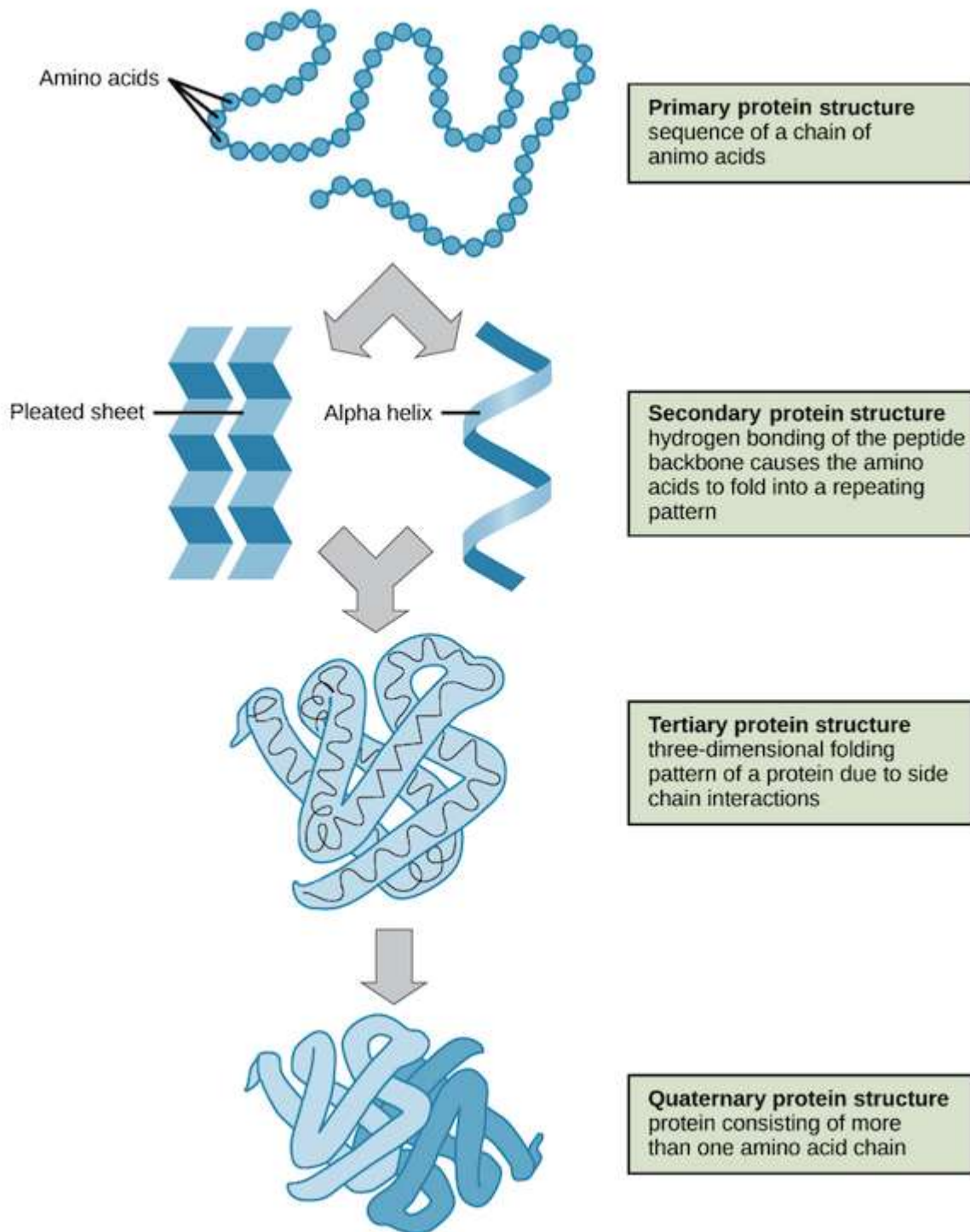
Image modified from OpenStax Biology.

Quaternary structure

Many proteins are made up of a single polypeptide chain and have only three levels of structure (the ones we've just discussed). However, some proteins are made up of multiple polypeptide chains, also known as subunits. When these subunits come together, they give the protein its **quaternary structure**.

We've already encountered one example of a protein with quaternary structure: haemoglobin. As mentioned earlier, haemoglobin carries oxygen in the blood and is made up of four subunits, two each of the α and β types. Another example is DNA polymerase, an enzyme that synthesizes new strands of DNA and is composed of ten subunits[5].

In general, the same types of interactions that contribute to tertiary structure (mostly weak interactions, such as hydrogen bonding and London dispersion forces) also hold the subunits together to give quaternary structure.

Flowchart depicting the four orders of protein structure.

Image modified from OpenStax Biology's modification of work by the National Human Genome Research Institute.

Denaturation and protein folding

Each protein has its own unique shape. If the temperature or pH of a protein's environment is changed, or if it is exposed to chemicals, these interactions may be disrupted, causing the

protein to lose its three-dimensional structure and turn back into an unstructured string of amino acids. When a protein loses its higher-order structure, but not its primary sequence, it is said to be **denatured**. Denatured proteins are usually non-functional.

For some proteins, denaturation can be reversed. Since the primary structure of the polypeptide is still intact (the amino acids haven't split up), it may be able to re-fold into its functional form if it's returned to its normal environment. Other times, however, denaturation is permanent. One example of irreversible protein denaturation is when an egg is fried. The albumin protein in the liquid egg white becomes opaque and solid as it is denatured by the heat of the stove, and will not return to its original, raw-egg state even when cooled down.

Researchers have found that some proteins can re-fold after denaturation even when they are alone in a test tube. Since these proteins can go from unstructured to folded all by themselves, their amino acid sequences must contain all the information needed for folding. However, not all proteins are able to pull off this trick, and how proteins normally fold in a cell appears to be more complicated. Many proteins don't fold by themselves, but instead get assistance from **chaperone** proteins (chaperonins).