## Examples of Protein Structures

Protein types

Proteins fall into three general classes, based on their overall three-dimensional structure and on their functional role: fibrous, membrane, and globular.

Fibrous proteins

Fibrous proteins tend to be long, narrow molecules. Fibrous proteins are used to construct macroscopic structures, especially structures outside of cells. Fibrous proteins tend to have a structural role, although some have more active functions as well.

Examples:

The protein **α-keratin** is used to form hair, nails, and skin. α-Keratin is a coiled-coil of α-helices ~300 residues long (corresponding to about 450 Å in length). α-Keratin is formed from two α-helices wrapped around one another (the backbone shown below is of a coiled-coil protein from PDB ID 3Q0X; this portion of the protein probably resembles the keratin coiled-coil structure)), and two of those structures wrapped around each other. Thus the keratin molecule contains a total of four intertwined helices.
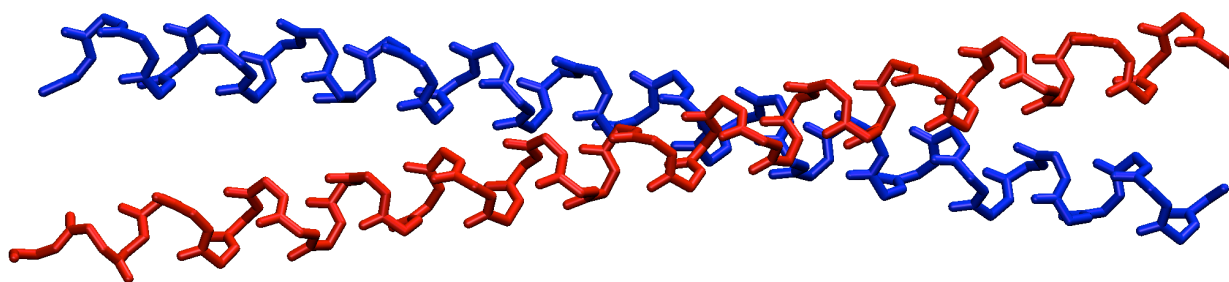
The helices in α-keratin are slightly distorted in order to allow the coiled-coil, but otherwise are formed from torsion angles similar to those present in normal α-helices. The approximate Φ / Ψ pair for the keratin helix is –64° / –42° compared to a normal α-helix of –57° / –47°.

The keratin helices within the protein are maintained in their relative locations by hydrogen bonds and disulfide bonds formed between side-chains of the helical residues. Hair and nail α-keratin molecules have a higher disulfide bond content than the form of keratin found in skin, and are tougher as a result.

High humidity allows facile rearrangement of hydrogen bonds, and can therefore allow the keratin molecules in hair to change their relative positions somewhat. Individuals with curly hair frequently find the degree of curling depends on the humidity; they also usually find that hair shaped while wet tends to retain its structure upon drying. A "permanent" results from reduction of the disulfide bonds between the chains, followed by shaping the hair, and by then oxidizing the free sulfhydryl groups to form new disulfide bonds that hold the hair in the new shape.

Smaller coiled-coil structures are also found at the interaction interface between some types of proteins that form less permanent complexes than keratin (especially in some transcription factors such as leucine zipper proteins).

The multiple types of collagen together comprise the most common animal protein (about 25% of the protein in animals is some form of collagen). Collagen is a triple helix formed from three polypeptide chains. The individual polypeptides form helices that are much more extended than an α-helix. A collagen helix contains 3.3 residues/turn compared to 3.6 residues/turn, but has a rise of 10 Å/turn, compared to 5.4 Å for an α-helix. An α-helix has a rise per residue of 1.5 Å; in contrast, a collagen helix has a rise of 3 Å per residue, and is therefore nearly as extended a structure as a β-sheet (3.5 Å/residue). The collagen triple helix is ~14 Å in diameter, compared to about 6 Å for an α-helix, due to the presence of three polypeptide chains.

The extended collagen helix is the result of its Φ / Ψ pair of –51° / 153°. The Φ angle is thus close to that of the α-helix, but the Ψ angle is quite different, and is close to the maximal extension angle (180°). The collagen triple helix has three of these extended helices wrapped around one another.[13]
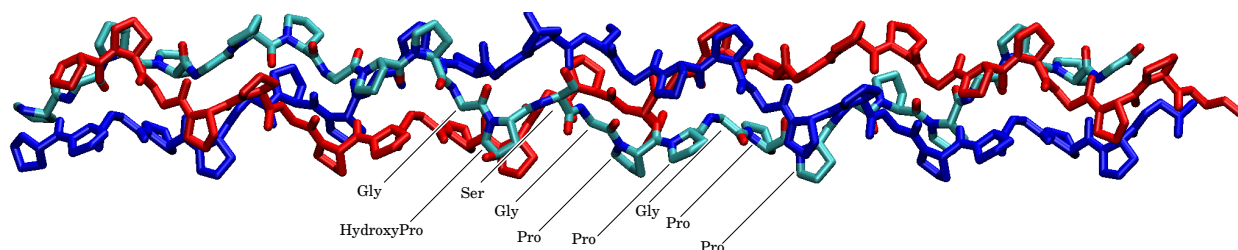
Each chain in the collagen triple helix is ~1000 residues long. The collagen molecule is therefore about 3000 Å = 0.3 μm long ([1000/3.3 residue/turn]·10 Å).

In collagen, the polypeptides have the typical sequence (Gly-X-Y)$_n$, where X residues are frequently proline and the Y residues are frequently hydroxyproline or hydroxylysine. Glycine fits in the center of triple helix (it is the only residue that is small enough to do so). Proline readily adopts the Φ / Ψ angles required by the triple helix. Note the proline rings in the strands shown above.

In collagen, the helices are held in register by several factors. One is hydrogen bonding involving hydroxyproline residues. The hydroxyl group is added to the proline residues following the synthesis of the collagen molecule in a vitamin C-dependent reaction. Some of the symptoms of scurvy (vitamin C deficiency) are due to inhibition of proline hydroxylation in collagen. Another factor that stabilizes the collagen structure is the result of covalent cross links formed between lysine, hydroxylysine, and histidine residues within the protein.[14]

---

[13] The drawing of collagen was created in VMD from PDB ID 3ADM; the structure was determined using trimers of a 27-residue artificial model peptide rather than a full-length collagen molecule.

[14] Since collagen molecules are much shorter than the macroscopic structures such as tendons and ligaments that are produced largely from collagen, it is apparent that the larger structures require

Membrane proteins

Membrane proteins comprise a unique class of proteins. For membrane proteins, a significant region of the protein must be stable in a hydrophobic environment. This is typically achieved by having non-polar side-chains on specific surface regions of the protein. Because of this exposed hydrophobic surface, and because many membrane proteins are destabilized by removal from the membrane, most membrane proteins are difficult to work with. As a result, structural information is available for only a relatively small number of these proteins, although new techniques have allowed three-dimensional structure determination for increasing numbers of membrane proteins in the last few years.

An example of a membrane protein is cytochrome *c* oxidase. Cytochrome *c* oxidase is one of the more important membrane proteins: it is the protein that donates electrons to oxygen in the electron transport chain. Cytochrome *c* oxidase is therefore the primary oxygen-utilization enzyme in aerobic organisms. Its importance stimulated the series of studies that resulted in the solution of its three-dimensional structure.

The region of the cytochrome *c* oxidase protein that interacts with the membrane is readily visible in the structural representation below.[15] In the figure, yellow residues are non-polar, light blue residues are polar, blue residues have basic side-chains, and red residues have acidic side-chains. Note the band of non-polar residues across the center of the protein. Note also that the region is predominantly α-helical. This is because α-helices satisfy all of the hydrogen bonding requirements of the peptide backbone, regardless of the polarity of the environment.
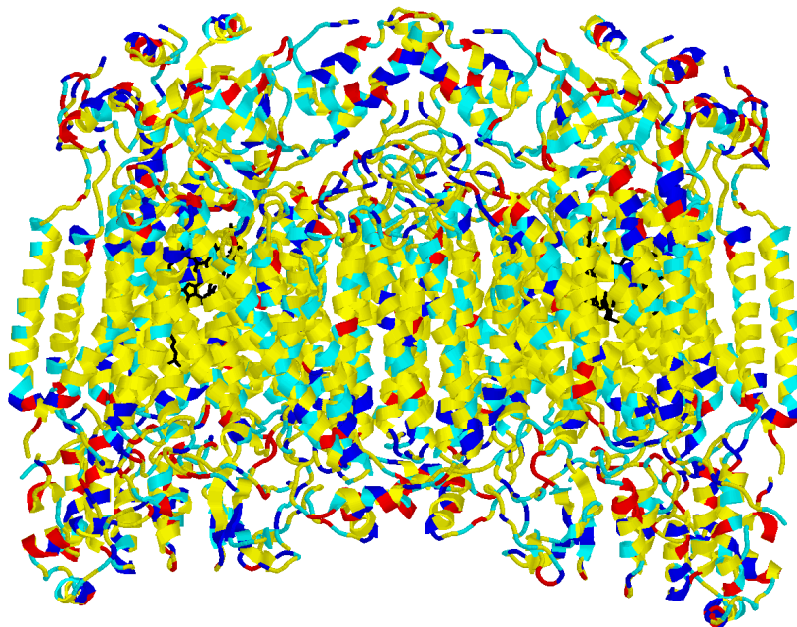
Cytochrome *c* oxidase

interactions between collagen molecules. Some of these interactions are also mediated by side chain hydrogen bonding and other links.

[15] The structural representation for cytochrome *c* oxidase was created in SwissPDB Viewer from PDB ID 2OCC (Yoshikawa, *et al.*, *Science* **280**: 1723-1729, 1998).

Globular proteins

Globular proteins comprise the most varied type of proteins. Globular proteins are soluble in aqueous solution. To achieve this, globular proteins generally have polar residues on the surface and hydrophobic residues on the interior.

Globular proteins frequently have stable structures that lend themselves to structure determination. This means that more three-dimensional structural information is available for globular proteins than for all other classes of proteins combined.

The structures of globular proteins are quite varied. Although early structural determination studies suggested that proteins have large amounts of α-helix (as was found for hemoglobin, shown below; hemoglobin is approximately 80% α-helical), the helical content of proteins can vary dramatically. The structure of human chorionic gonadotropin, for example, contains large amounts of β-sheet, but only a single, short, α-helix.

Hemoglobin
(from PDB ID 1B86)

human chorionic gonadotropin (hCG, from PDB ID 1HCN)
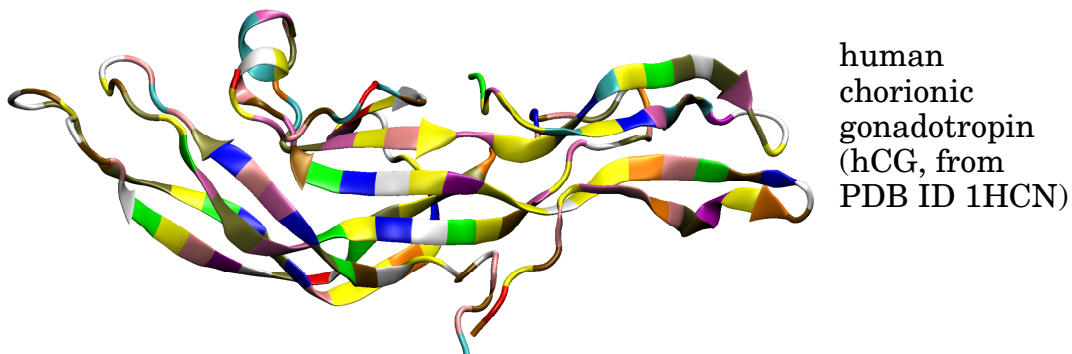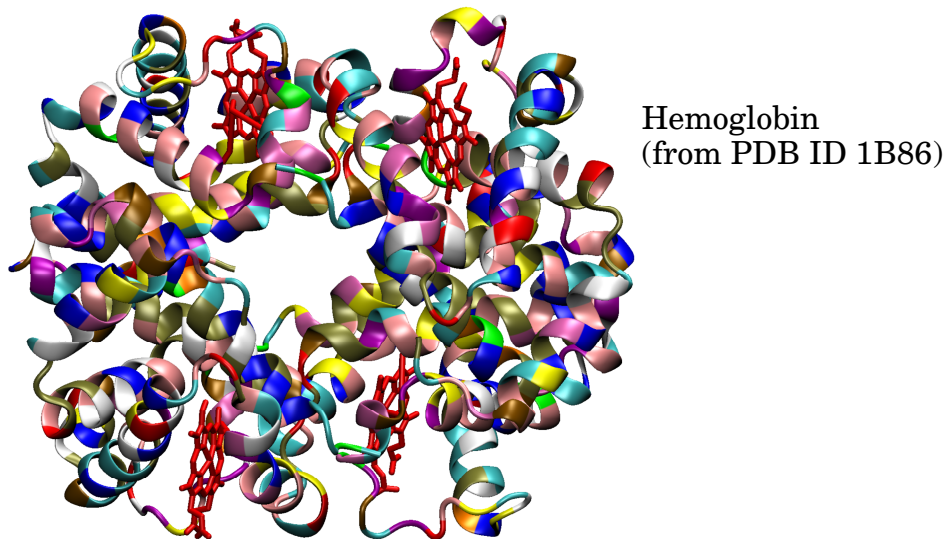
The above structure for hCG shows a protein that is quite elongated, and is far from spherical. However, hCG is considered a globular protein; the term "globular" refers more to the property of solubility in aqueous solution than to the roughly spherical shape of many proteins in this class (such as hemoglobin).

Representations of protein structure

When structural information is available, the protein is often simply presented as a snapshot of the three-dimensional structure. These snapshots can be of considerable use, but often obscure important details. One way around the problem of presenting complex three-dimensional information is to use computer programs that allow the viewer to examine the structure from many perspectives. However, even with computer graphics programs, the amount of information present in a three-dimensional representation is too large to interpret readily. Most computer visualization software packages have methods for simplifying the overall structure (compare, for example, the representations of hCG shown below).
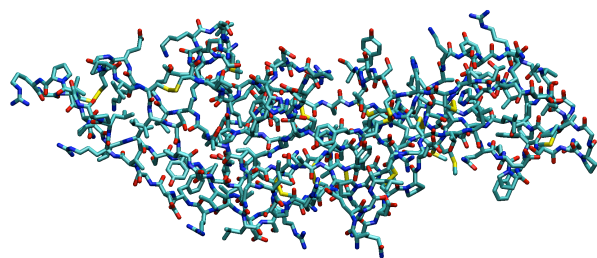

Licorice representation, showing all hCG polypeptide non-hydrogen atoms, color-coded by atom type. (These diagrams were both created in VMD.)
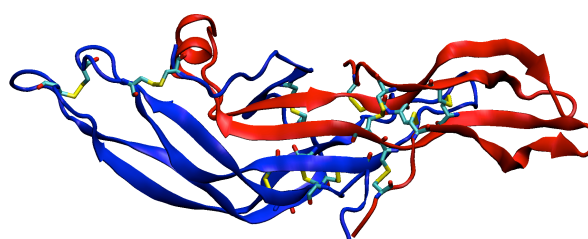
Ribbon diagram of the hCG backbone trace, color-coded by polypeptide chain (α chain in red, β chain in blue), with cysteine residues in licorice representation)

Although the diagrams above show the three-dimensional structure accurately, more stylized representations of structures are sometimes more useful. These allow emphasis of important features, and usually present the information more clearly. However, this may come at the cost of modifying the appearance of the actual structure considerably. For example, the stylized diagram for the hCG structure emphasizes the arrangement of β-structure in the protein at the cost of distorting the structure considerably.[16]

---
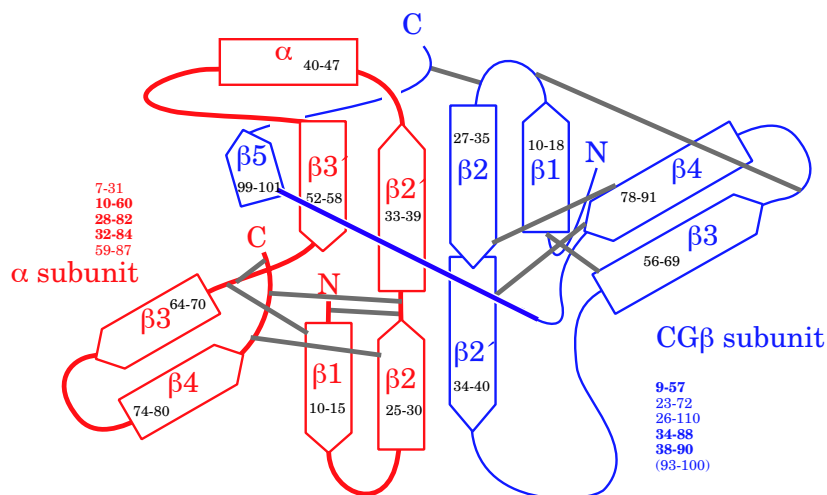
[16] The secondary structure diagram was modified from information and drawings in Wu, *et al.*, *Structure* **2**: 545-558 (1994) and Lapthorn, *et al.*, *Nature* **369**: 455-461 (1994)

Summary
Proteins can be classified according to their structural properties. This scheme divides proteins into three main types: fibrous, membrane, and globular.

Fibrous proteins are long narrow molecules. Most are involved in forming macroscopic structural elements, including keratin and collagen.

Membrane proteins typically have a hydrophobic region that interacts with the non-polar interior of membranes. These proteins are much less well understood than those of other classes, because they are difficult to purify and study. Membrane spanning regions of membrane proteins are frequently α-helical, although some exceptions have been observed.

Globular proteins are a diverse class of soluble proteins. Many of the most heavily studied proteins are members of this class of proteins.