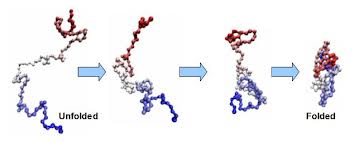
## What is a Protein?

Protein is a giant, complex molecule that plays an essential role in our body. It does most of the cell’s work and is required to structure, function, and regulate the body’s tissues and organs.

Protein is made of hundreds or thousands of long-chain smaller amino acids. **The sequence of amino acids determines the structure and function of the protein**. Twenty different types of amino acids are combined to create a protein.

**Protein Folding**

Protein folding is the physical process by which a linear polypeptide folds into its characteristic and functional three-dimensional structure. Folding of a polypeptide chain is strongly influenced by the **solubility of the AA R-groups in water**. Each protein exists as an unfolded polypeptide or random coil when translated from a sequence of mRNA to a linear chain of amino acids. This polypeptide lacks any stable (long-lasting) three-dimensional structure (the left hand side of the neighboring figure). **Amino acids interact with each other to produce a well-defined three-dimensional structure,** the folded protein (the right hand side of the figure), known as the native state. All the information for the native fold appears therefore to be contained within the primary structure (Anfinsen received the Nobel Prize for this), and proteins are self-folding (although *in vivo*, polypeptide folding is often assisted additional molecules known as molecular chaperones).



**Minimizing the number of hydrophobic side-chains exposed to water (the hydrophobic effect) is an important driving force behind the folding process**. **Intramolecular hydrogen bonds also contribute to protein stability** (think of their importance in secondary structures). **Ionic interactions** (attraction between unlike electric charges of ionized R-groups) also contribute to the stability of tertiary structures. **Disulfide bridges** (covalent bonds) **between neighboring cysteine residues** can also stabilize three-dimensional structures. Note that disulfide bonds are rarely observed in intracellular proteins because of the reducing intracellular environment.

The correct 3D structure of a protein is essential to its function, although some parts of functional proteins may remain unfolded. Failure to fold into native structure generally produces inactive proteins, but in some instances misfolded proteins have modified or toxic functionality (think prions & amyloid fibrils). Consistent with their functional importance, three-dimensional structures of proteins are more conserved during evolution time than are the primary amino-acid sequences.

## Key Points

* Protein folding is a process in which a linear chain of amino acids attains a defined three-dimensional structure, but there is a possibility of forming misfolded or denatured proteins, which are often inactive.
* Proteins must also be located in the correct part of the cell in order to function correctly; therefore, a signal sequence is often attached to direct the protein to its proper location, which is removed after it attains its location.
* Protein misfolding is the cause of numerous diseases, such as mad cow disease, Creutzfeldt-Jakob disease, and cystic fibrosis.

## Key Terms

* **prion**: a self-propagating misfolded conformer of a protein that is responsible for a number of diseases that affect the brain and other neural tissue
* **chaperone**: a protein that assists the non-covalent folding/unfolding of other proteins

## Protein Folding

After being translated from mRNA, all proteins start out on a ribosome as a linear sequence of amino acids. This linear sequence must “fold” during and after the synthesis so that the protein can acquire what is known as its native conformation. The native conformation of a protein is a stable three-dimensional structure that strongly determines a protein’s biological function. When a protein loses its biological function as a result of a loss of three-dimensional structure, we say that the protein has undergone denaturation. Proteins can be denatured not only by heat, but also by extremes of pH; these two conditions affect the weak interactions and the hydrogen bonds that are responsible for a protein’s three-dimensional structure. Even if a protein is properly specified by its corresponding mRNA, it could take on a completely dysfunctional shape if abnormal temperature or pH conditions prevent it from folding correctly. The denatured state of the protein does not equate with the unfolding of the protein and randomization of conformation. Actually, denatured proteins exist in a set of partially-folded states that are currently poorly understood. Many proteins fold spontaneously, but some proteins require helper molecules, called chaperones, to prevent them from aggregating during the complicated process of folding.

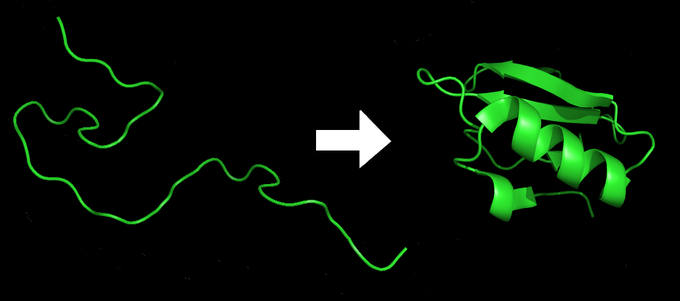


Figure 15.12.1: Protein folding: A protein starts as a linear sequence of amino acids, then folds into a 3-dimensional shape imbued with all the functional properties required inside the cell.

## Protein Modification and Targeting

During and after translation, individual amino acids may be chemically modified and signal sequences may be appended to the protein. A signal sequence is a short tail of amino acids that directs a protein to a specific cellular compartment. These sequences at the amino end or the carboxyl end of the protein can be thought of as the protein’s “train ticket” to its ultimate destination. Other cellular factors recognize each signal sequence and help transport the protein from the cytoplasm to its correct compartment. For instance, a specific sequence at the amino terminus will direct a protein to the mitochondria or chloroplasts (in plants). Once the protein reaches its cellular destination, the signal sequence is usually clipped off.

## Misfolding

It is very important for proteins to achieve their native conformation since failure to do so may lead to serious problems in the accomplishment of its biological function. Defects in protein folding may be the molecular cause of a range of human genetic disorders. For example, cystic fibrosis is caused by defects in a membrane-bound protein called cystic fibrosis transmembrane conductance regulator (CFTR). This protein serves as a channel for chloride ions. The most common cystic fibrosis-causing mutation is the deletion of a Phe residue at position 508 in CFTR, which causes improper folding of the protein. Many of the disease-related mutations in collagen also cause defective folding.

A misfolded protein, known as prion, appears to be the agent of a number of rare degenerative brain diseases in mammals, like the mad cow disease. Related diseases include kuru and Creutzfeldt-Jakob. The diseases are sometimes referred to as spongiform encephalopathies, so named because the brain becomes riddled with holes. Prion, the misfolded protein, is a normal constituent of brain tissue in all mammals, but its function is not yet known. Prions cannot reproduce independently and not considered living microoganisms. A complete understanding of prion diseases awaits new information about how prion protein affects brain function, as well as more detailed structural information about the protein. Therefore, improved understanding of protein folding may lead to new therapies for cystic fibrosis, Creutzfeldt-Jakob, and many other diseases.

Where does protein folding Occurs?

In all eukaryotic cells, the endoplasmic reticulum (ER) is an intracellular organelle where folding and assembly occurs for proteins destined to the extracellular space, plasma membrane, and the exo/endocytic compartments (Kaufman 1999). As a protein-folding compartment, the ER is exquisitely sensitive to alterations in homeostasis, and provides stringent quality control systems to ensure that only correctly folded proteins transit to the Golgi and unfolded or misfolded proteins are retained and ultimately degraded.

A number of biochemical and physiological stimuli, such as perturbation in calcium homeostasis or redox status, elevated secretory protein synthesis, expression of misfolded proteins, sugar/glucose deprivation, altered glycosylation, and overloading of cholesterol can disrupt ER homeostasis, impose stress to the ER, and subsequently lead to accumulation of unfolded or misfolded proteins in the ER lumen. The ER has evolved highly specific signaling pathways called the unfolded protein response (UPR) to cope with the accumulation of unfolded or misfolded proteins.

What are the 7 types of proteins?

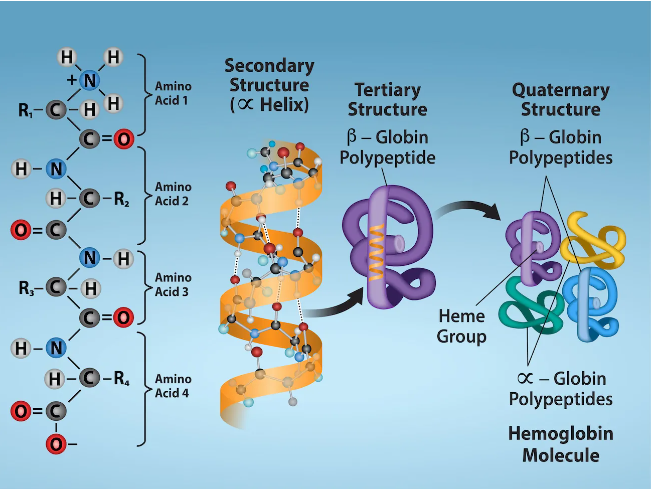
There is a total of seven different protein types under which all proteins fall. These include **antibodies, contractile proteins, enzymes, hormonal proteins, structural proteins, storage proteins, and transport proteins**.

# Hierarchical Structure of Proteins

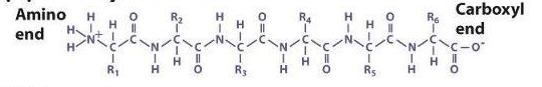
One of the earliest insights from structural biology was that many of the protein chains in cells fold into defined 3-dimensional structures. The structure of myoglobin revealed this for the first time, and since then, many of the features of folded proteins have been revealed as more and more structures have been determined. One of the central discoveries is that proteins have a hierarchical structure. The term "hierarchy" refers to something that is made of ranked levels, where simpler levels are arranged into more complex levels.

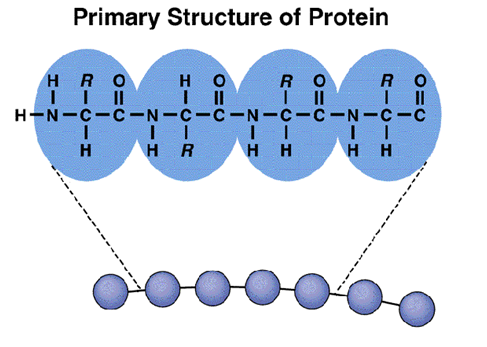
#### Hierarchical Structure

Protein structure is usually broken down into four hierarchical levels of organization



**Primary Structure**. This is the ordering of amino acids in the protein chain. In all living cells, this order is encoded in the organism's genome, and the protein is built by ribosomes by connecting amino acids in the proper order.





**Secondary Structure**. Nearly all folded proteins have structural elements that are formed in local regions of the protein chain. There are two common secondary structures: **alpha helices** and **beta strands**. Both of these secondary structures arrange the chain so that most of the main chain atoms form hydrogen bonds with themselves in a very efficient way--alpha helices form hydrogen bonds within the helix, and beta strands form hydrogen bonds with neighboring strands when arranged into beta sheets. Other specialized secondary structures are also observed, including defined structures for small loops and other types of helices.

Alpha-Helix and Beta-Pleated sheets are types of the secondary structure of the protein.

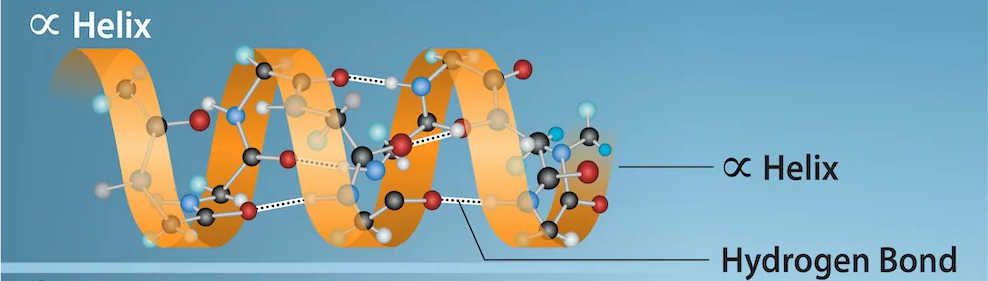
**Alpha-Helix Protein,**

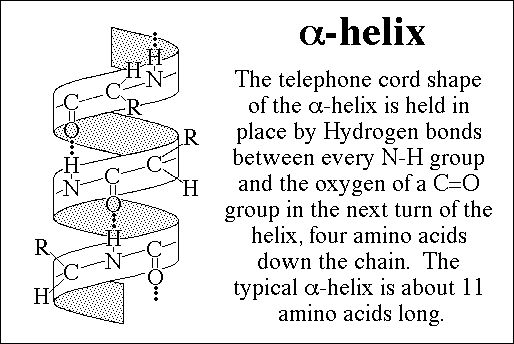
The most common type of secondary structure of a protein is the alpha-helix.

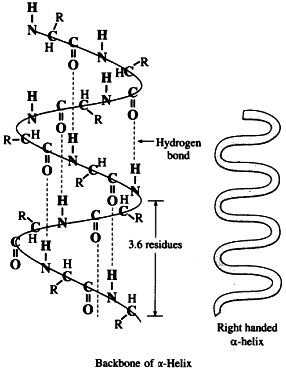
Linus Pauling predicted the structure of the alpha-helix protein. The prediction was confirmed when the first three-dimensional structure of protein myoglobin was determined by X-ray crystallography.

In the alpha-helix protein, a hydrogen bond is formed between the N−H group to the C=O group of the amino acid.

The alkyl groups of the alpha-helix chain are not involved in the H bonds but maintain the alpha-helix structure. Every winding turn in an alpha helix has 3.6 amino acids residues.







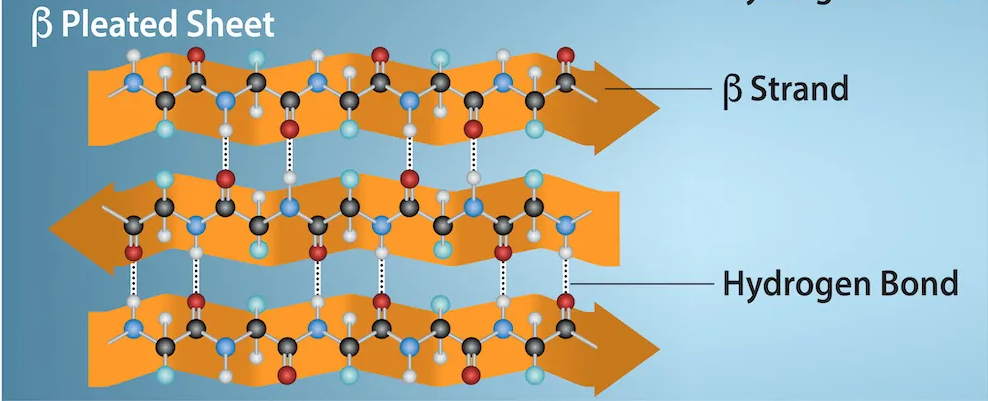
* The alpha helix involves regularly spaced Hydrogen bonds between residues along a chain.
* The amide hydrogen and the carbonyl oxygen of the peptide bond are forming hydrogen bond donors and acceptors respectively:
* The alpha helix is a common type of secondary structure found in proteins.
* It is a spiral-shaped structure held together by hydrogen bonds.
* The alpha helix has a regular, repeating pattern with a pitch of 5.4 Å (angstroms) per turn, and a rise of 1.5 Å per amino acid.
* The structure is stabilized by hydrophobic interactions between the nonpolar side chains of the amino acids.
* The alpha helix is a stable, compact structure that is found in many proteins, including structural proteins and enzymes.
* The alpha helix is right-handed when the chain is followed from the amino to the carboxyl direction. (The helical nomenclature is easily visualized by pointing the thumb of the right hand upwards—this is the amino to carboxyl direction of the helix. The helix then turns in the same direction as the fingers of the right-hand curve.)
* As the helix turns, the carbonyl oxygen of the peptide bond point upwards toward the downward-facing amide protons, making the hydrogen bond.
* The R groups of the amino acids point outwards from the helix.
* Helices are characterized by the number of residues per turn.
* In the alpha helix, there is not an integral number of amino acid residues per turn of the helix.
* There are 3.6 residues per turn in the alpha helix; in other words, the helix will repeat itself every 36 residues, with ten turns of the helix in that interval.

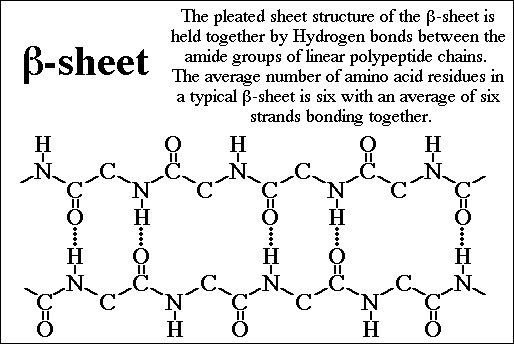
**Beta-Pleated Sheets of Protein**

The second essential type of secondary structure of a protein is the Beta-Pleated Sheets of Protein. It consists of various beta strands linked by hydrogen bonds between adjacent strands. Three to ten amino acids are combined to create a beta-strand polypeptide.

Beta sheets are involved in forming the fibrils and protein aggregates observed in amyloidosis.

Alike alpha-helix, the residue hydrogen bond between the adjacent strands is separate from each other.





* The beta sheet is a common type of secondary structure found in proteins.
* It is a flat structure formed by hydrogen bonds between adjacent strands of the polypeptide chain.
* The beta sheet can be either parallel or antiparallel, depending on the direction of the peptide bonds in the strands.
* The beta sheet is stabilized by hydrophobic interactions between the non-polar side chains of the amino acids.
* The beta sheet is a flexible structure that is found in many proteins, including structural proteins and enzymes.
* The beta sheet involves H-bonding between backbone residues in adjacent chains.
* In the beta sheet, a single chain forms H-bonds with its neighbouring chains, with the donor (amide) and acceptor (carbonyl) atoms pointing sideways rather than along the chain, as in the alpha helix.
* Beta sheets can be either parallel, where the chains point in the same direction when represented in the amino- to carboxyl- terminus, or anti-parallel, where the amino- to carboxyl- directions of the adjacent chains point in the same direction.

## Differences between Alpha-Helix and Beta-Sheet

| **S No.** | **Alpha-Helix** | **Beta-Sheet** |
| --- | --- | --- |
| **1** | **Amino acids exist in the right-handed coiled rod-like structure.** | **Amino acids exist in an almost entirely extended conformation, i.e. linear or sheet-like structure.** |
| **2** | **Intramolecular hydrogen bonding forms within the polypeptide chain to create a spiral structure.** | **Beta sheets are formed by linking two or more beta strands by intermolecular hydrogen bonds.** |
| **3** | **3.6 amino acid residues are winded to form an alpha-helix polypeptide.** | **Three to ten amino acids are combined to form a beta-strand polypeptide.** |
| **4** | **Alpha-Helix can be a single chain polypeptide.** | **Beta-Sheet cannot be in a single chain Polypeptide. There must be two or more beta-strands.** |
| **5** | **Alkyl groups of alpha-helix are oriented outside of the helix.** | **Alkyl groups are oriented both inside and outside of the sheet.** |
| **6** | **Example: Keratin, Myoglobin and Haemoglobin.** | **Example: Skin Fibres or Fibroin.** |

## Importance of secondary structure in protein function

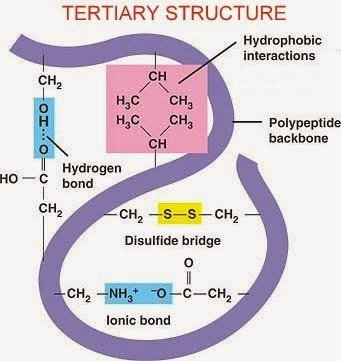
* The secondary structure of a protein plays a crucial role in its overall function.
* It determines the shape and stability of the protein, which is necessary for the protein to perform its specific function in the body.
* The secondary structure also determines the accessibility of the protein’s active site, which is the region of the protein responsible for its catalytic or regulatory activity.
* Changes in the secondary structure of a protein can alter its function, and this can have significant consequences for the overall function of the protein.
* For example, changes in the secondary structure of an enzyme can affect its catalytic activity, while changes in the secondary structure of a structural protein can affect its mechanical properties.
* The stability and function of a protein can also be affected by the presence of other molecules, such as ligands or other proteins, which can bind to specific sites on the protein and alter its conformation.

## Techniques for studying secondary structure

The shape and function of a protein are based on its secondary structure, which is an important part of its structure. There are several techniques that can be used to study the secondary structure of proteins, including spectroscopic techniques, X-ray crystallography, and nuclear magnetic resonance (NMR) spectroscopy.

* **Infrared spectroscopy and circular dichroism** are spectroscopic techniques that measure the absorption of light by the protein. These techniques can provide information about the secondary structure and conformation of the protein by analyzing the vibrations and rotations of the bonds within the protein.
* **X-ray crystallography**is a powerful technique that involves crystallizing the protein and using X-rays to determine its three-dimensional structure. This technique can provide highly detailed information about the secondary structure of the protein, as well as its tertiary and quaternary structures.
* **NMR spectroscopy** is another technique that can be used to study the secondary structure of proteins. It works by measuring the magnetic properties of the nuclei of the atoms in the protein, and can provide detailed information about the structure and conformation of the protein.
* In addition to these methods, **biochemical assays** and **computational modelling** can also be used to study the secondary structure of proteins. Biochemical assays, like enzyme assays, can give information about how enzymes work, while computational modelling can use computer simulations to predict the structure and conformation of the protein.

**Tertiary Structure**. The whole chain, with all of these local secondary structures, is then folded into the overall tertiary structure of the protein chain. This may include bundles of helices, beta strands stacked side-by-side into beta sheets that are then sandwiched together, and numerous other combinations.

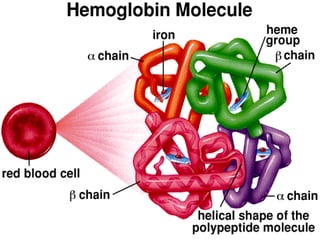


Tertiary structure is held together by four different bonds and interactions:

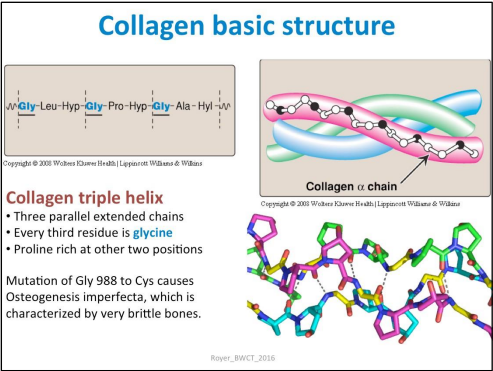
* + **Disulphide Bonds** - Where two **Cysteine** amino acids are found together, a strong double bond (S=S) is formed between the Sulphur atoms within the Cysteine monomers.
  + **Ionic Bonds** - If two oppositely charged ‘R’ groups (+ve and -ve) are found close to each other, and ionic bond forms between them.
  + **Hydrogen Bonds** - Your typical everyday Hydrogen bonds.
  + **Hydrophobic and Hydrophilic Interactions** - Some amino acids may be hydrophobic while others are hydrophilic. In a water based environment, a globular protein will orientate itself such that it’s hydrophobic parts are towards its centre and its hydrophilic parts are towards its edges

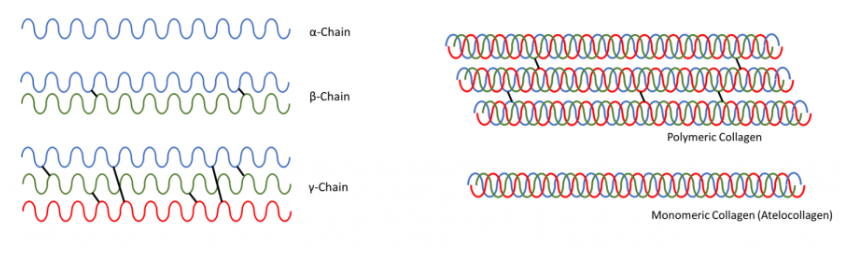
**Quaternary Structure**. Finally, two or more folded chains can associate to form a functional assembly, sometimes with an inorganic component, to form a **Quaternary Structure** protein. The individual proteins may be identical (homo-oligomers) or of many different types (hetero-oligomers). The proteins typically form very specific interfaces that bind them together in specific orientations. In the PDB archive the functional quaternary structure is termed the ["Biological Assembly”](https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/biological-assemblies).

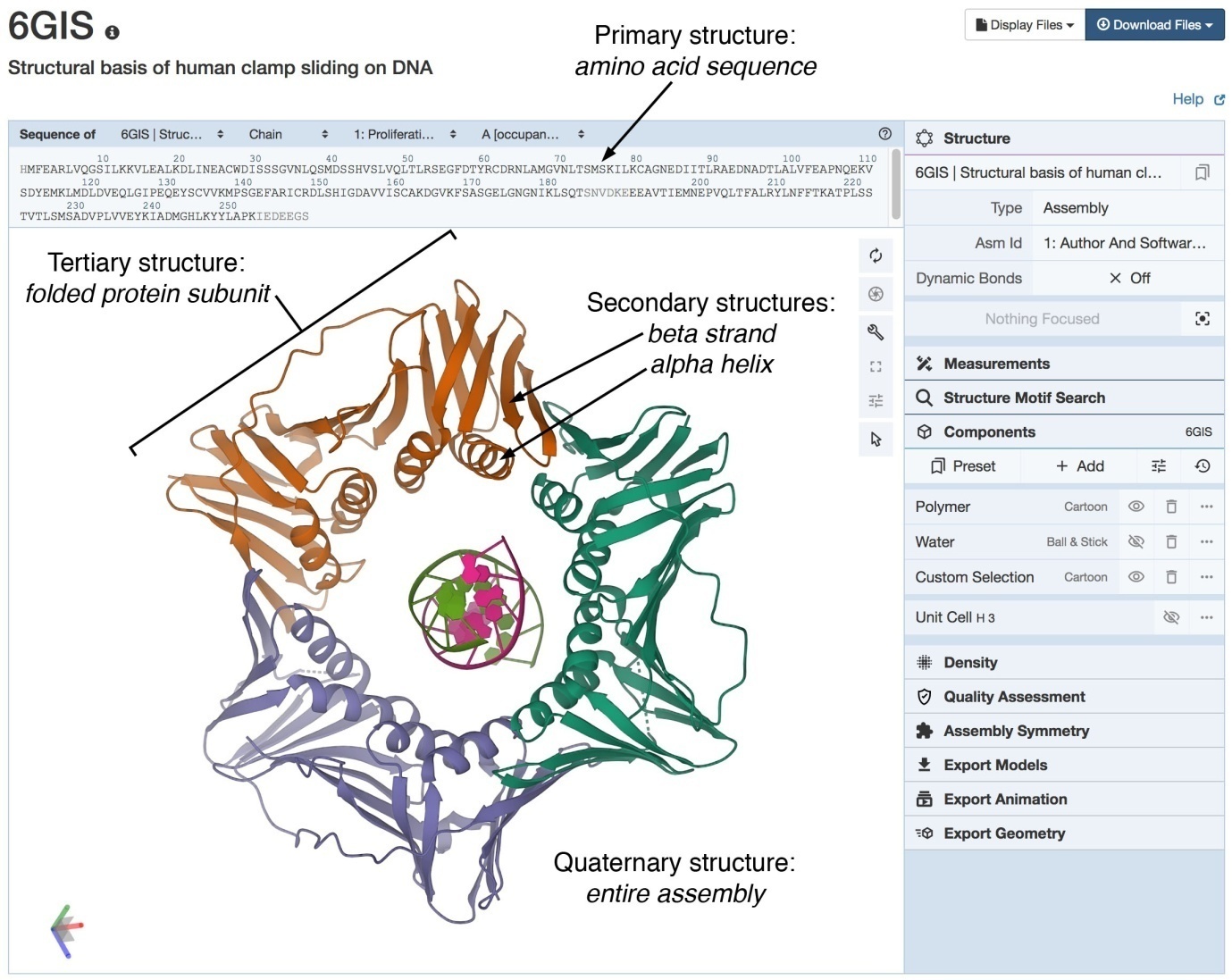
* proteins are made up of multiple polypeptide chains, sometimes with an inorganic component (*for example, a haem group in haemoglogin*) called a **Prosthetic Group**. These proteins will only be able to function if all subunits are present.
* **Haemoglobin** is a water soluble globular protein which is composed of two α polypeptide chains, two β polypeptide chains and an inorganic prosthetic haem group. Its function is to carry oxygen around in the blood, and it is facilitated in doing so by the presence of the haem group which contains a Fe2+ ion, onto which the oxygen molecules can bind.



* **Collagen** is a fibrous protein consisting of three polypeptide chains wound around each other. Each of the three chains is a coil itself. Hydrogen bonds form between these coils, which are around 1000 amino acids in length, which gives the structure strength. This is important given collagen’s role, as structural protein. This strength is increased by the fact that collagen molecules form further chains with other collagen molecules and form **Covalent Cross Links** with each other, which are staggered along the molecules to further increase stability. Collagen molecules wrapped around each other form **Collagen Fibrils** which themselves form **Collagen Fibres**.







**SUMMARY of Protein Structure**

**PRIMARY STRUCTURE**

1. The primary structure of a protein refers to the linear sequence of amino acids in the protein chain.
2. The specific order of amino acids is determined by the DNA sequence of the gene that codes for the protein.
3. The primary structure of a protein is critical to its function, as it determines the chemical and physical properties of the protein.
4. The primary structure can be determined experimentally using techniques such as protein sequencing and mass spectrometry.
5. The primary structure of a protein can be altered by mutations in the DNA sequence of the gene that codes for the protein. These mutations can have significant effects on the structure and function of the protein.

**SECONDARY STRUCTURE**

1. The secondary structure of a protein refers to the local folding of the protein chain into regular, repeating patterns.
2. The two most common types of secondary structures are alpha-helices and beta-sheets.
3. Alpha-helices are formed by the twisting of the protein chain into a right-handed helix. This helix is stabilized by hydrogen bonds between the carbonyl group of one amino acid and the amino group of an amino acid four positions down the chain.
4. Beta-sheets are formed by the folding of the protein chain into a flat, sheet-like structure. This structure is stabilized by hydrogen bonds between the carbonyl and amino groups of adjacent amino acids in the protein chain.
5. The secondary structure of a protein is stabilized by hydrogen bonds between nearby amino acids.
6. The secondary structure of a protein can have significant effects on the function of the protein, as it determines the specific interactions between the protein and other molecules in the cell.
7. The secondary structure can be predicted computationally using various algorithms, based on the primary sequence of the protein.
8. The secondary structure of a protein can be altered by mutations in the primary sequence of the protein, which can have significant effects on the stability and function of the protein.

**TERTIARY STRUCTURE**

1. The tertiary structure of a protein refers to the three-dimensional arrangement of the protein chain.
2. The tertiary structure is stabilized by a complex interplay of interactions between different regions of the protein, including hydrogen bonds, hydrophobic interactions, and disulfide bonds.
3. The tertiary structure is critical to the function of the protein, as it determines the specific interactions between the protein and other molecules in the cell.
4. The tertiary structure can be visualized using various techniques, including X-ray crystallography and nuclear magnetic resonance spectroscopy.
5. The tertiary structure of a protein can be altered by mutations in the primary sequence of the protein, which can have significant effects on the stability and function of the protein.
6. The tertiary structure of a protein can also be influenced by external factors such as pH, temperature, and the presence of other molecules in the cell.
7. Proteins with similar tertiary structures often have similar functions, even if their primary and secondary structures are different.
8. The tertiary structure of a protein can be predicted computationally using various algorithms, based on the primary sequence of the protein.

**QUATERNARY STRUCTURE**

1. The quaternary structure of a protein refers to the arrangement of multiple protein subunits to form a functional protein complex.
2. The subunits in a protein complex can be identical or different, and they interact with each other through various types of bonds and interactions, including hydrogen bonds, disulfide bonds, hydrophobic interactions, and electrostatic interactions.
3. The quaternary structure is critical to the function of the protein complex, as it determines the specific interactions between the protein complex and other molecules in the cell.
4. The quaternary structure of a protein complex can be visualized using various techniques, including X-ray crystallography and electron microscopy.
5. The quaternary structure of a protein complex can be altered by mutations in the genes that code for the subunits, which can have significant effects on the stability and function of the complex.
6. The quaternary structure of a protein complex can also be influenced by external factors such as pH, temperature, and the presence of other molecules in the cell.
7. Proteins with similar quaternary structures often have similar functions, even if their primary, secondary, and tertiary structures are different.
8. The quaternary structure of a protein complex can be predicted computationally using various algorithms, based on the primary sequences of the subunits and their known interactions.

