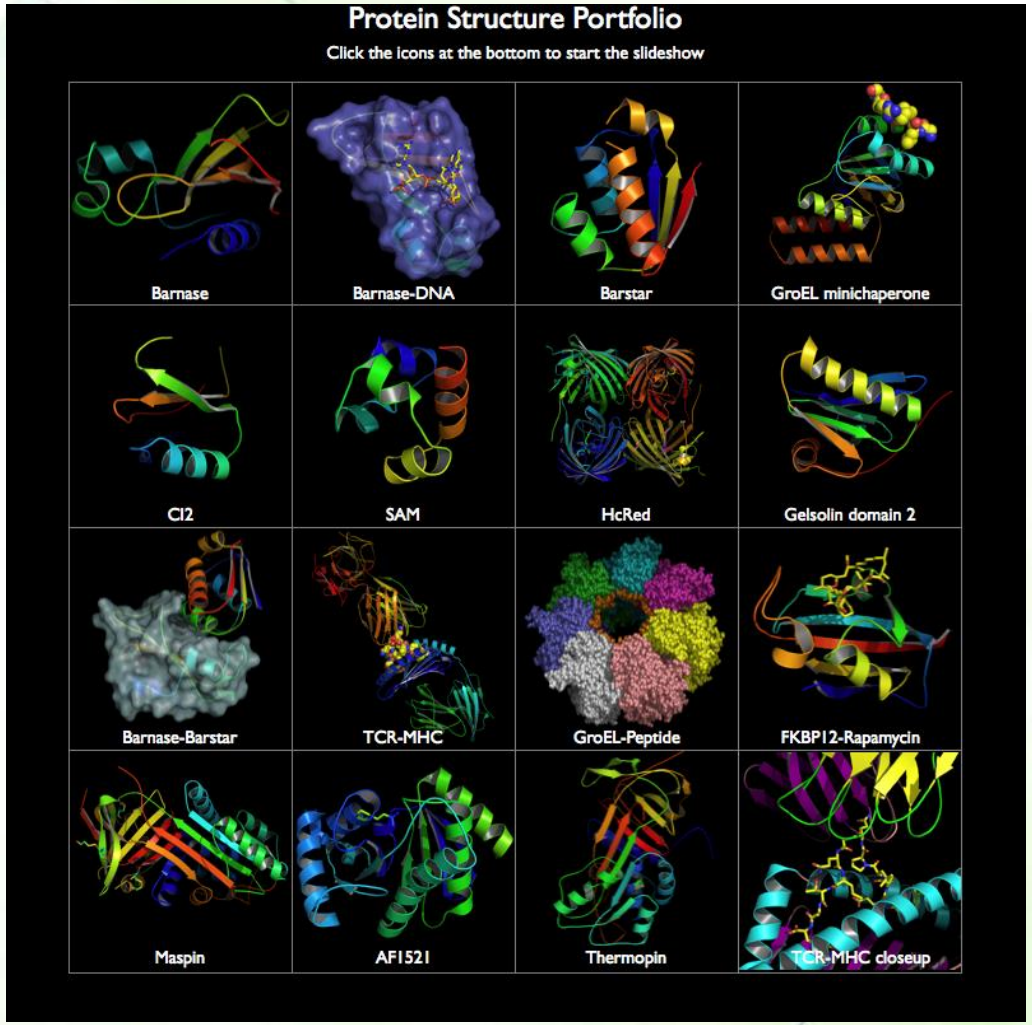
Protein structure

Summer 2024

Overview of proteins

- Proteins have different structures, and some have repeating inner structures, other do not.
- A protein may have *gazillion* possibilities of structures, but a few would be active.
- These active structures are known as native conformations (the 3-dimensioanl structure of a properly folded and functional protein).



Tunyasuvunakool, K., Adler, J., Wu, Z. et al. Highly accurate protein structure prediction for the human proteome. Nature (2021). https://doi.org/10.1038/s41586-021-03828-1

Highly accurate protein structure prediction for the human proteome

Kathryn Tunyasuvunakool, Jonas Adler, [...]Demis Hassabis

Nature (2021) | Cite this article

New public database of AI-predicted protein structures could transform biology

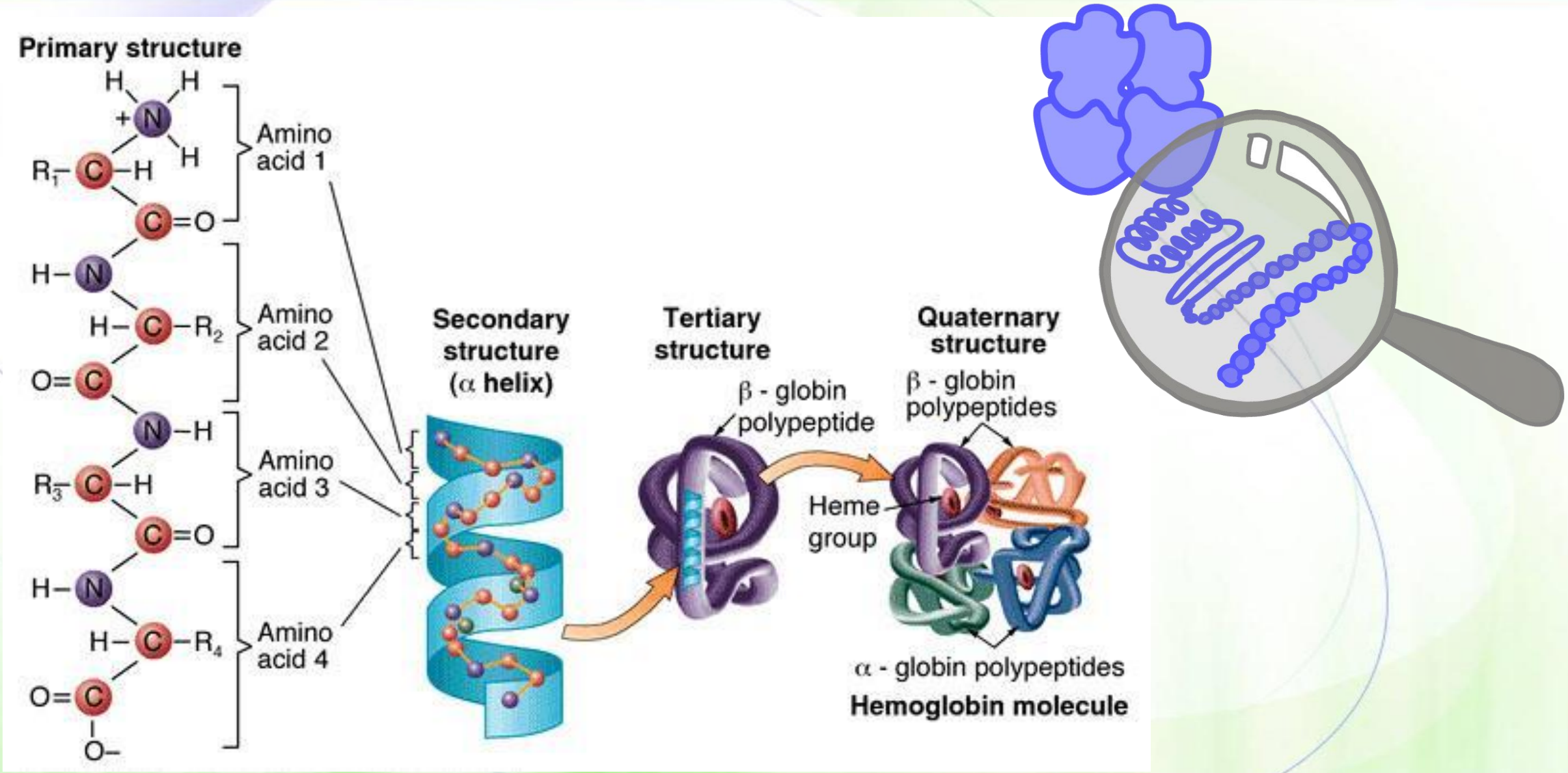
By Robert F. Service | Jul. 22, 2021 , 11:00 AM

AI MACHINE LEARNING & DATA SCIENCE POPULAR RESEARCH

DeepMind's AlphaFold2 Predicts Protein Structures with Atomic-Level Accuracy

Levels of protein structure

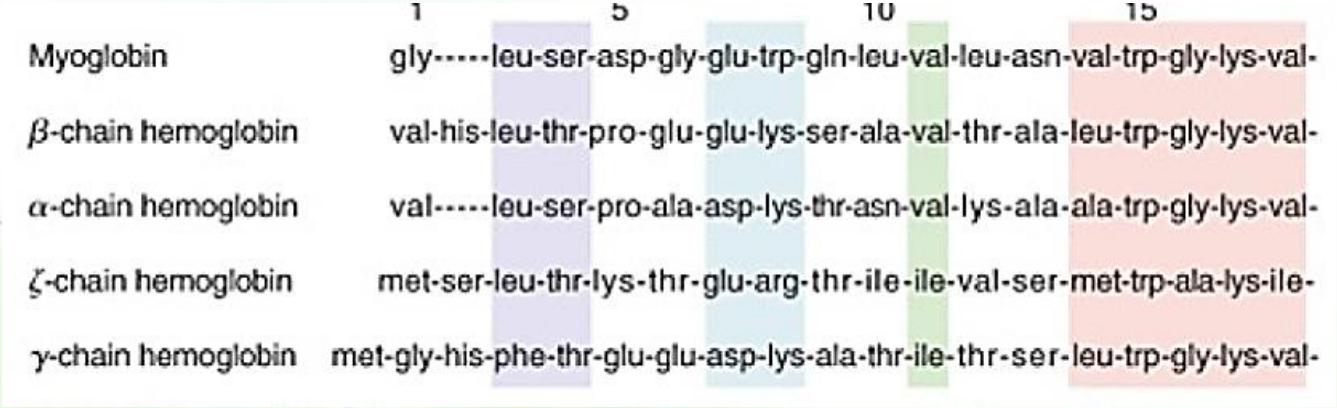
- Primary structure: the sequence of amino acid residues
- Secondary structure: the localized organization of parts of a polypeptide chain
- Tertiary structure: the three-dimensional structure and/or arrangement of all the amino acids residues of a polypeptide chain
- Some proteins are made of multiple polypeptides crosslinked (connected) with each other. These are known as multimeric proteins. Quaternary structure describes the number and relative positions of the subunits in a multimeric protein


Primary structure
Amino acid 1
Amino acid 2
Amino acid 3
Amino acid 4
Secondary structure (α helix)
Tertiary structure
β - globin polypeptide
Heme group
Quaternary structure
β - globin polypeptides
α - globin polypeptides
Hemoglobin molecule

Primary structure

What is primary structure?

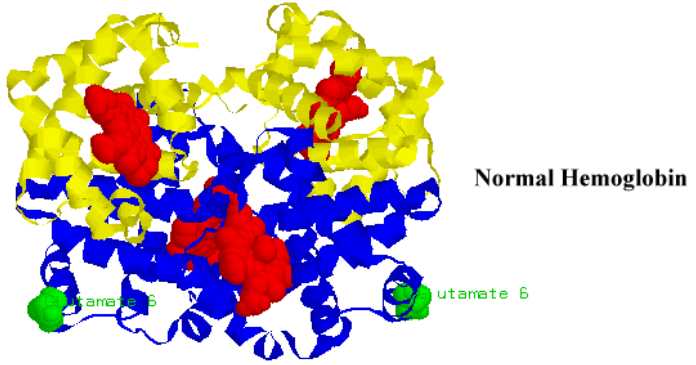
- The order in which the amino acids are covalently linked together.
  - Example: Leu—Gly—Thr—Val—Arg—Asp—His
- The primary structure of a protein determines the other levels of structure.
- Proteins that differ somewhat in primary structure and properties from tissue to tissue, but that retain essentially the same function, are called tissue-specific isoforms or isozymes.

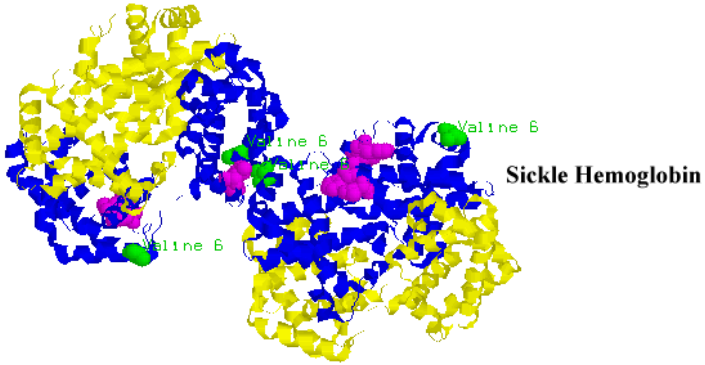


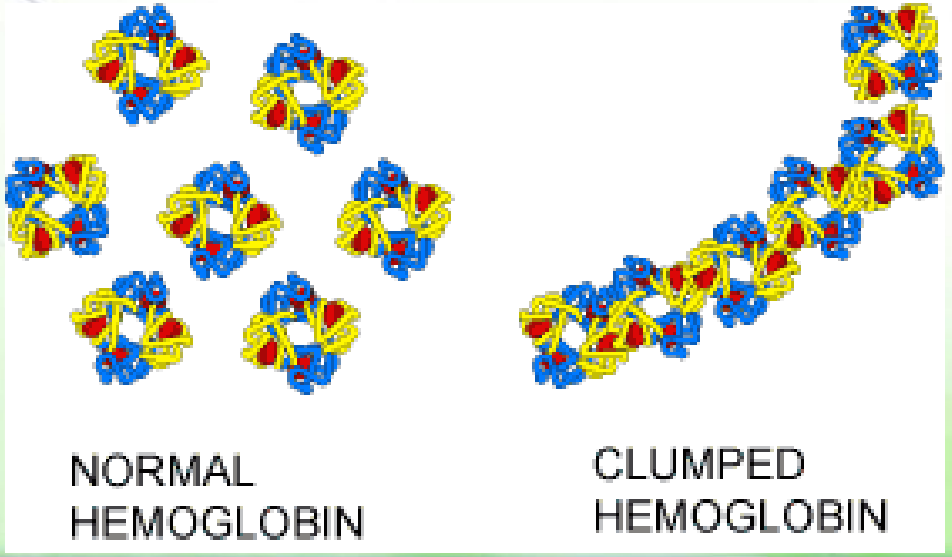
Why is it the primary structure important?

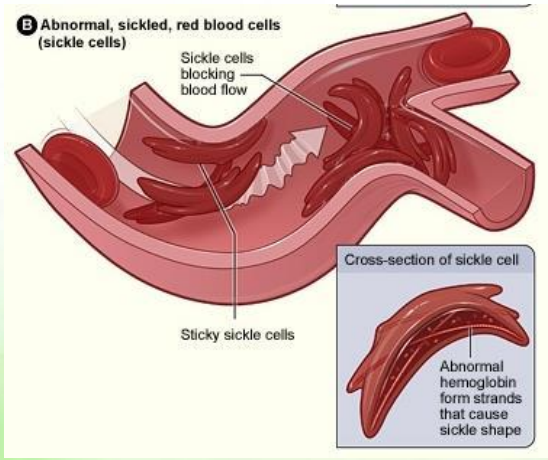
Sickle hemoglobin (HbS) as an example

- A single amino acid substitution can give rise to a malfunctioning protein, as is the case with sickle-cell anemia.
- It is caused by a change of amino acids in the 6th position of β globin (Glu to Val).
- The mutation results in: 1) arrays of aggregates of hemoglobin molecules, 2) deformation of the red blood cell, and 3) clotting in blood vessels and tissues.









Secondary Structure

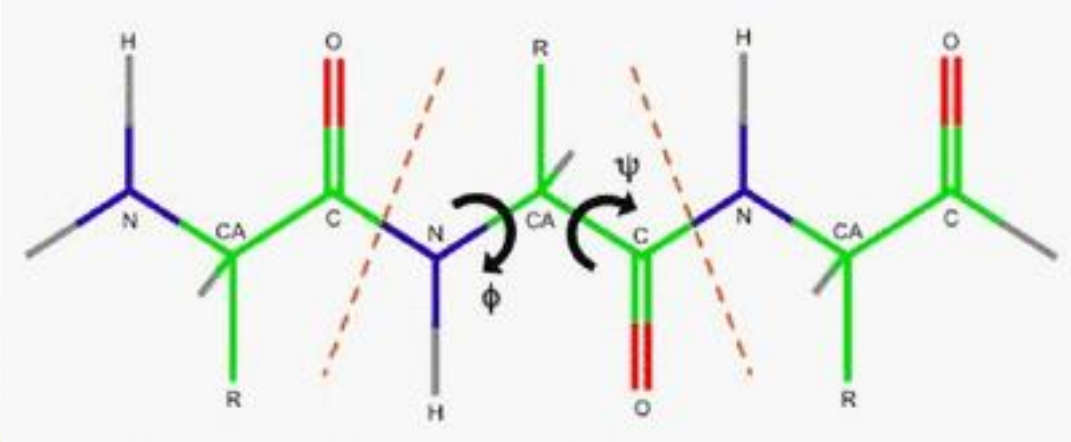
What is a secondary structure? How is it formed?

- Two bonds within each amino acid residue can freely rotate.
  - the bond between the α-carbon and the amino nitrogen
  - the bond between the α-carbon and the carboxyl carbon
- A secondary structure is hydrogen-bonded, locally arranged structures of the backbone of a polypeptide chain.
- Examples:
  - Alpha helix
  - Beta-pleated sheet

  } Regular secondary structures

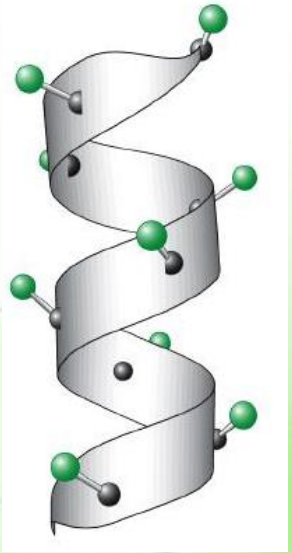
  - Turns
  - Loops
  - Bends
  - Coils

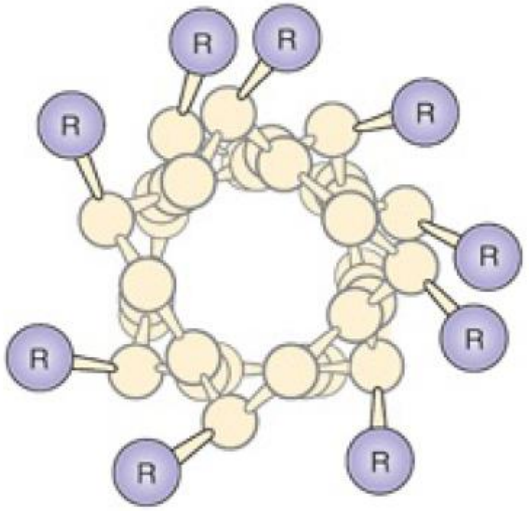
  } Nonregular secondary structures

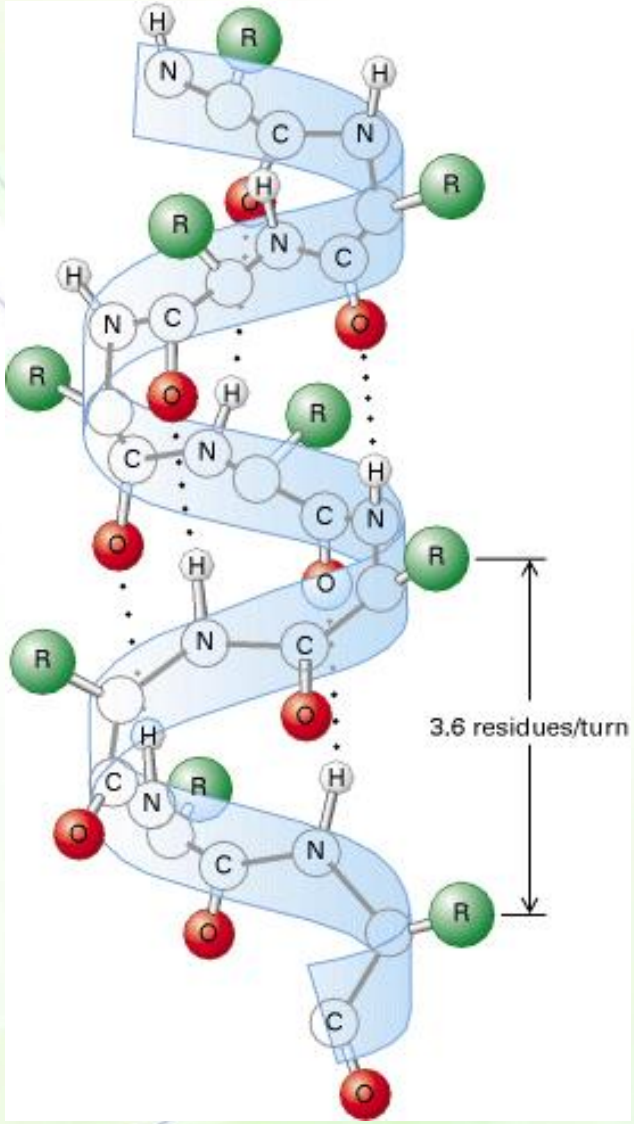


The α helix

- It looks like a helical rod.
- The helix has an average of 3.6 amino acids per turn.
- It is very stable because of the linear hydrogen bonds.
- The side chains of the amino acids project outward from the helix in order to avoid molecular congestion (known as steric hindrance) with the polypeptide backbone and with each other.

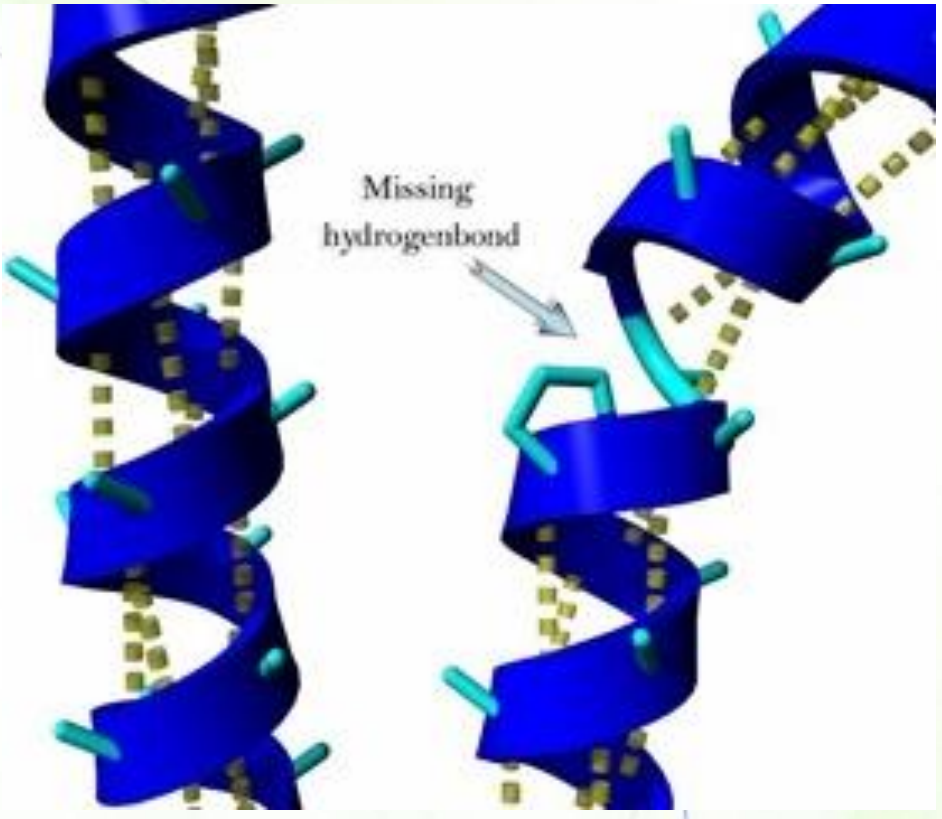


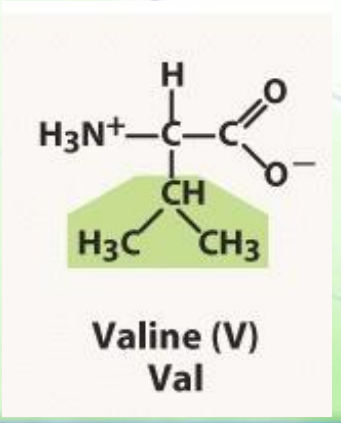


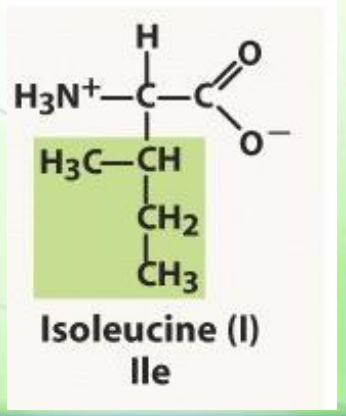


Amino acids NOT found in α-helix

- Glycine: too small
- Proline
  - No rotation around N-C bond
  - No hydrogen bonding of α-amino group
- Close proximity of a pair of charged amino acids with similar charges
- Amino acids with branches at the β-carbon atom (valine, threonine, and isoleucine)

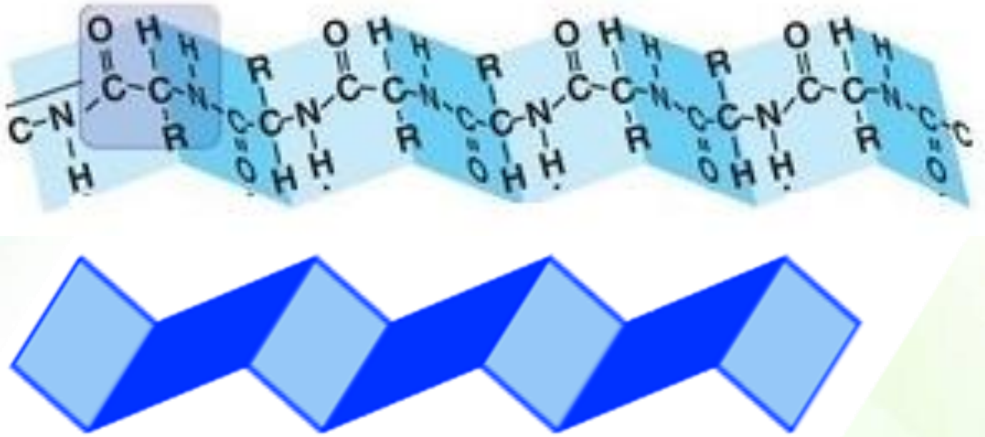






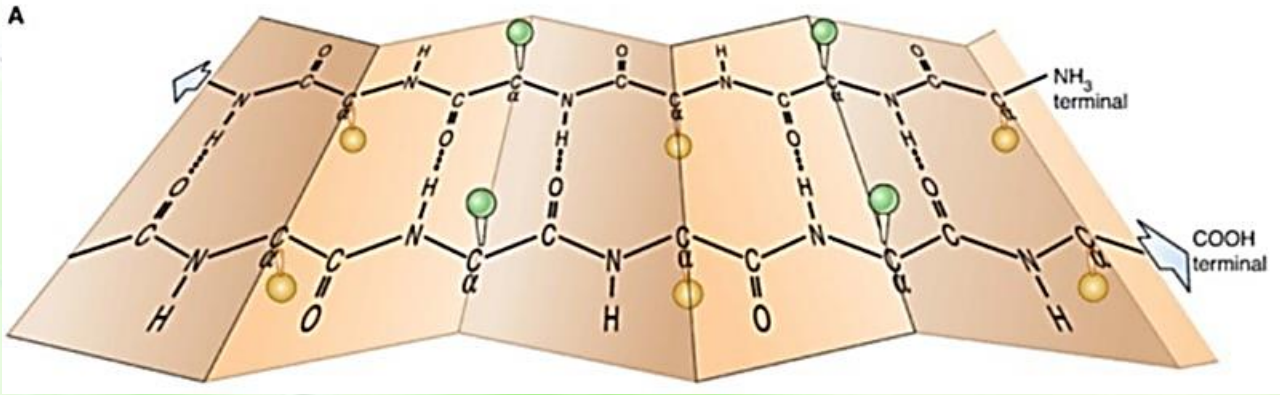
β-pleated sheet (β sheet)

- They are composed of two or more straight chains (β strands) that are hydrogen bonded side by side.



β strand

- Optimal hydrogen bonding occurs when the sheet is bent or folded (pleated) to form β-pleated sheets.

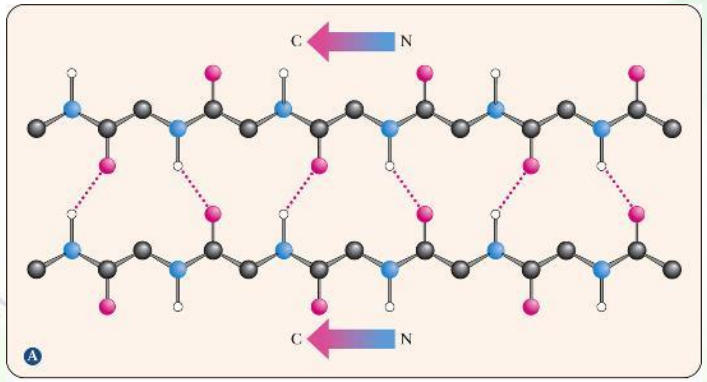


A pleated skirt

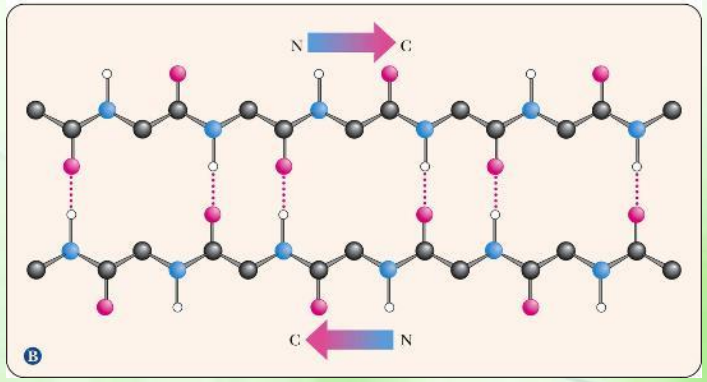
More on β-sheets

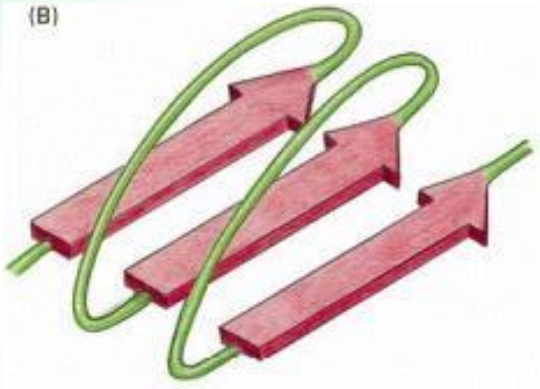
- β sheets can form between many strands, typically 4 or 5 but as many as 10 or more.
- Such β sheets can be purely antiparallel, purely parallel, or mixed.
- Proline tends to disrupt β strands. Why?

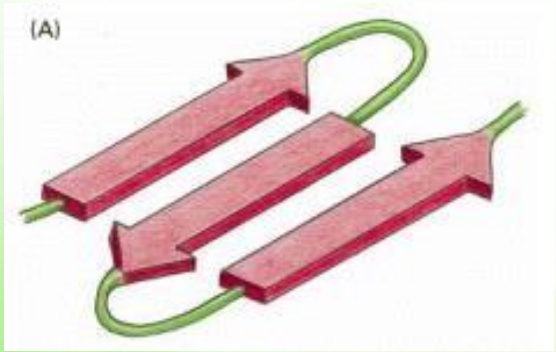
Parallel

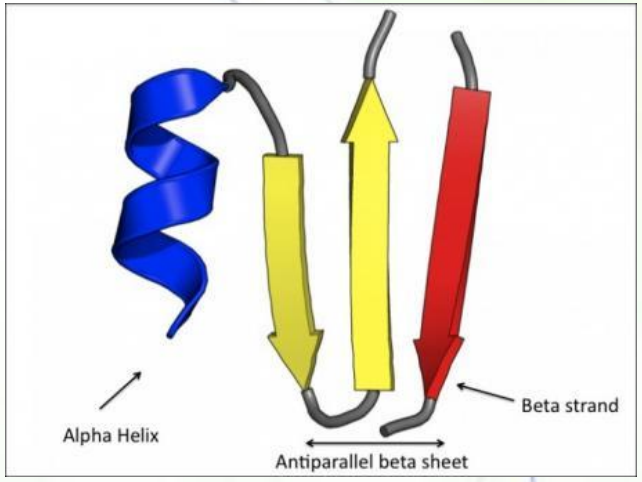


Antiparallel



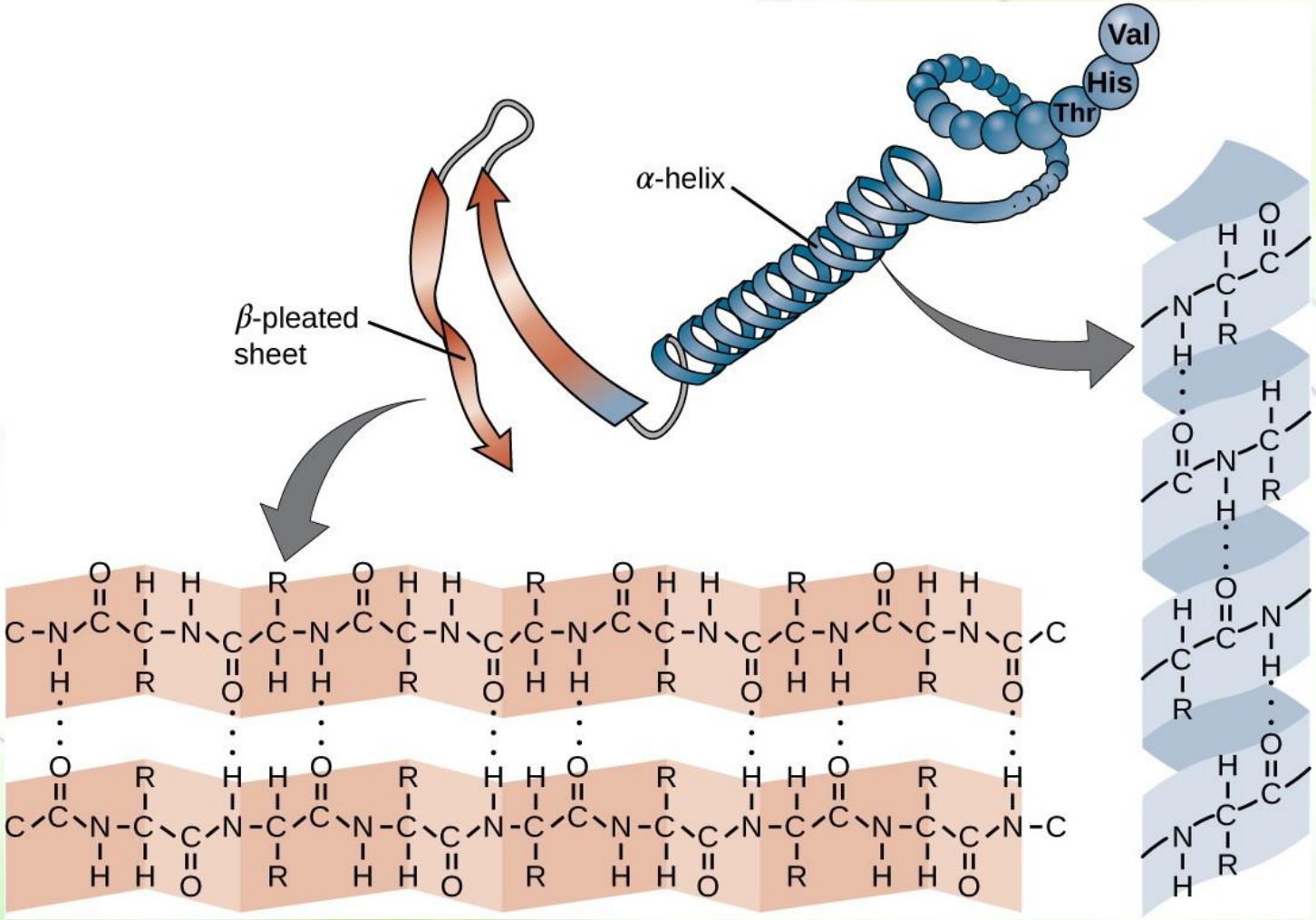






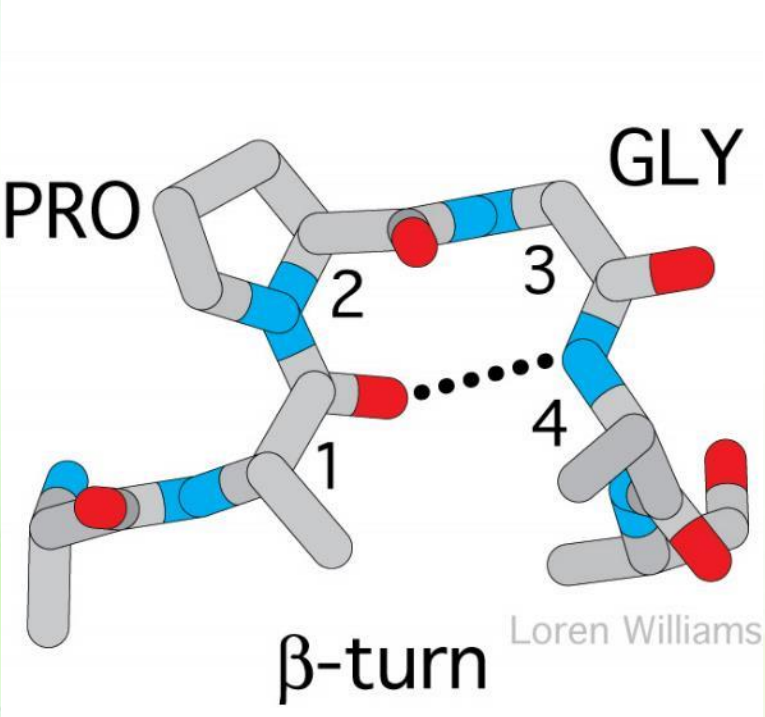
Based on hydrogen bonding pattern, which do you think is more stable: parallel or anti-parallel sheets?

How are they illustrated/drawn?



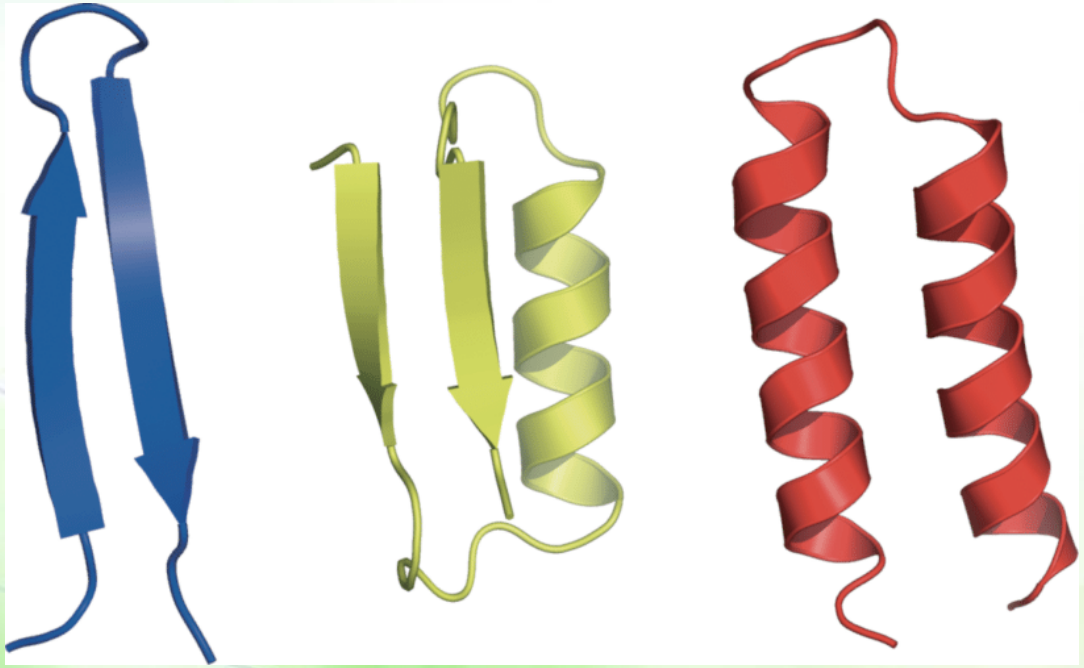
β-turns

- Turns are compact, U-shaped secondary structures.
- They are also known as β turns or hairpin bend.
- What are they used for? How are they stabilized?
- Glycine and proline are commonly present in turns. Why?



Loops and coils

- Loops are a diverse class of secondary structures in proteins with irregular structure and that connect the main secondary structures.
- They are found on surface of molecule and provide flexibility to proteins.
- Amino acids in loops are often not conserved.

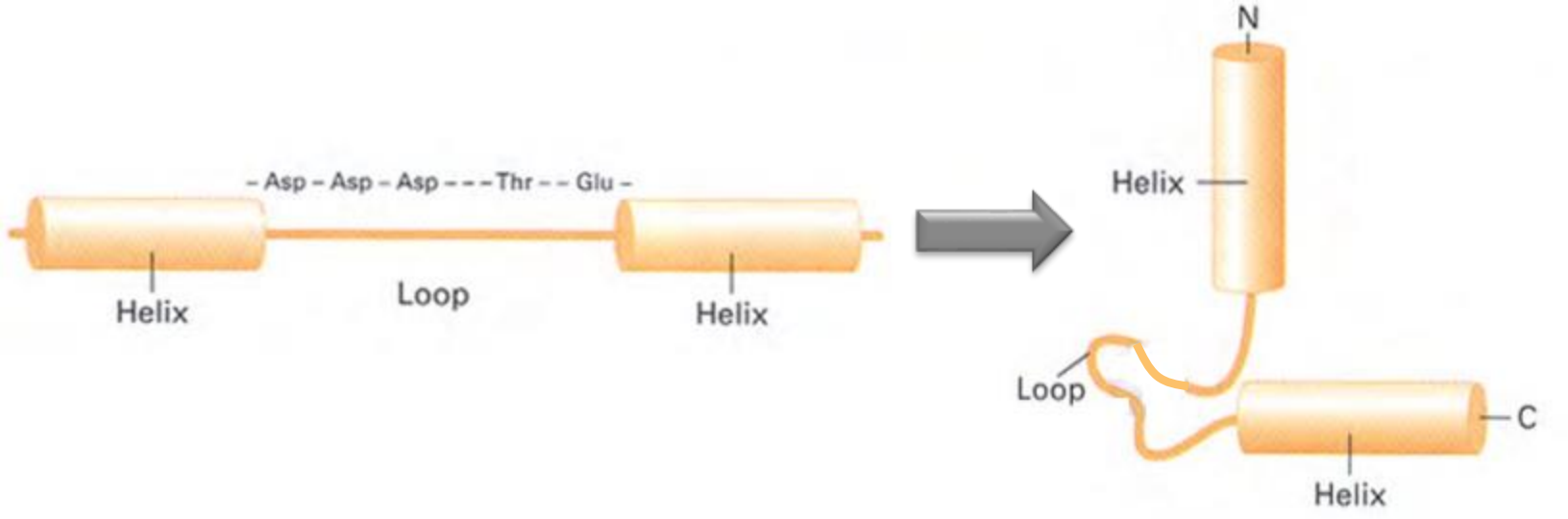


Super-secondary structures

- They are regions in proteins that contain an ordered organization of secondary structures.
- Examples:
  - Motifs
  - Domains

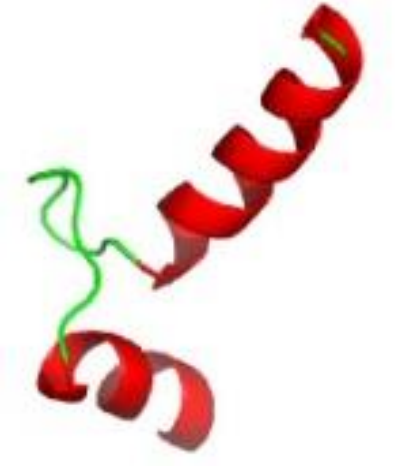
A motif (a module)

- A motif is made of multiple, repetitive or consecutive (connected) secondary structures, that can be small or large.
- They usually constitute a small portion of a protein (typically less than 20 amino acids).
- In general, motifs may provide us with information about the folding of proteins, but not the biological function of the protein.

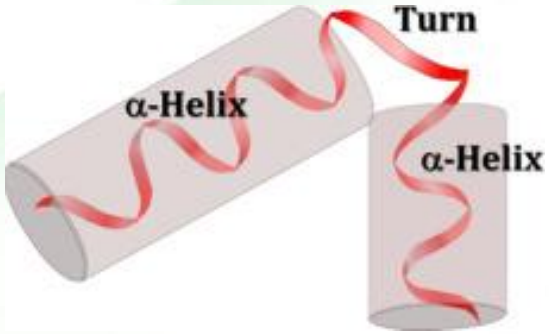


Examples of motifs

Helix-loop-helix: Two α-helices connected by a loop. It is found in DNA-binding proteins.

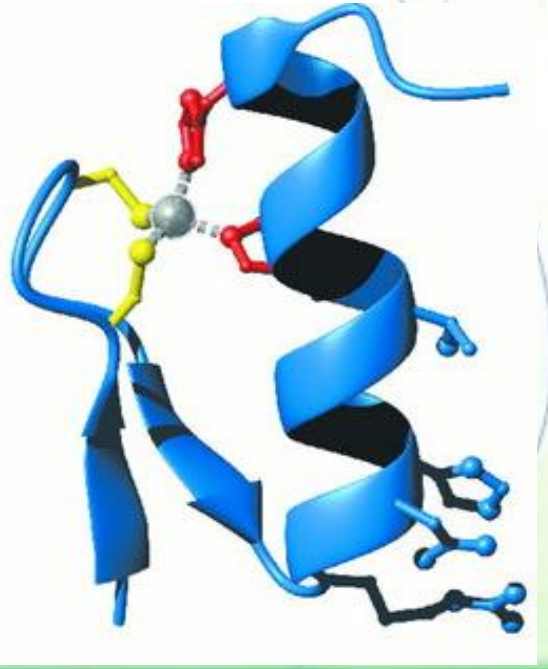


Helix-turn-helix: Two α helices joined by a short strand of amino acids. It is found in DNA-binding proteins.



Beta hairpin: Two antiparallel beta strands connected by a turn.

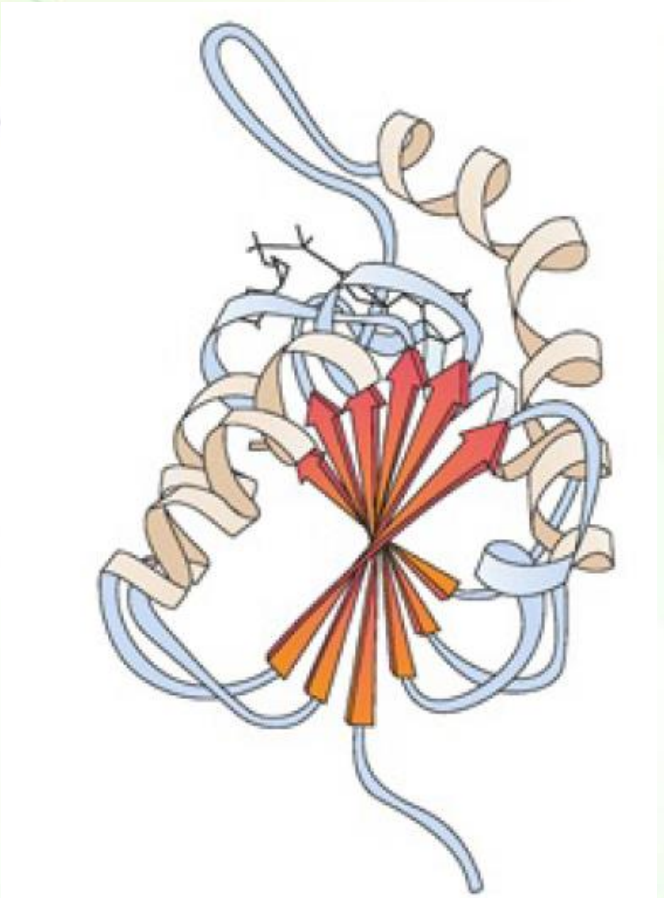
Zinc finger: Two beta strands with an alpha helix end folded over to bind a zinc ion. Important in DNA binding proteins.



Tertiary structure

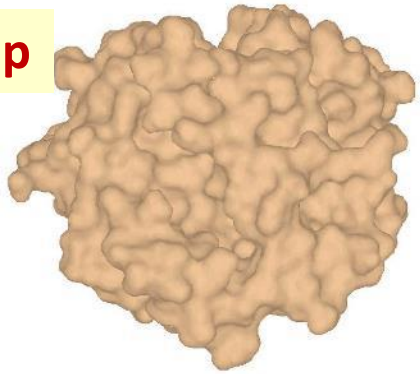
What is tertiary structure?

- The overall conformation of a polypeptide chain
- The three-dimensional arrangement of all the amino acids residues
- The spatial arrangement of amino acid residues that are far apart in the sequence

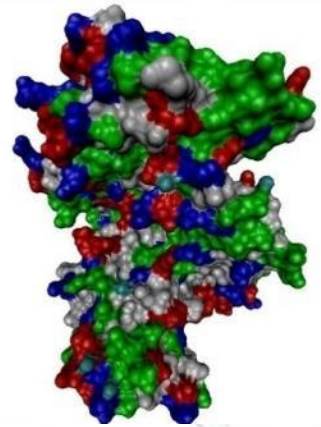


How to look at proteins…

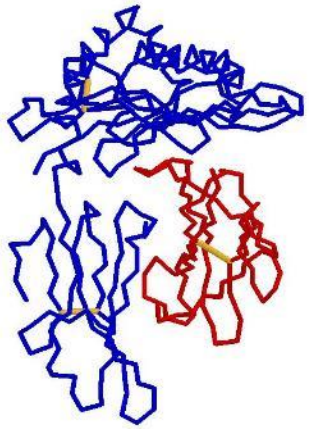
Protein surface map



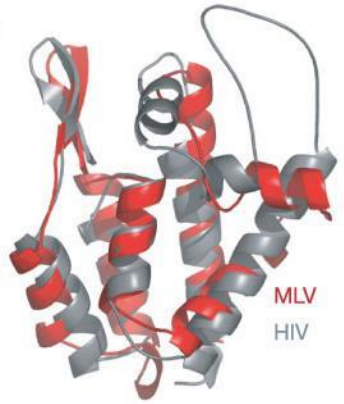
Space-filling structure



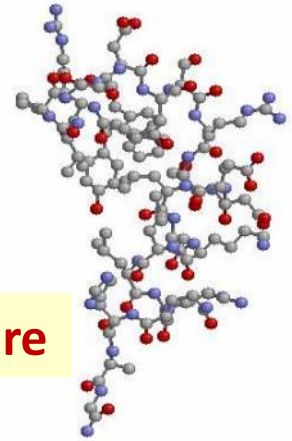
Trace structure



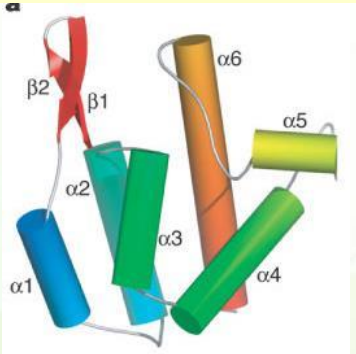
Ribbon structure



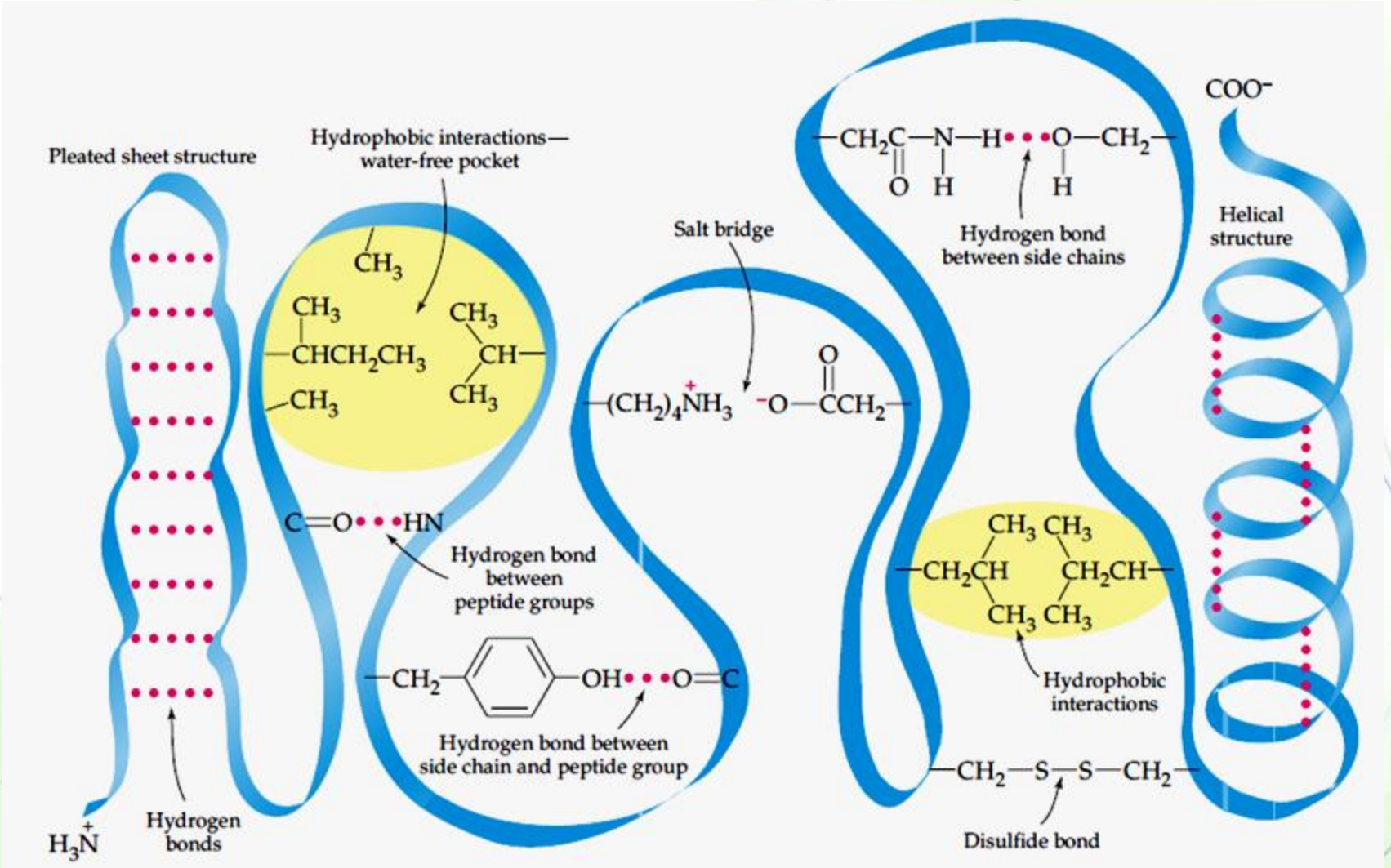
Ball and stick structure



Cylinder structure

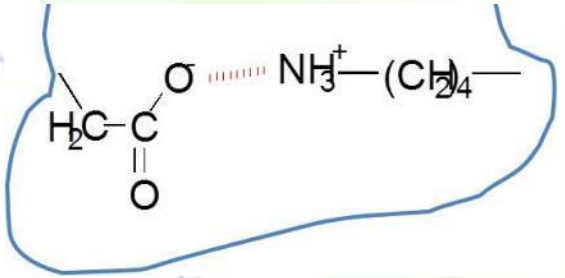


Shape-determining forces

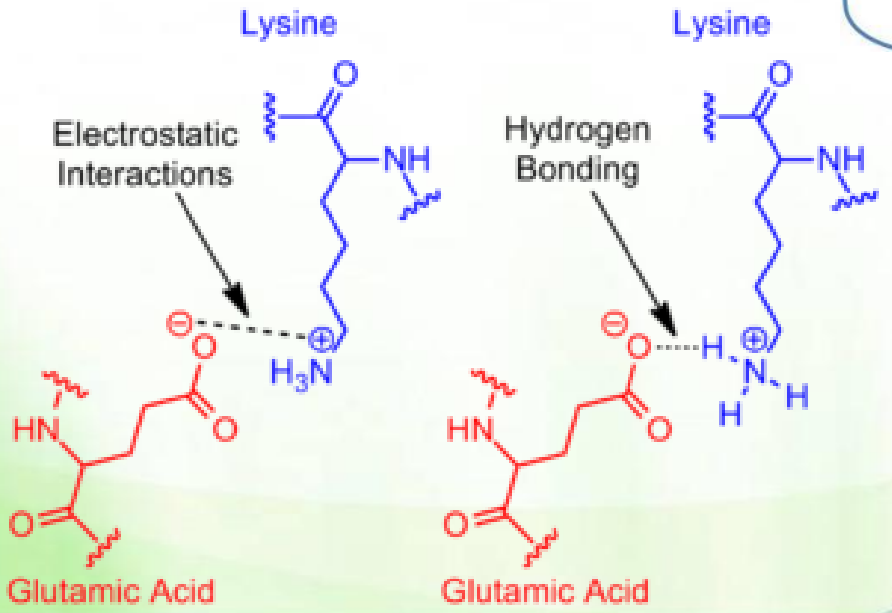


Non-covalent interactions

- Hydrogen bonds occur not only within and between polypeptide chains but with the surrounding aqueous medium.
- Charge-charge interactions (salt bridges) occur between oppositely charged R-groups of amino acids.
  - Charge-dipole interactions form between charged R groups with the partial charges of water.

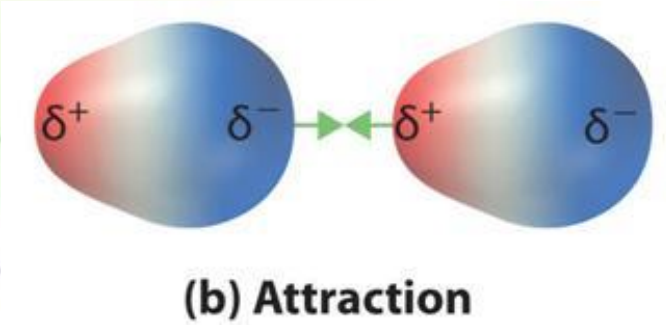


The same charged group can form either hydrogen bonding or electrostatic interactions

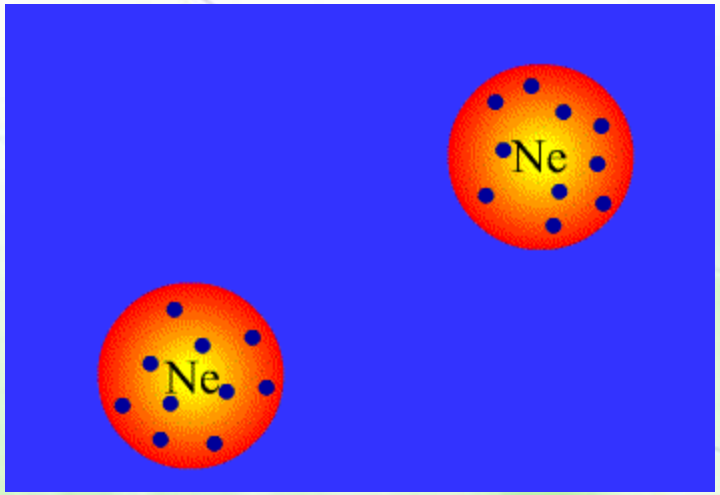


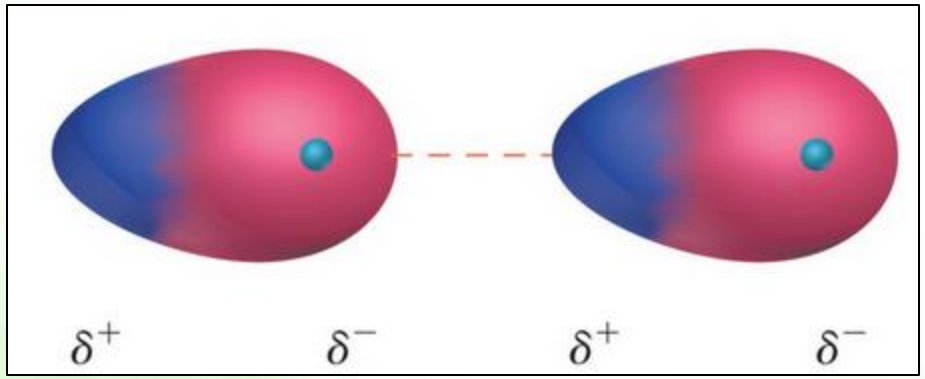
van der Waals attractions

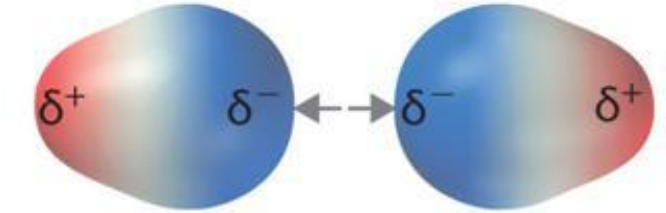
- There are both attractive and repulsive van der Waals forces that control protein folding.
- Although van der Waals forces are extremely weak, they are significant because there are so many of them in large protein molecules.



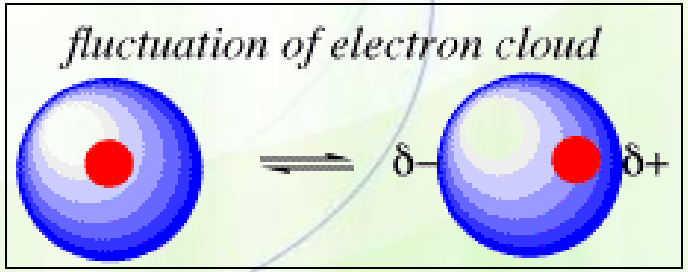
(b) Attraction





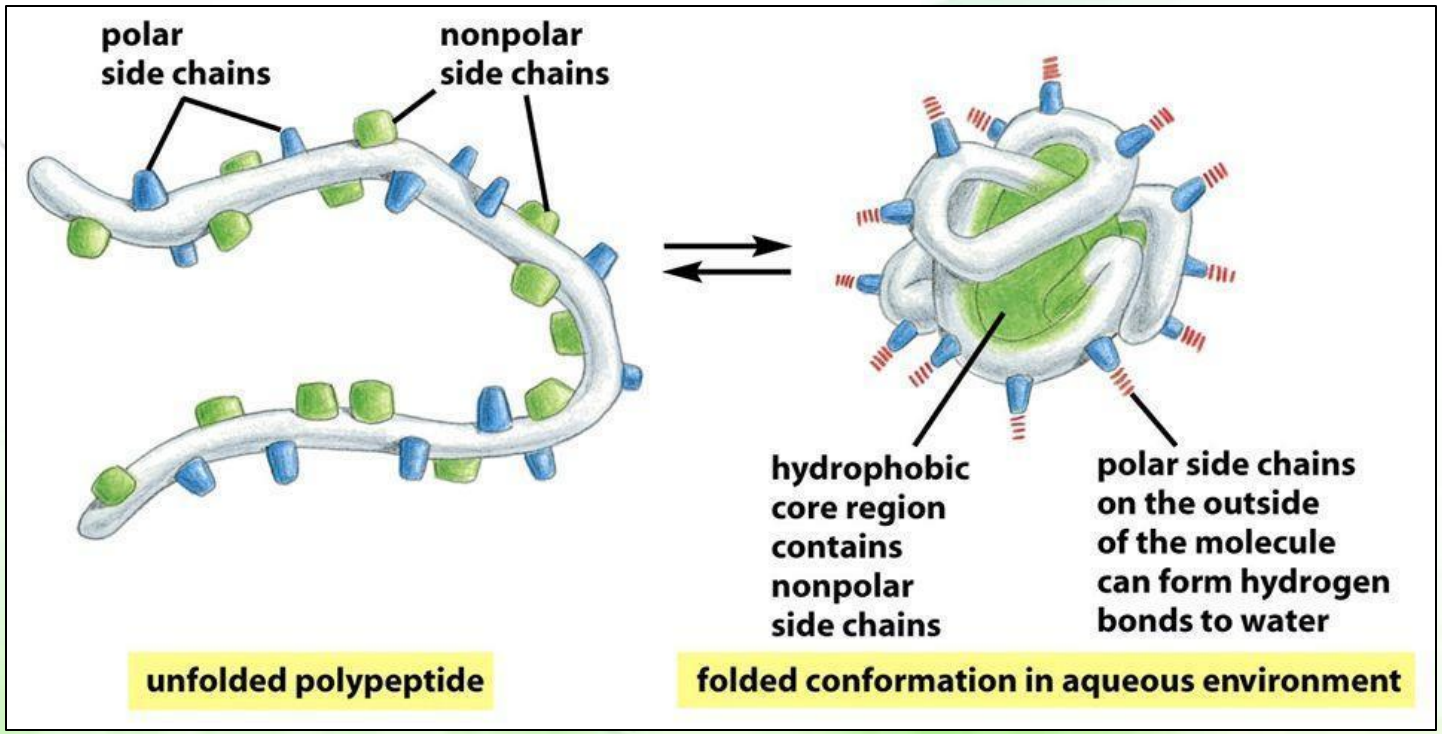


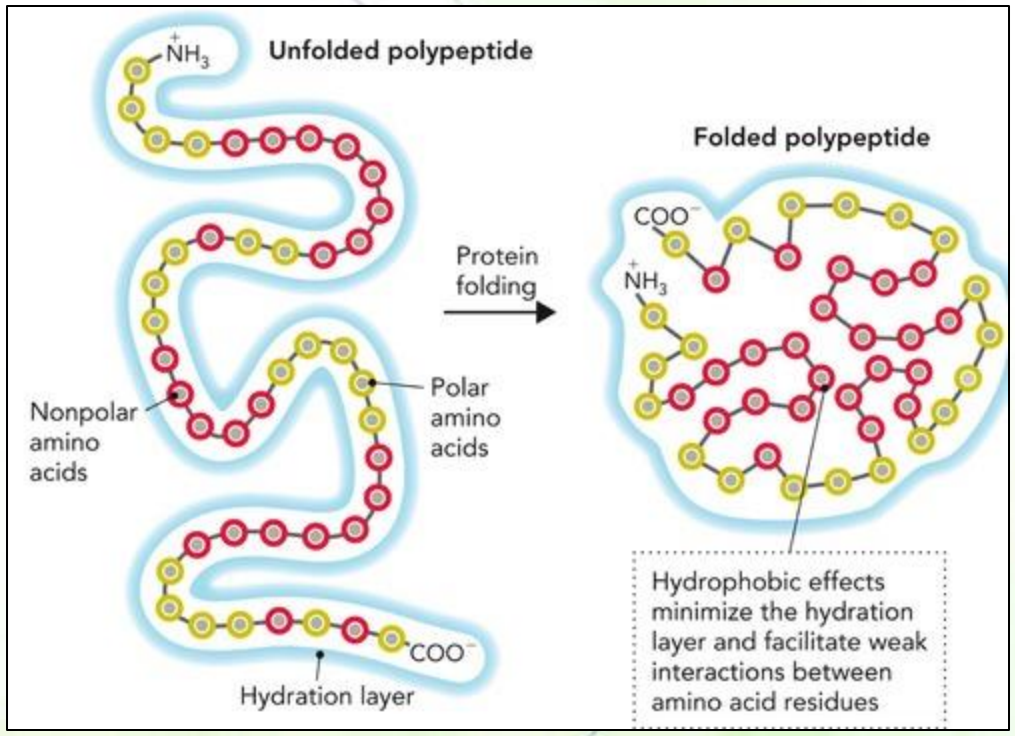
(d) Repulsion



Hydrophobic interactions

- A system is more thermodynamically (energetically) stable when hydrophobic groups are clustered together rather than extended into the aqueous surroundings.

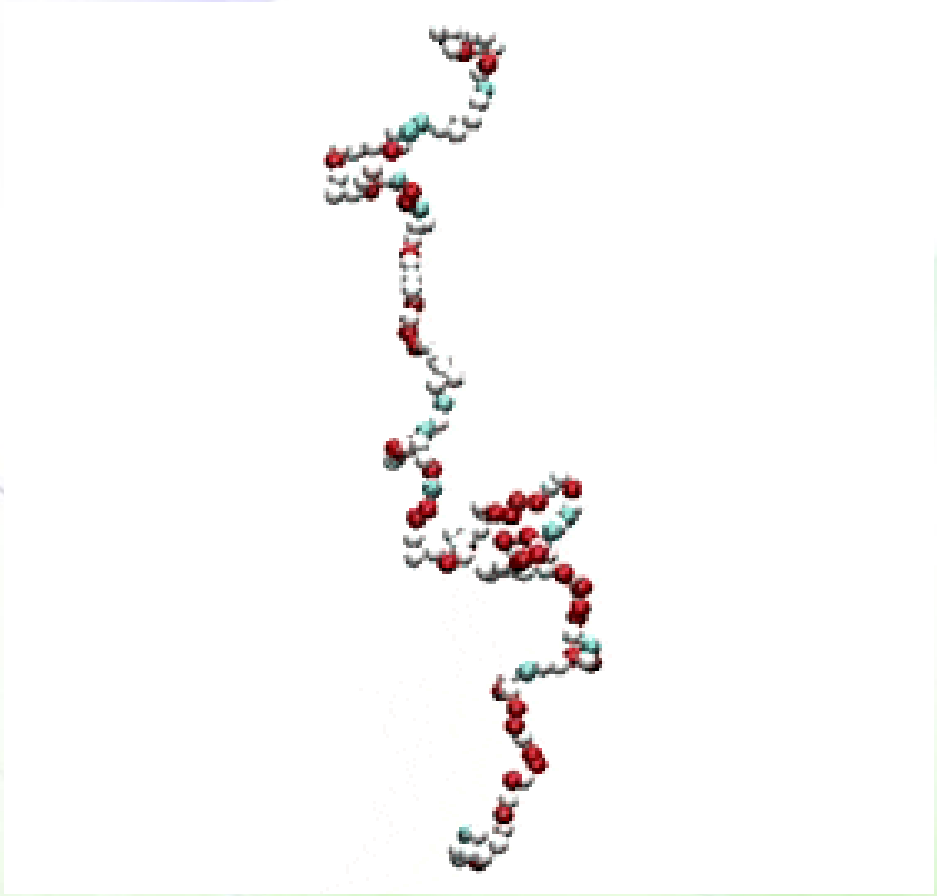


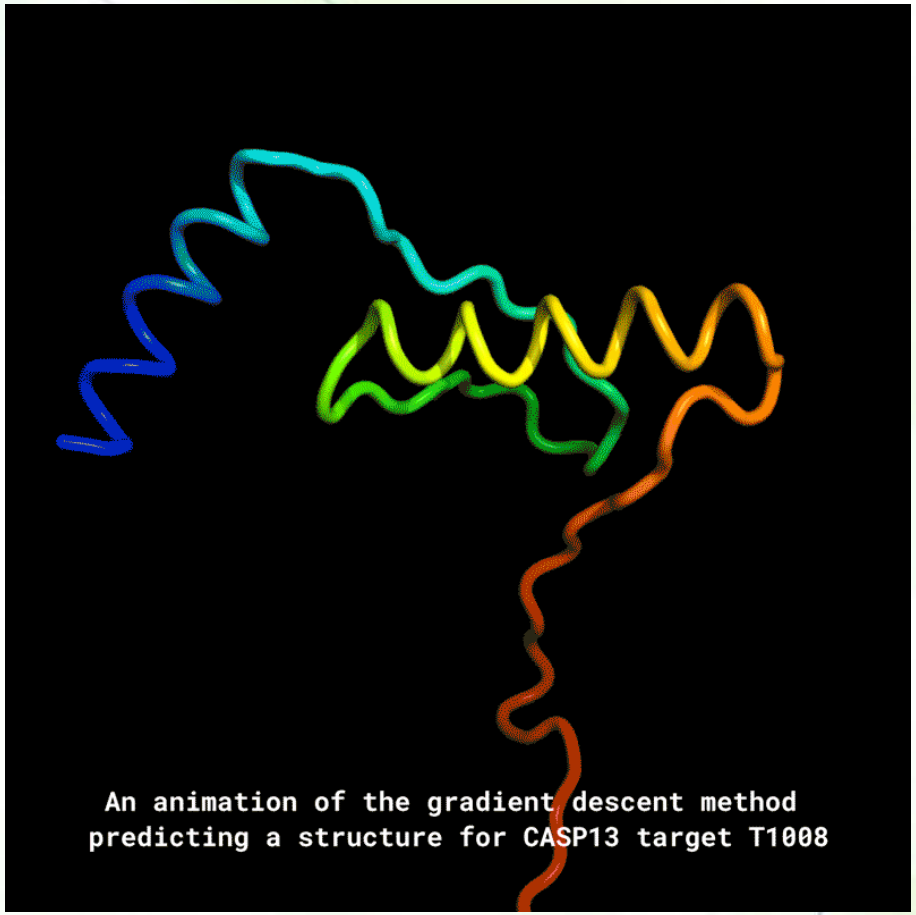


Can polar amino acids be found in the interior?...YES

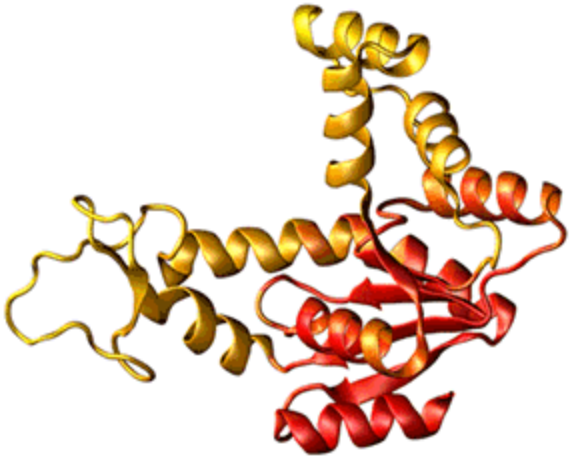
- Polar amino acids can be found in the interior of proteins
- In this case, they form hydrogen bonds to other amino acids or to the polypeptide backbone
- They play important roles in the function of the protein

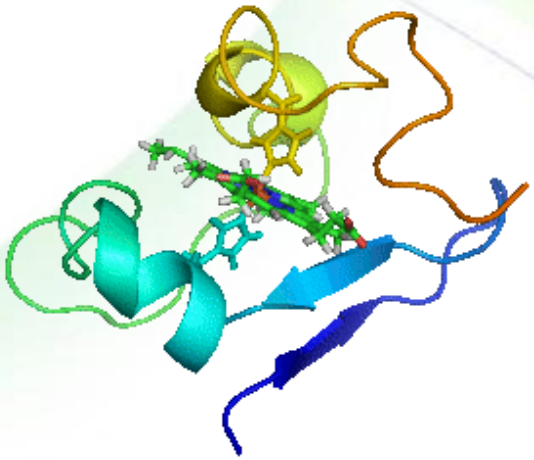
A hypothetical look at protein folding

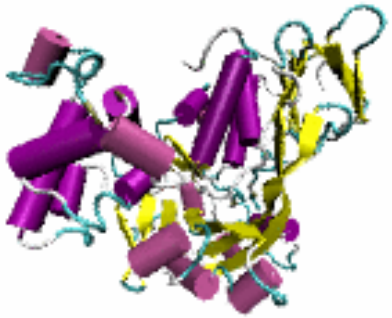


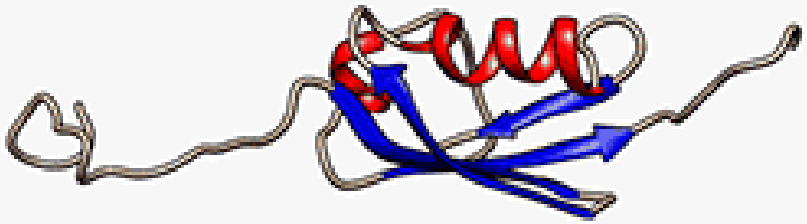


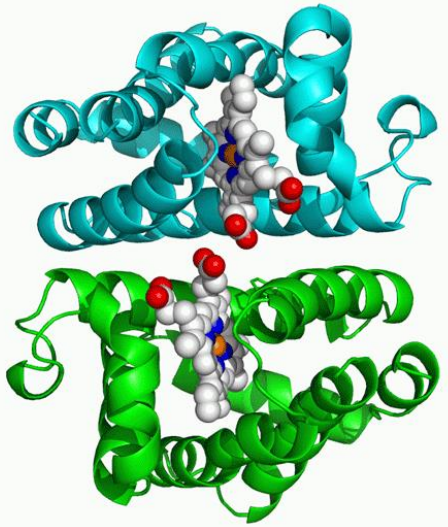
Protein are NOT static











Stabilizing factors

- There are two forces that do not determine the three-dimensional structure of proteins, but stabilize these structures:
  - Disulfide bonds
  - Metal ions

Disulfide bonds

- The side chain of cysteine contains a reactive sulfhydryl group (—SH), which can oxidize to form a disulfide bond (—S—S—) to a second cysteine.
- The crosslinking of two cysteines to form a new amino acid, called cystine.

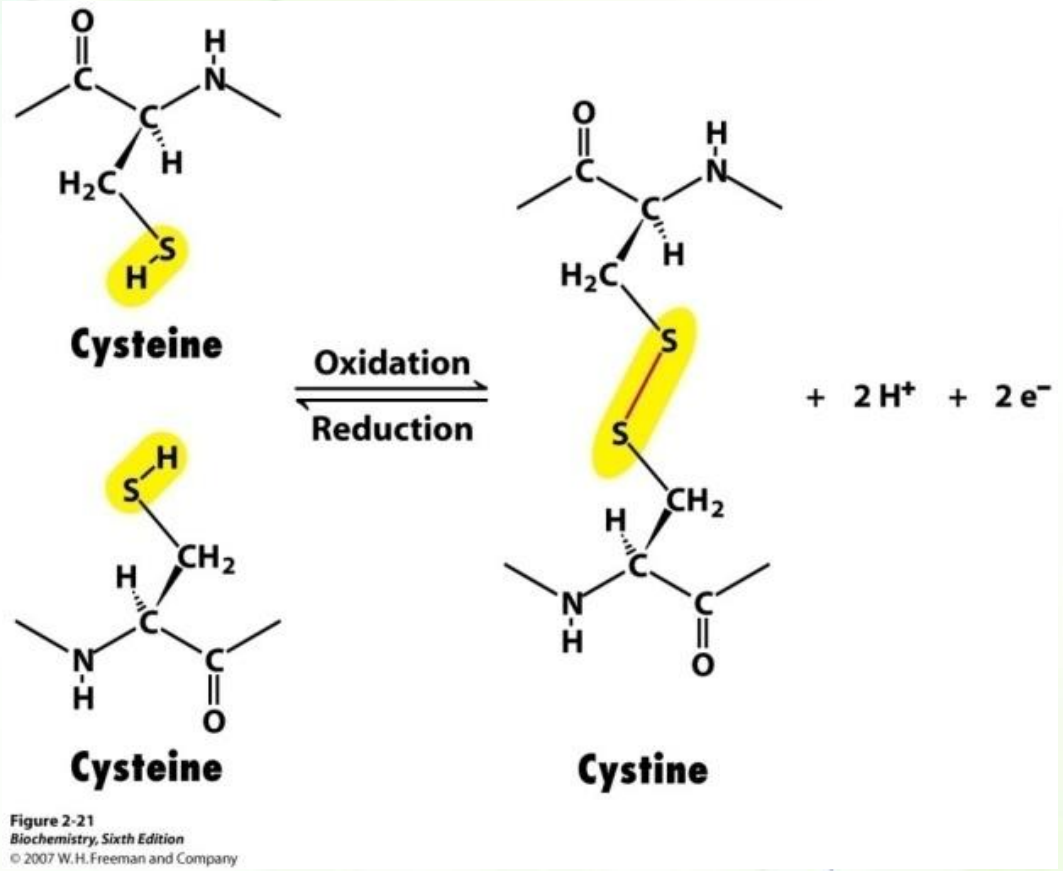
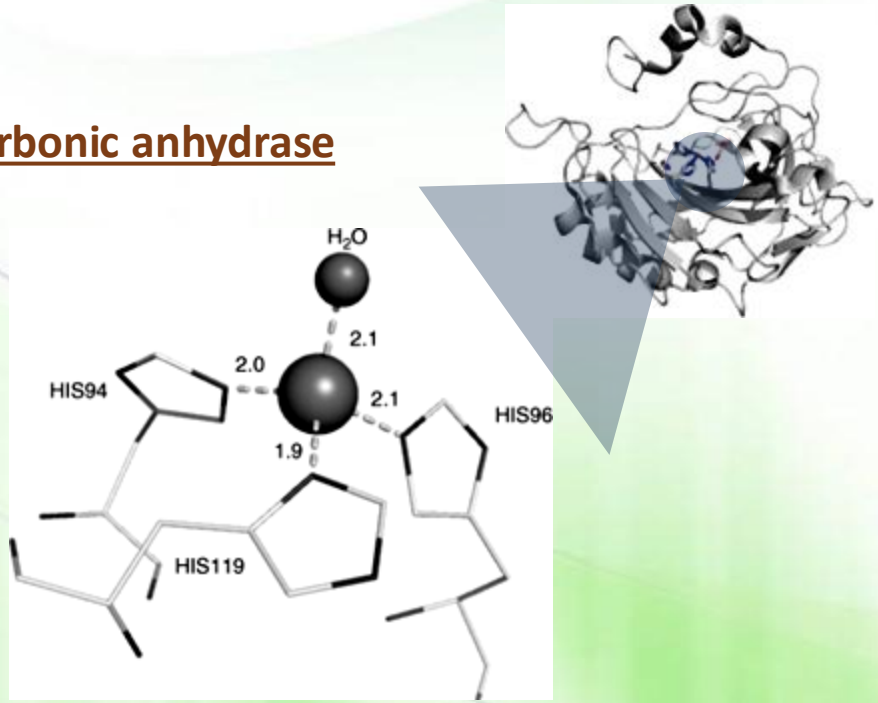


Figure 2-21
Biochemistry, Sixth Edition

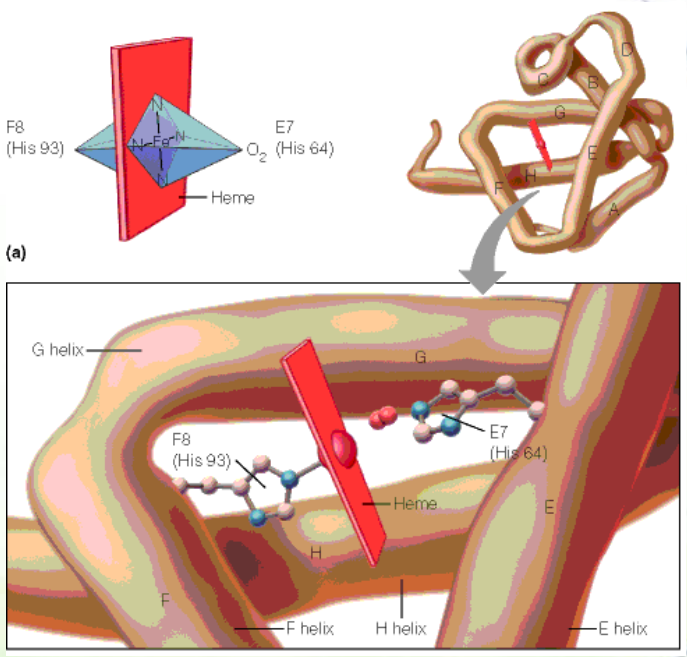
metal ions

- Several proteins can be complexed to a single metal ion that can stabilize protein structure by forming:
  - Covalent interaction (myoglobin)
  - Salt bridges (carbonic anhydrase)

Carbonic anhydrase

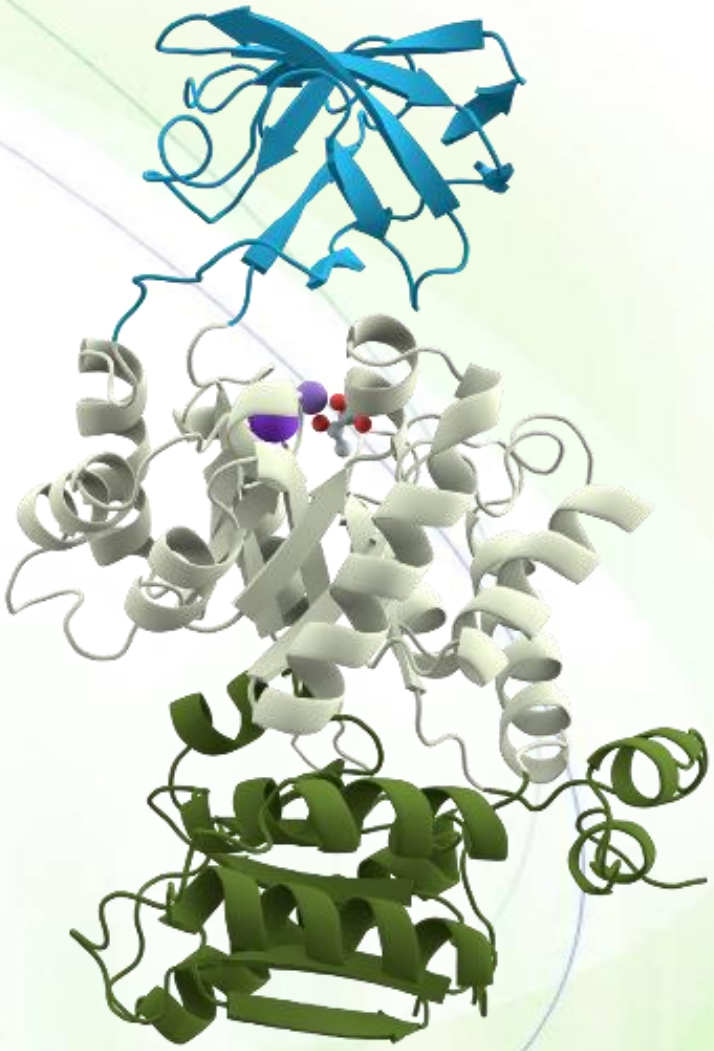


Myoglobin



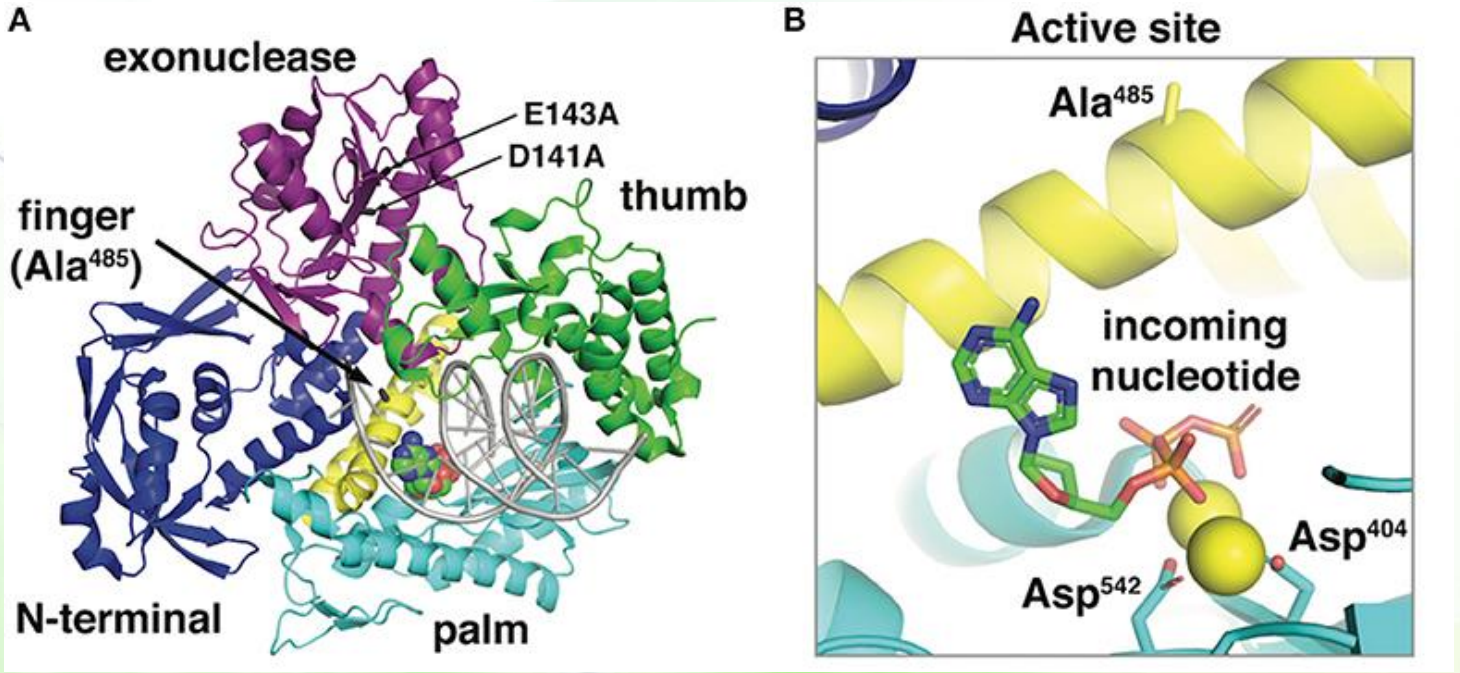
Domains

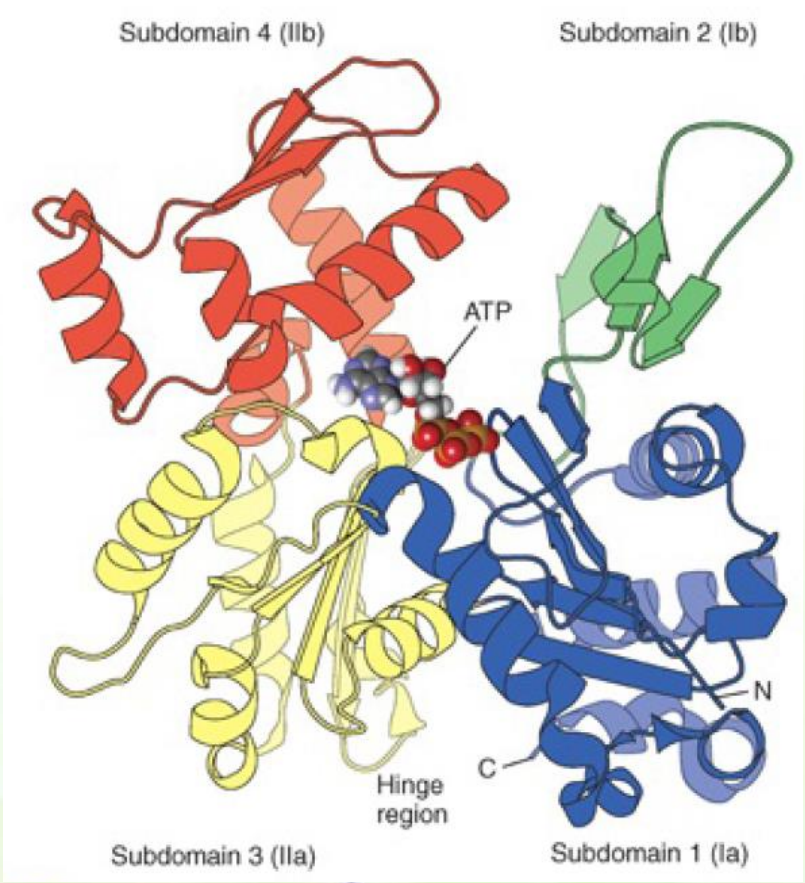
- A domain is a combination of α helices and/or β sheets that are connected to each other via turns, loops, and coils and are organized in a specific three-dimensional structure.
  - A domain may consist of 100–200 residues.
- Domains fold independently of the rest of the protein or of other domains within the same protein.
- Similar domains can be found in proteins with similar function and/or structure and can be present in different proteins
- Domains may also be defined in functional terms
  - Enzymatic activity
  - Binding ability (e.g., a DNA-binding domain)



Folds

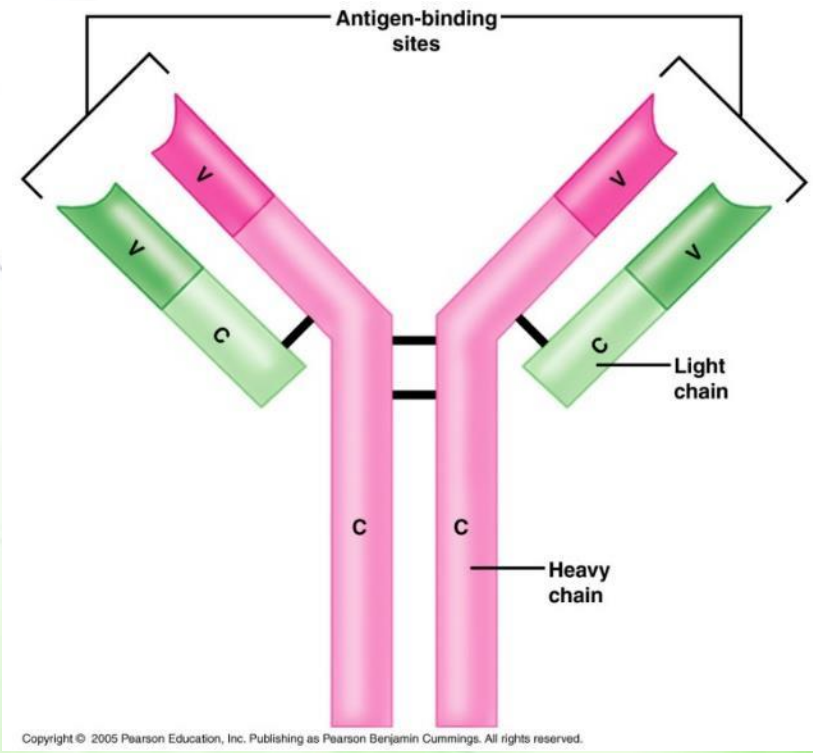
- When large patterns of secondary structures or multiple domains within a protein possess specific functions, they are known as **Folds**.
  - The actin fold
  - The nucleotide-binding fold

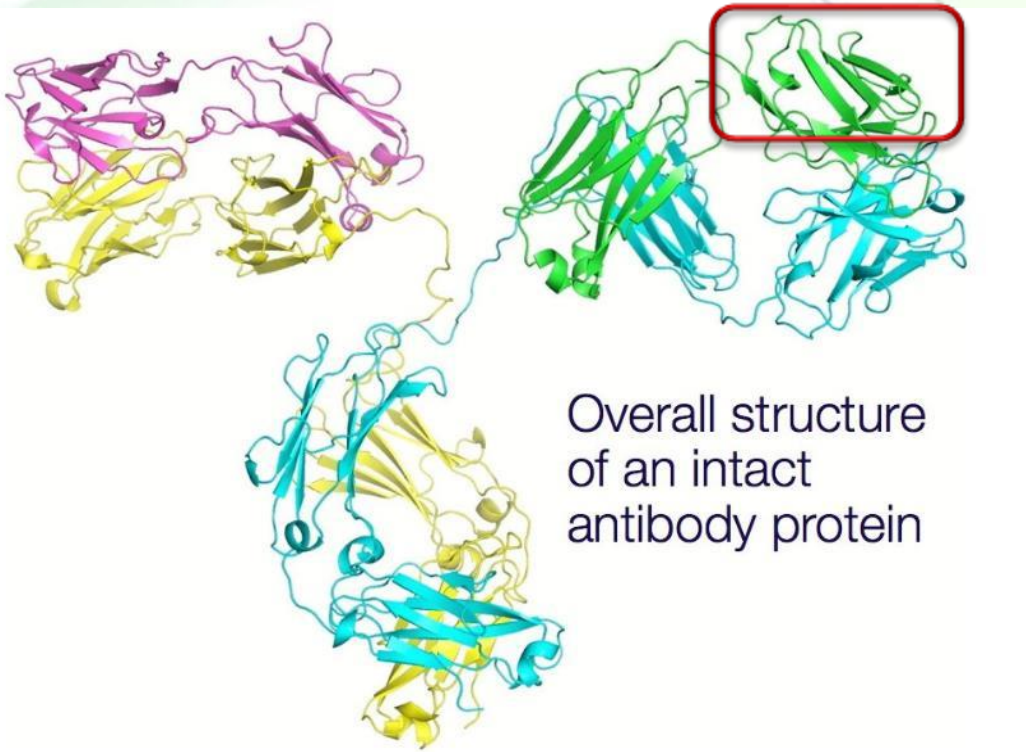


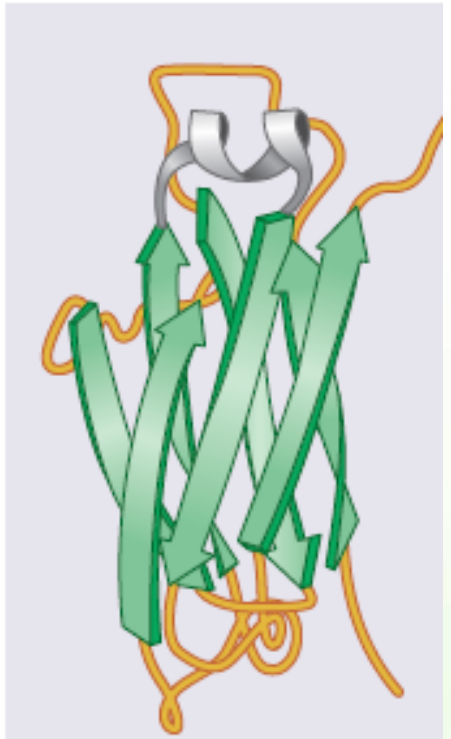


A more complex motif/domain/fold is…

- The immunoglobulin fold or module that enables interaction with molecules of various structures and sizes.

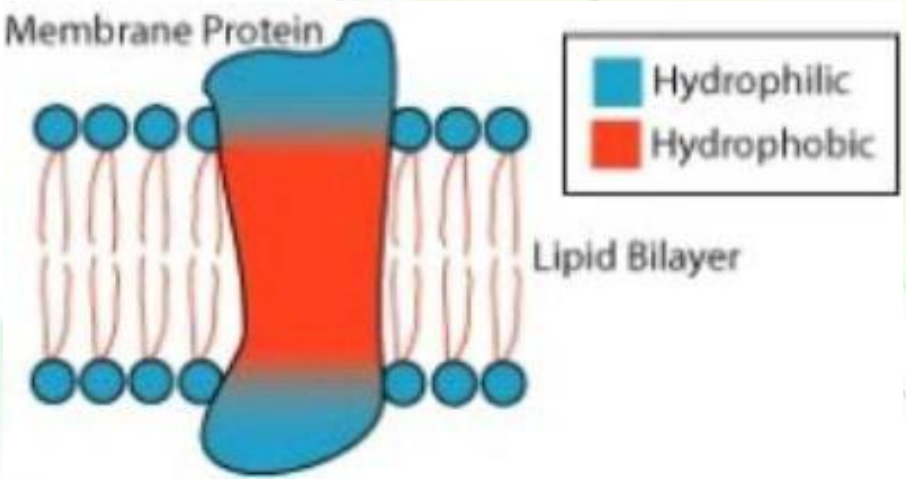


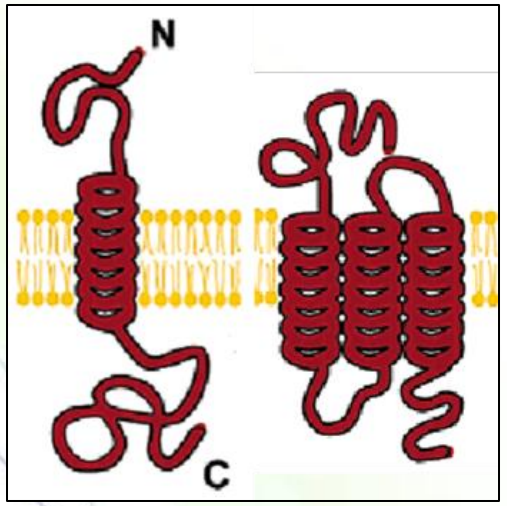


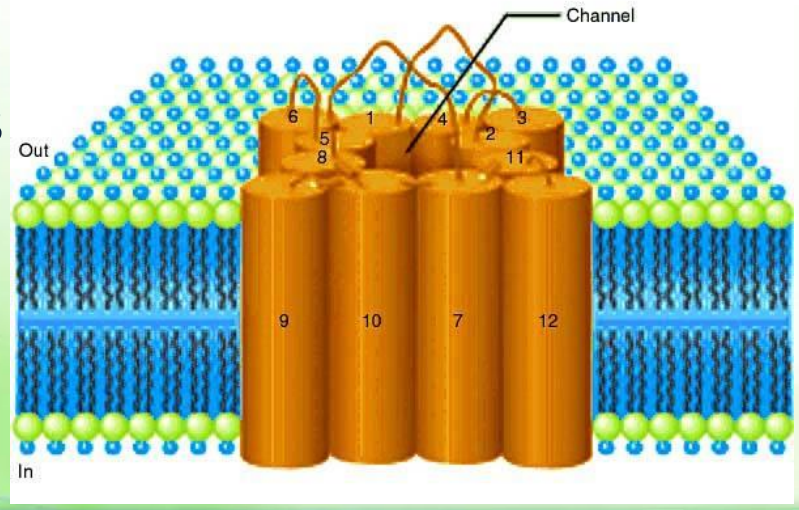


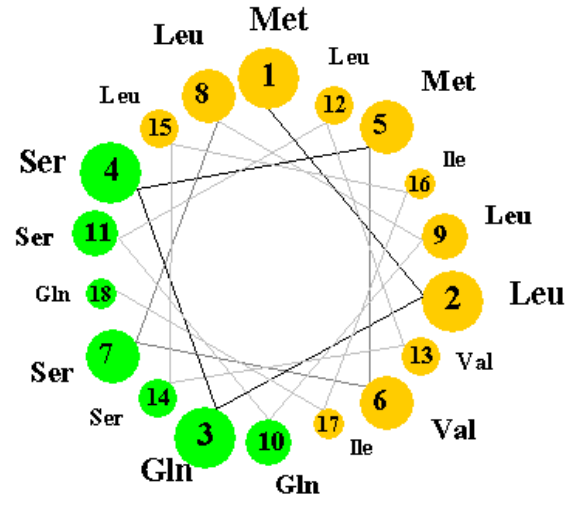
α-helices as transmembrane domains

- Membrane-spanning proteins contain a transmembrane domain that is an α–helix made of hydrophobic amino acids.
- The α–helices of proteins with multiple transmembrane domains are connected by loops containing hydrophilic amino acids located outside of the membrane.
- Membrane ion channel proteins contain amphipathic α–helices.









Quaternary structure

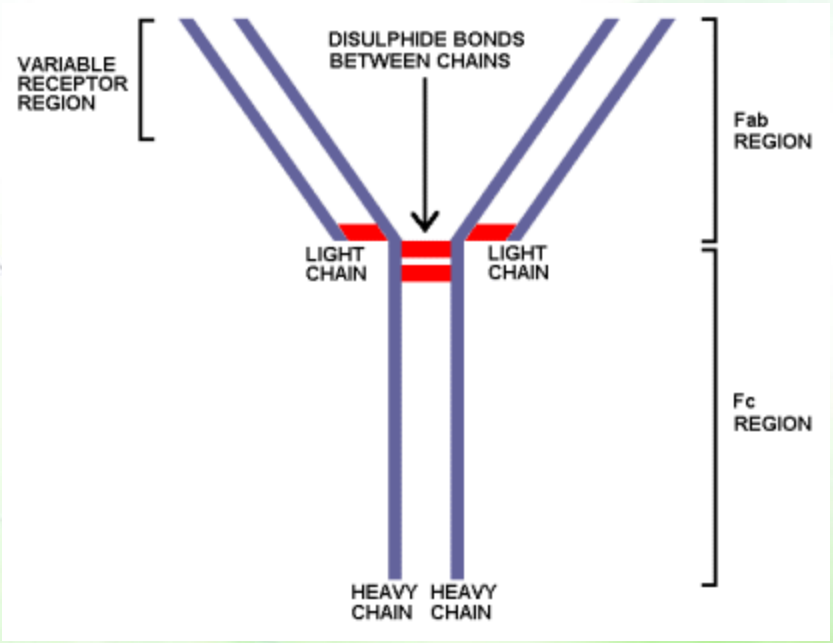
What is it?

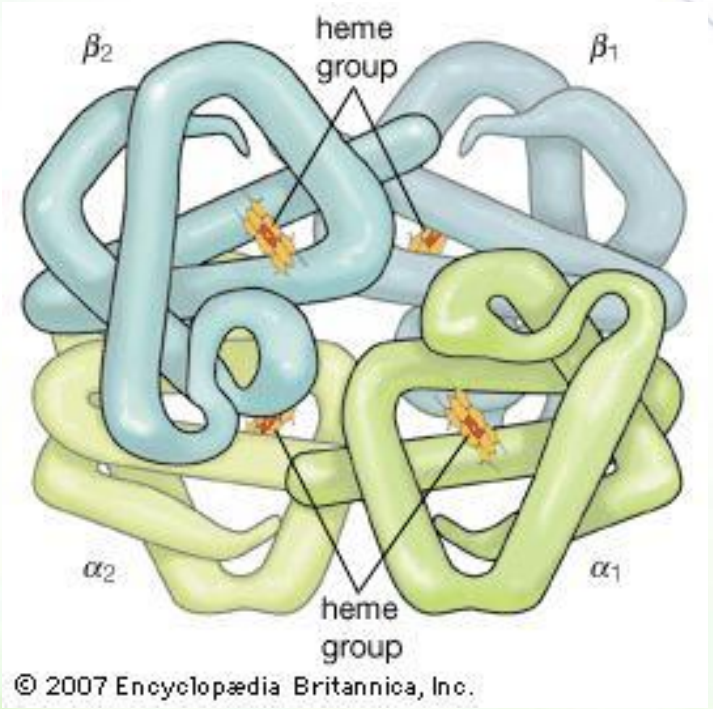
- Proteins have a quaternary structure if they are composed of more than one polypeptide chain.
  - They are oligomeric proteins (oligo = a few or small or short; mer = part or unit)
- *A quaternary structure Is the spatial arrangement of subunits and the nature of their interactions.*
- A protein can be a:
  - Monomer: One subunit
  - Dimer: Two subunits
    - The simplest: a homodimer
  - A trimer: Three subunits
  - A tetramer: Four subunits
  - …etc

- Each polypeptide chain is called a subunit.
- Oligomeric or multimeric proteins are made of multiple polypeptides that are
  - identical → homooligomers (homo = same)
  - different → heterooligomers (hetero = different)

How are the subunits connected?

- Sometimes subunits are disulfide-bonded together, other times, noncovalent bonds stabilize interactions between subunits



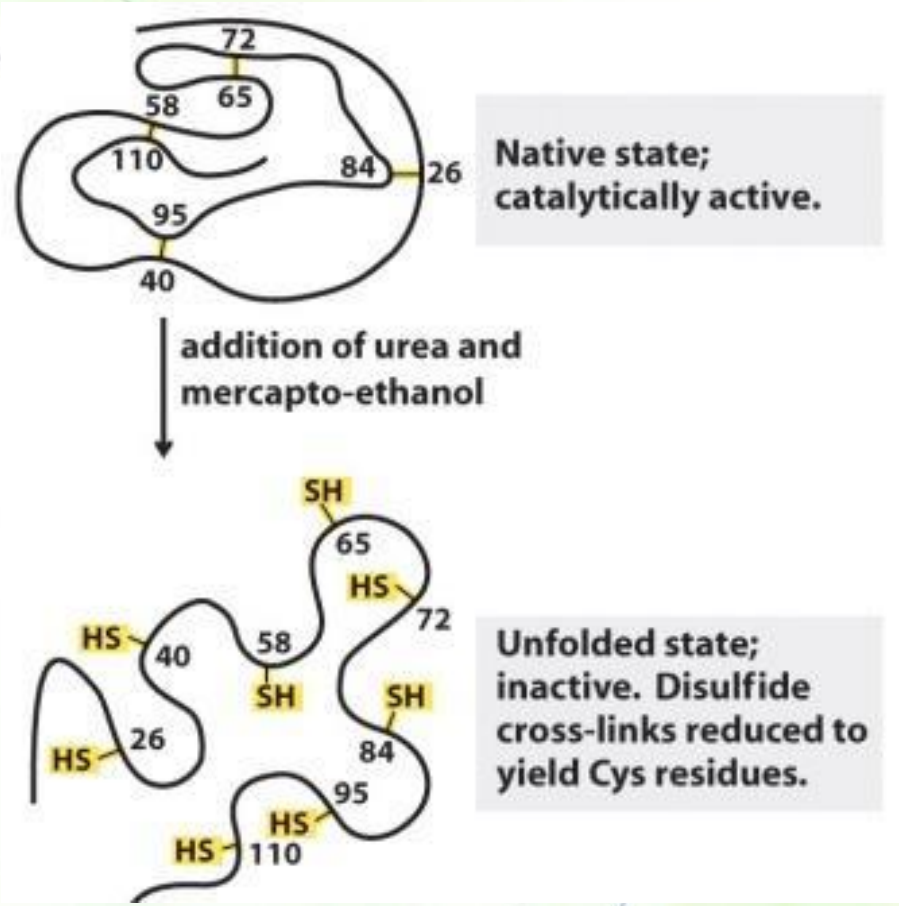


Properties of Proteins:

Denaturation and Renaturation

Denaturation

- Denaturation is the disruption of the native conformation of a protein via breaking the noncovalent bonds that determine the structure of a protein
- Complete disruption of tertiary structure is achieved by reduction of the disulfide bonds in a protein
- The denatured protein loses its properties such as activity and become insoluble.



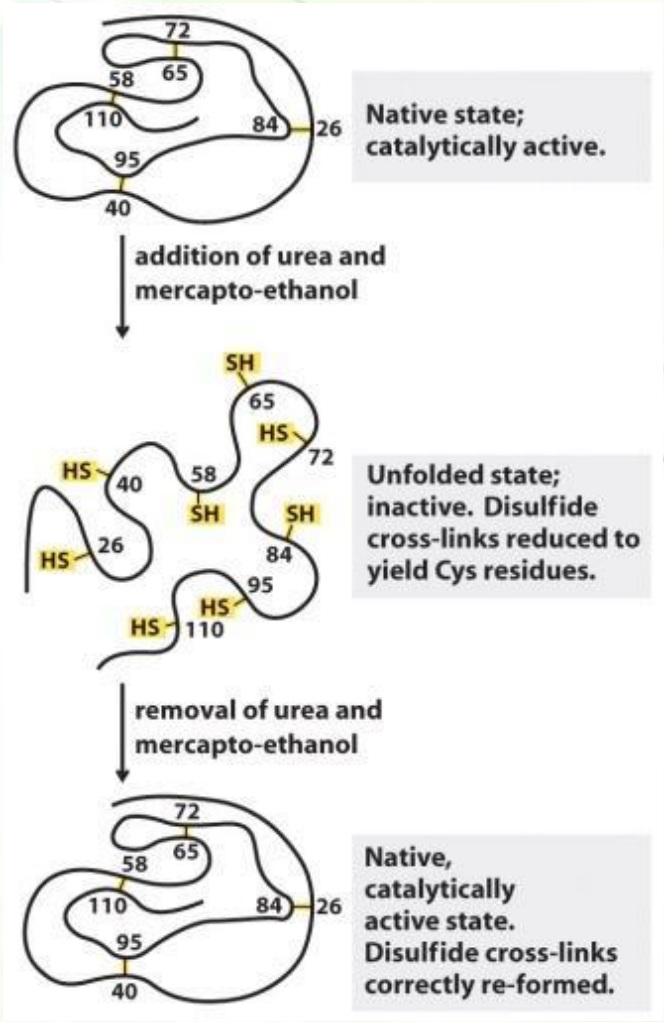
Denaturing agents

- Heat disrupts low-energy van der Waals forces in proteins
- Extremes of pH: change in the charge of the protein's amino acid side chains (electrostatic and hydrogen bonds).
- Detergents: Triton X-100 (nonionic, uncharged) and sodium dodecyl sulfate (SDS, anionic, charged) disrupt the hydrophobic forces.
  - SDS also disrupt electrostatic interactions.
- Urea and guanidine hydrochloride disrupt hydrogen bonding and hydrophobic interactions.

- Reducing agents: β-mercaptoethanol (β-ME) and dithiothreitol (DTT)
  - Both reduce disulfide bonds.

Renaturation

- Renaturation is the process in which the native conformation of a protein is re-acquired.
- Renaturation can occur quickly and spontaneously, and disulfide bonds are formed correctly.
- If a protein is unfolded, it can refold to its correct structure placing the S-S bonds in the right orientation (adjacent to each other prior to formation), then the correct S-S bonds are reformed.
- This is particularly true for small proteins.

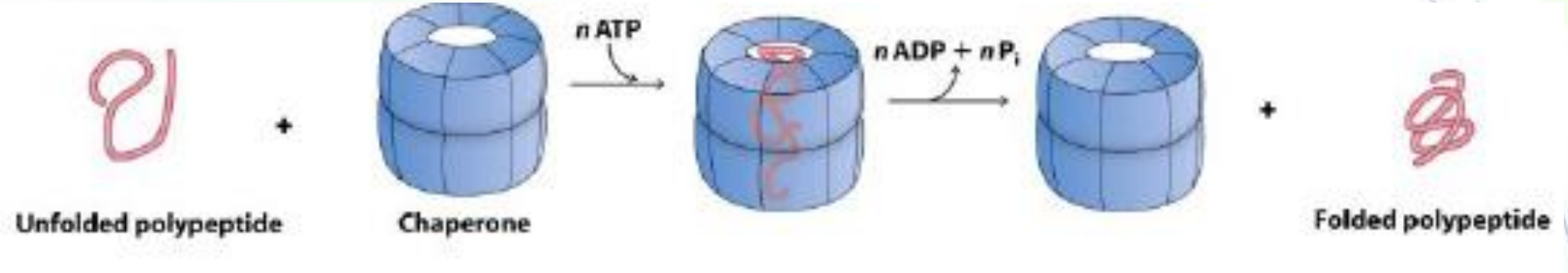


Factors that determine protein structure

- The least amount of energy needed to stabilize the protein. This is determined by:
  - The amino acid sequence (the primary structure), mainly the internal residues.
  - The proper angles between the amino acids
    - The different sets of weak noncovalent bonds that form between the mainly the R groups.
    - Non-protein molecules.

Problem solvers: chaperones

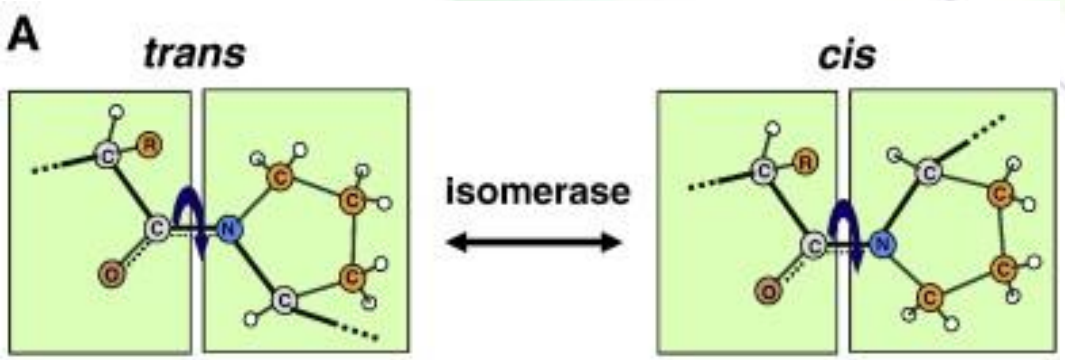
- These proteins bind to polypeptide chains and help them fold with the most energetically favorable folding pathway.
- Chaperones also prevent the hydrophobic regions in newly synthesized protein chains from associating with each other to form protein aggregates .



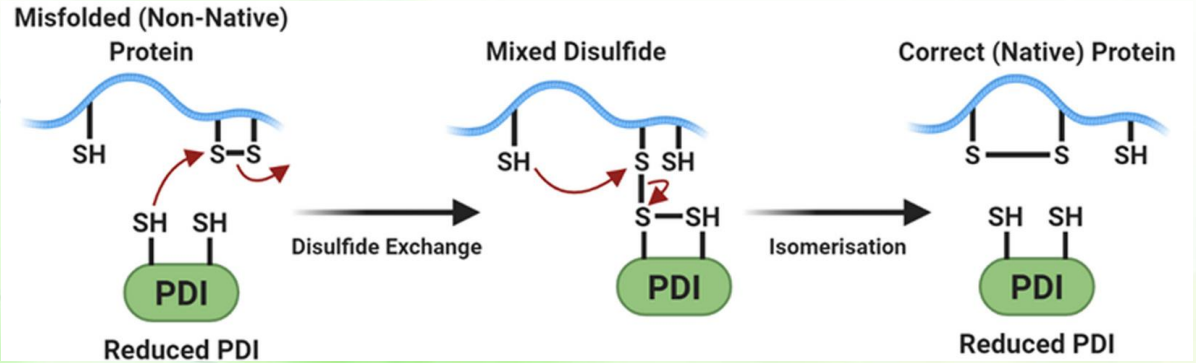
Many diseases are the result of defects in protein folding.

Other players

- A cis–trans isomerase converts a trans peptide bond preceding a proline into the cis conformation, which is well-suited for making hairpin turns.



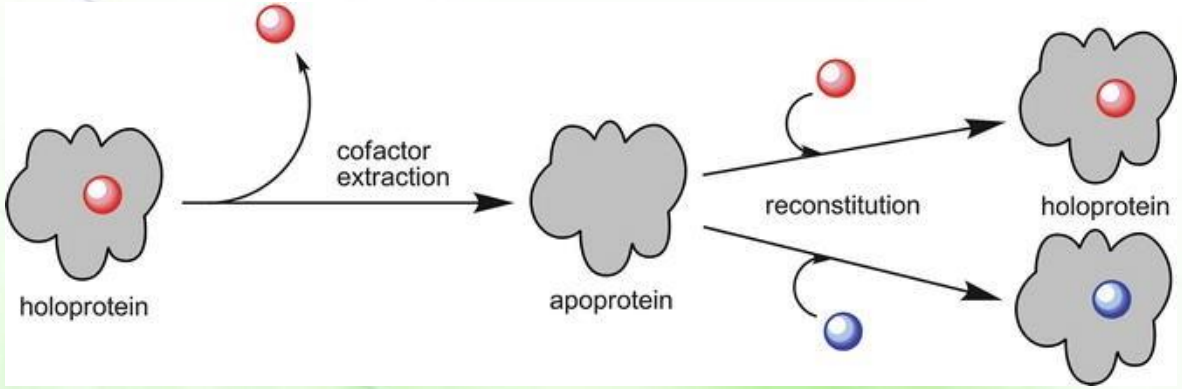
- A protein disulfide isomerase, after the protein has folded, breaks and reforms disulfide bonds between the –SH groups of two cysteine residues.



Complex protein structures

Holo- and apo-proteins

- When a protein is conjugated to any associated non-protein components, such as prosthetic groups or metal ions, the protein is known as a **holoprotein** (also known as (AKA) a conjugated protein).
- If the non-protein component is removed, the protein is known as an **apoprotein**.
  - In other words, it is the protein portion of a conjugated protein without the attached non-protein group.

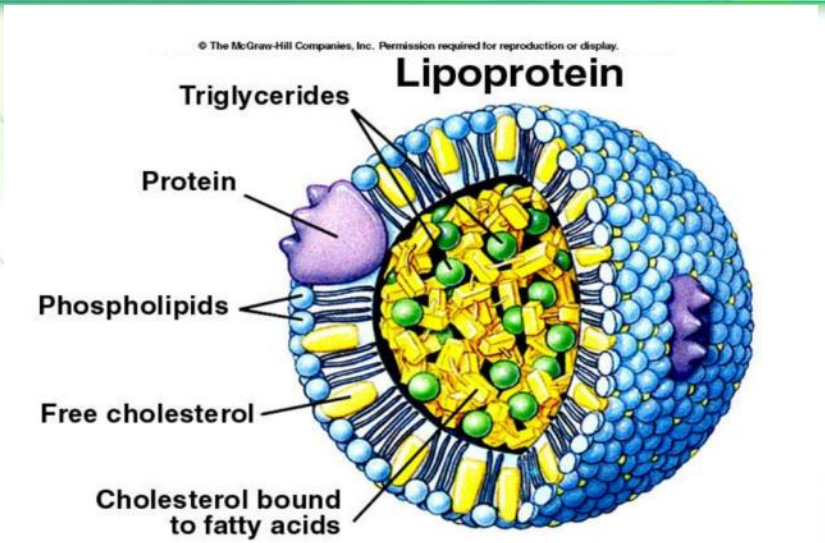


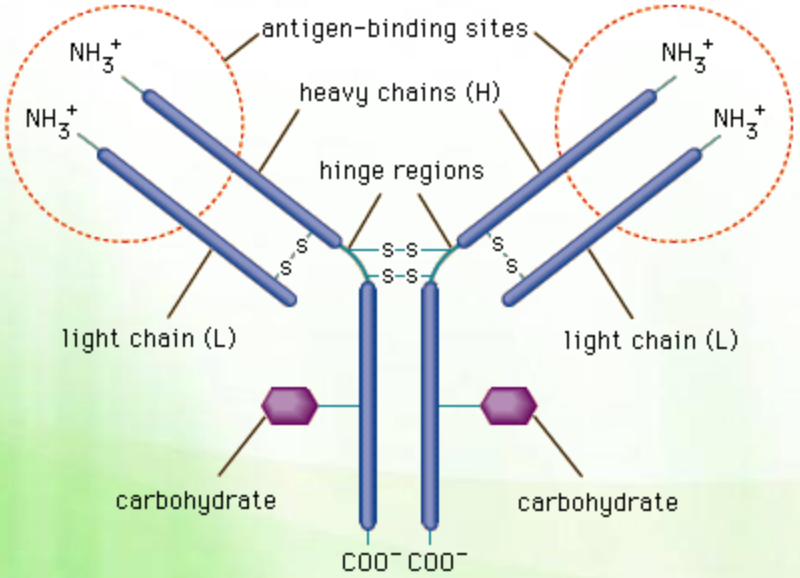
Coenzymes: complex organic molecules that assist enzymes in catalyzing biochemical reactions

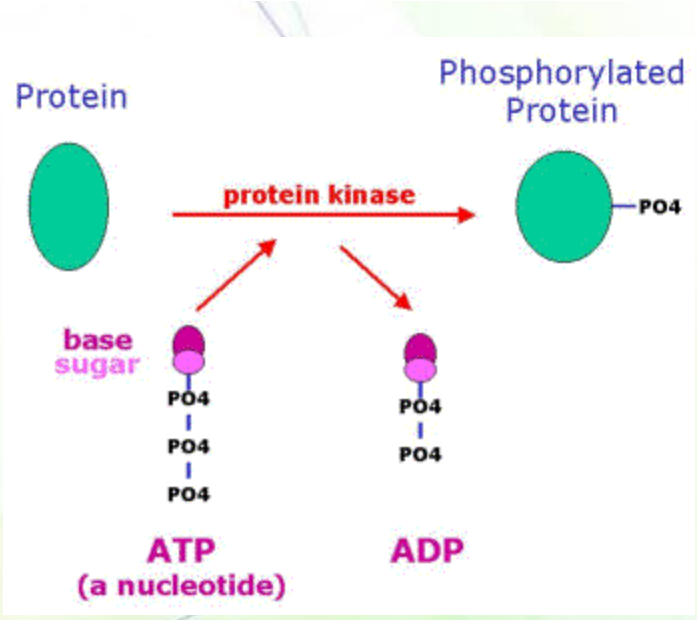
Prosthetic groups: Coenzymes or metals that are tightly (covalently) bound to proteins

Other names of conjugated proteins

- Lipoproteins: Proteins associated with lipids
- Phosphoproteins: proteins that are phosphorylated
- Hemoproteins: proteins with heme
- Nucleoproteins: proteins with a nucleic acid
- Glycoproteins: proteins with carbohydrate groups

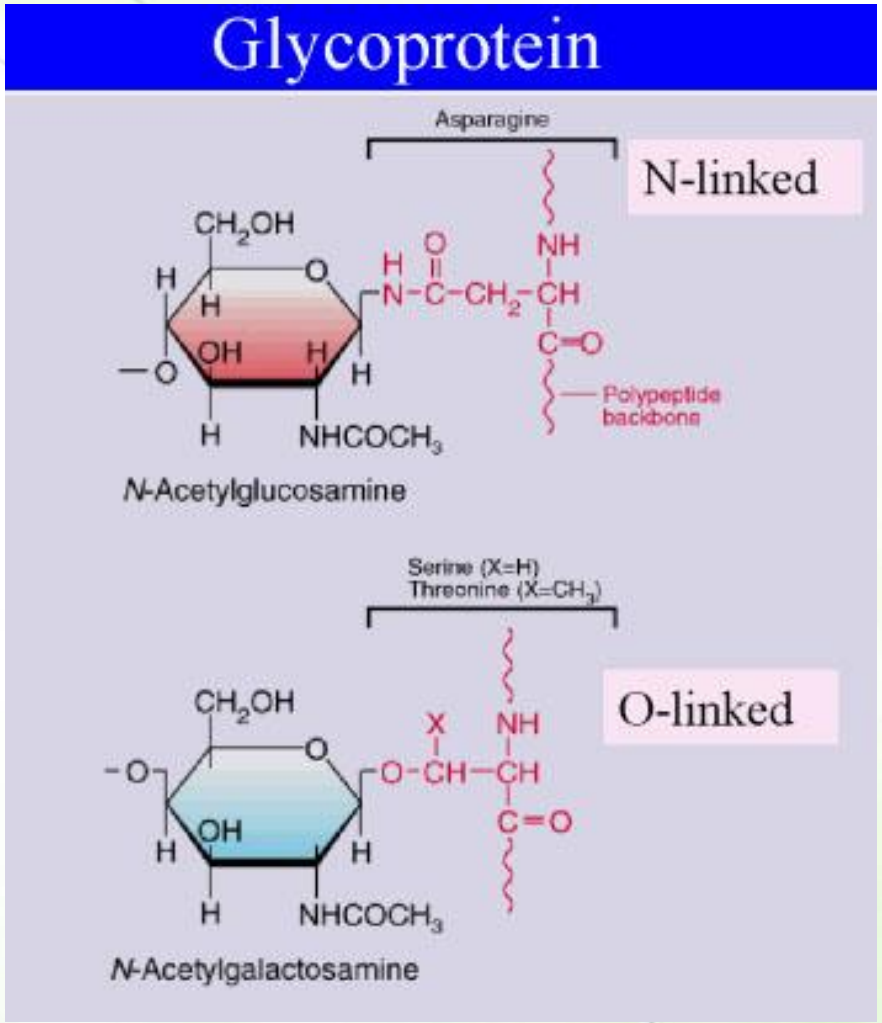






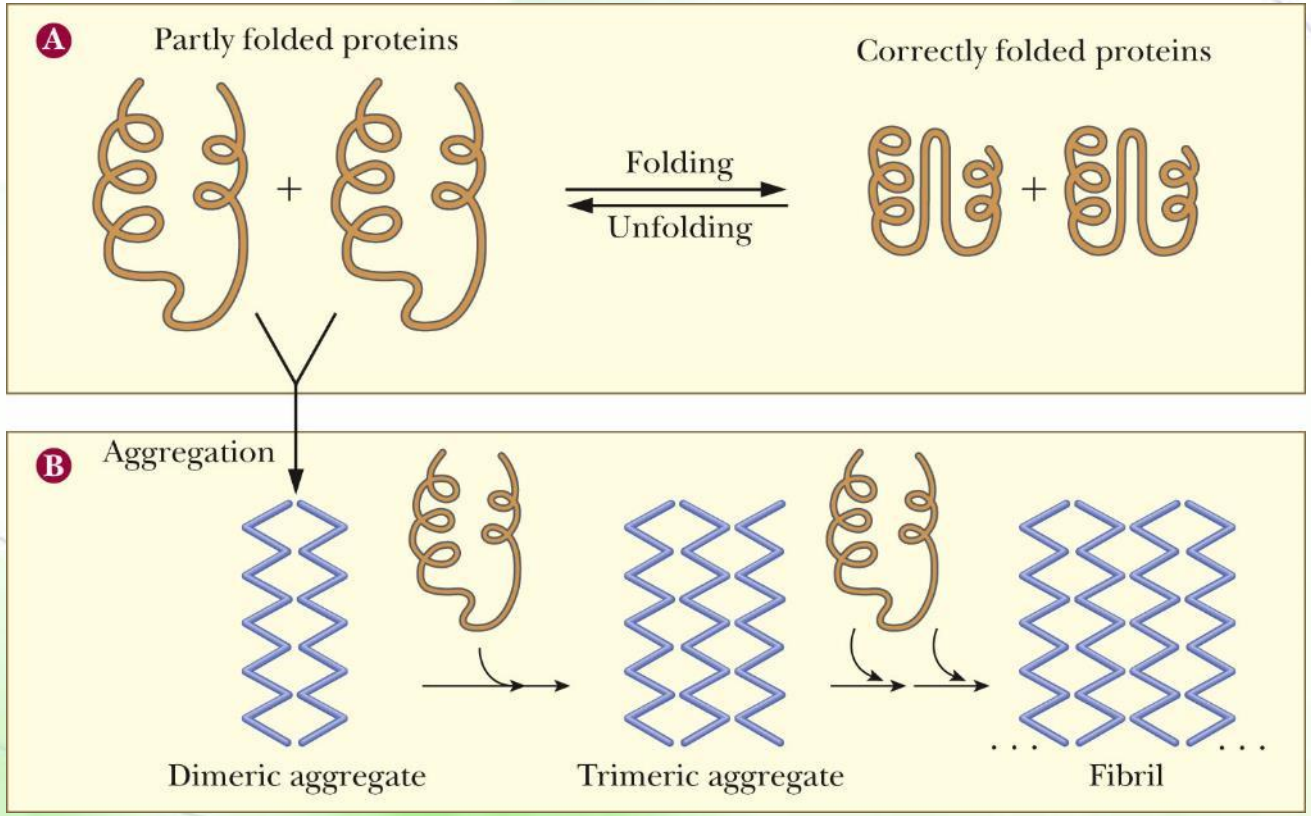
Classes of glycoproteins

- N-linked sugars
  - The amide nitrogen of the R-group of asparagine
- O-linked sugars
  - The hydroxyl groups of either serine or threonine
  - Occasionally to hydroxylysine such as in collagen



The problem of misfolding

- When proteins do not fold correctly, their internal hydrophobic regions become exposed and interact with other hydrophobic regions on other molecules, and form aggregates.

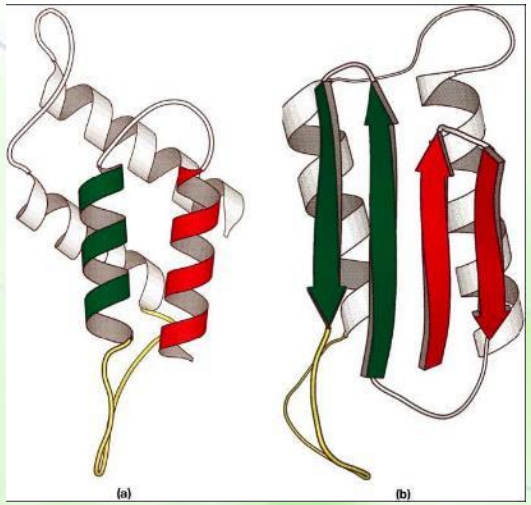


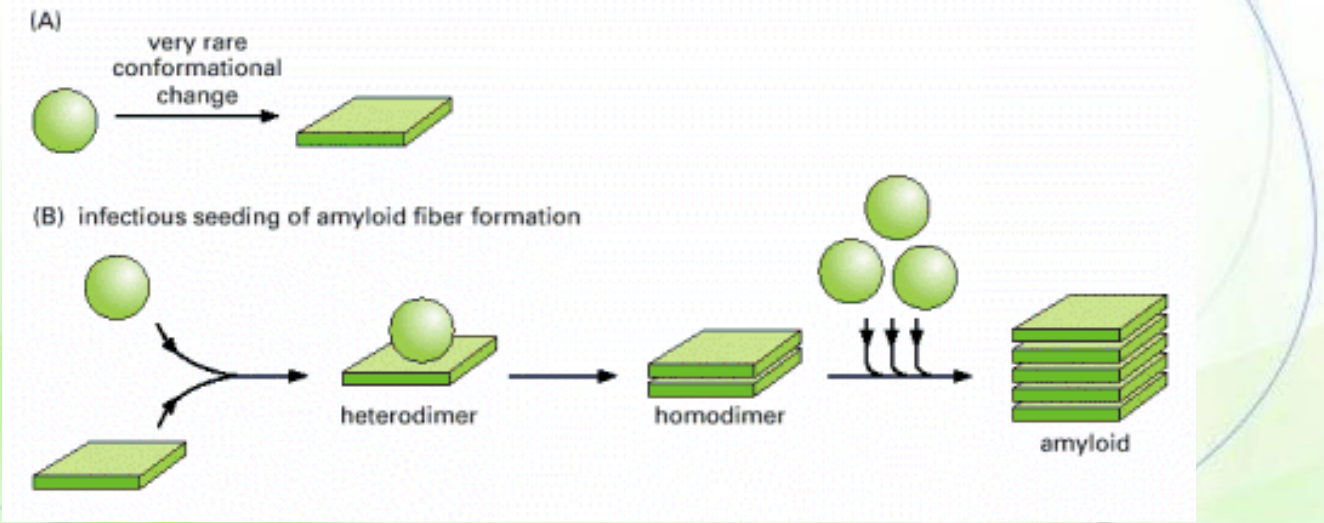
Outcome of protein misfolding

- Partly folded or misfolded polypeptides or fragments may sometimes associate with similar chains to form aggregates.
- Aggregates vary in size from soluble dimers and trimers up to insoluble fibrillar structures (amyloid).
- Both soluble and insoluble aggregates can be toxic to cells.

Prion disease

- Striking examples of protein folding-related diseases are prion diseases, such as Creutzfeldt-Jacob disease (in humans), and mad cow disease (in cows), and scrapie (in sheep).
- Pathological conditions can result if a brain protein known to as prion protein (PrP) is misfolded into an incorrect form called PrPsc.
- PrPC has a lot of α-helical conformation, but PrPsc has more β strands forming aggregates.





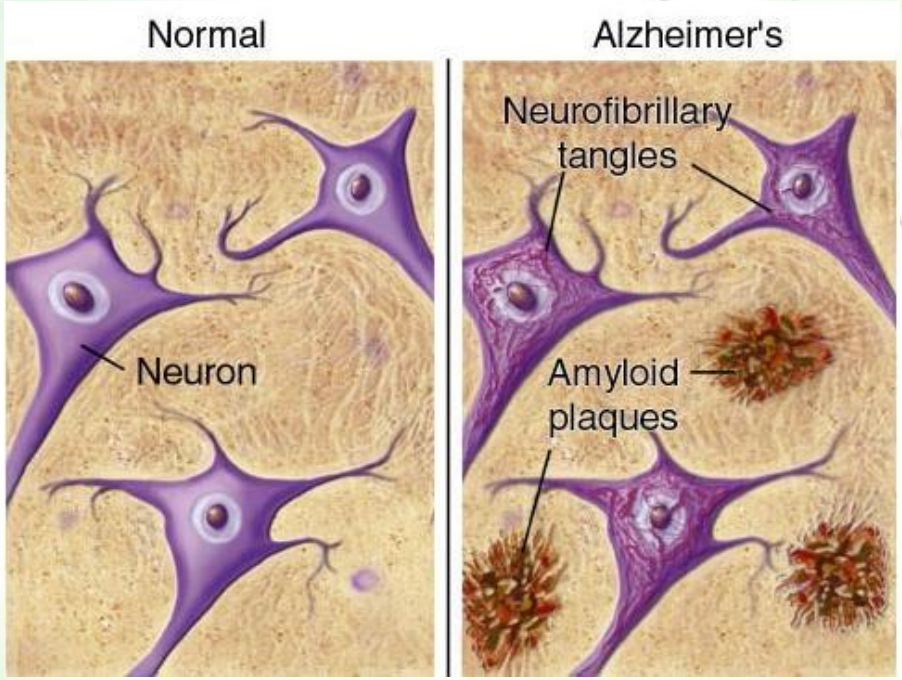
The prion protein

- The disease is caused by a transmissible agent
- Abnormal protein can be acquired by
  - Infection
  - Inheritance
  - Spontaneously

Alzheimer's Disease

- Not transmissible between individuals

  - Extracellular plaques of protein aggregates of a protein called tau and another known as amyloid peptides (Aβ) damage neurons.



Formation of plaques