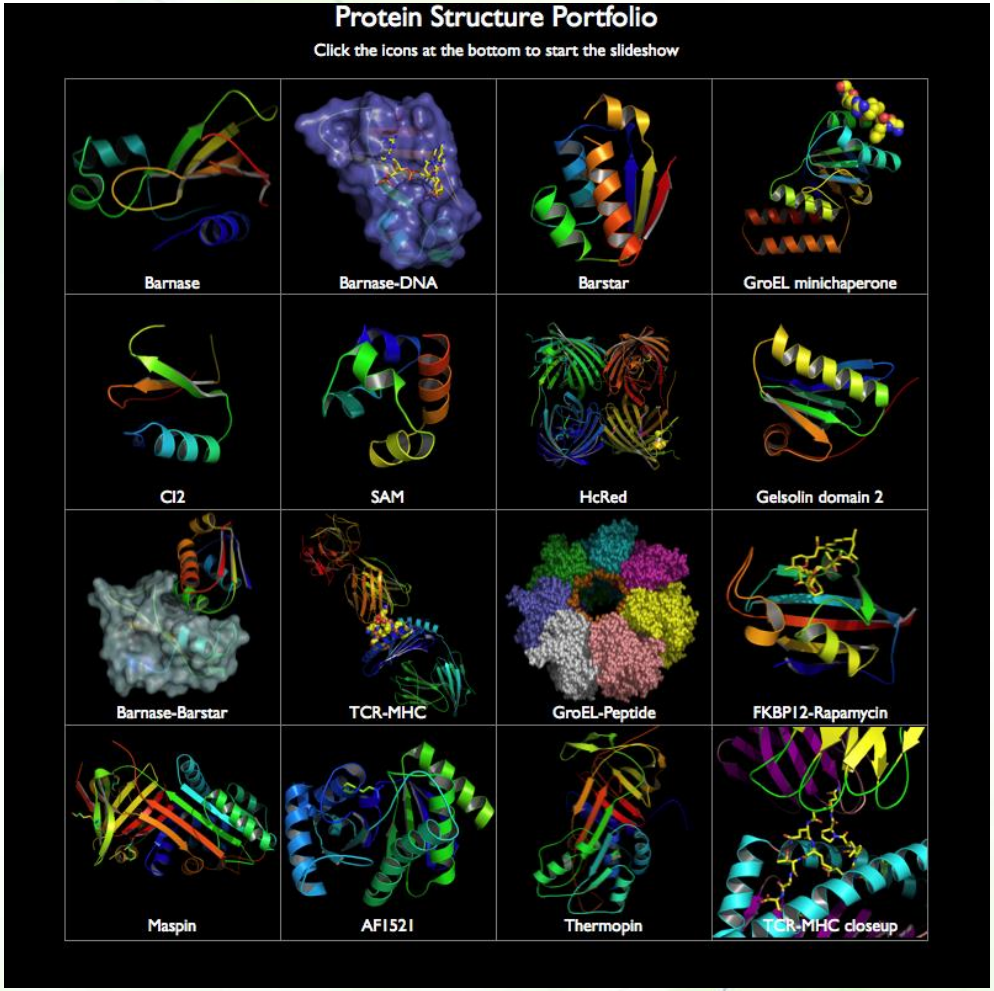
Protein Structure

Dr. Diala Abu Hassan

Overview of proteins

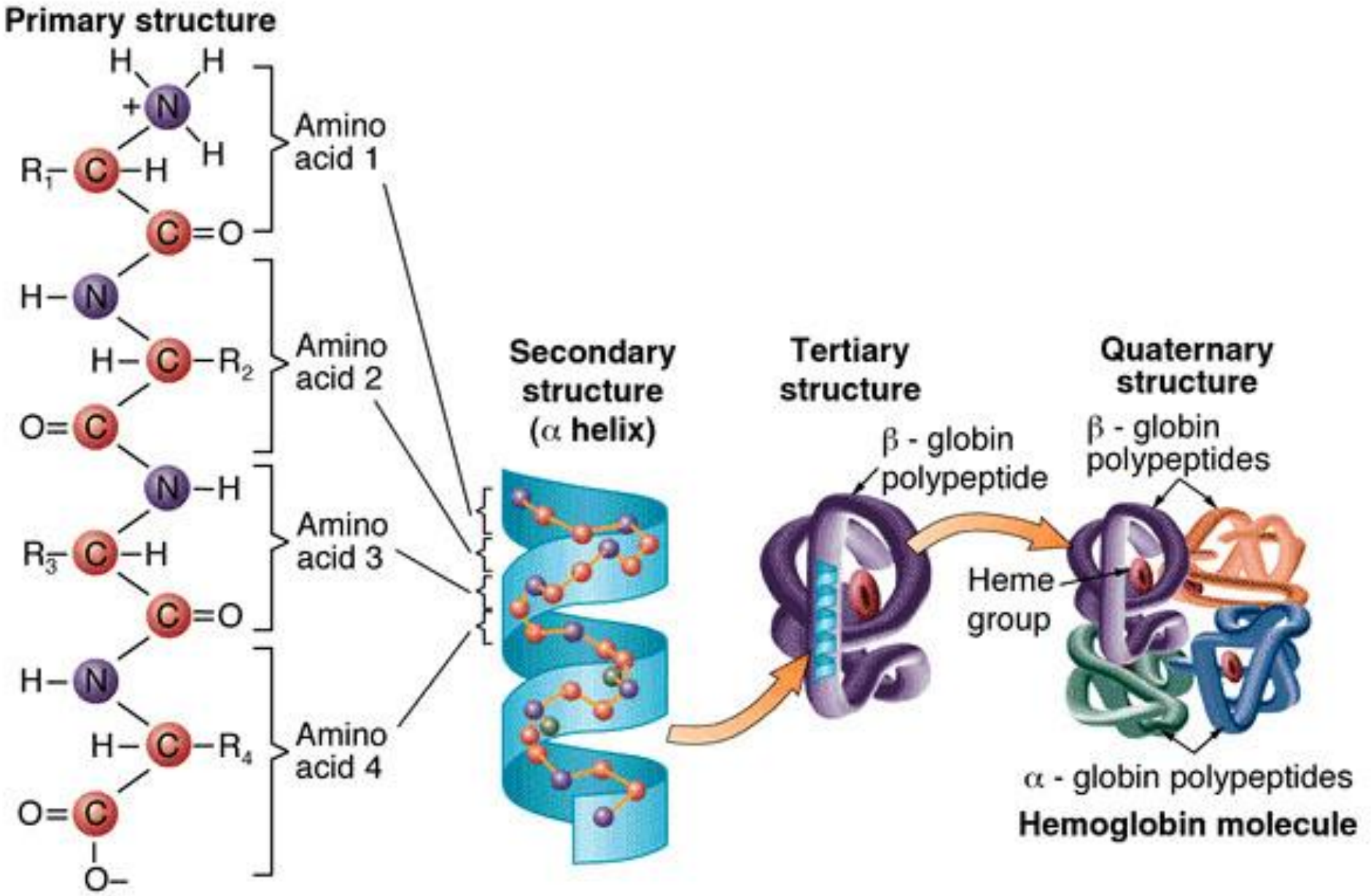
- Proteins have different structures and some have repeating inner structures, other do not.
- A protein may have *gazillion* possibilities of structures, but a few would be active.
- These active structures are known as **native conformations** (the 3-dimensioanl structure of a properly folded and functional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



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

Protein structure



Primary structure

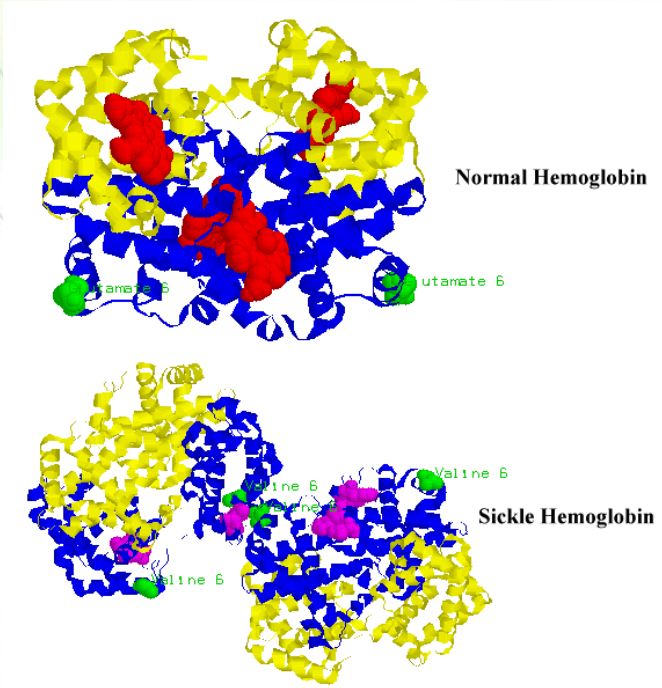
What is primary structure?

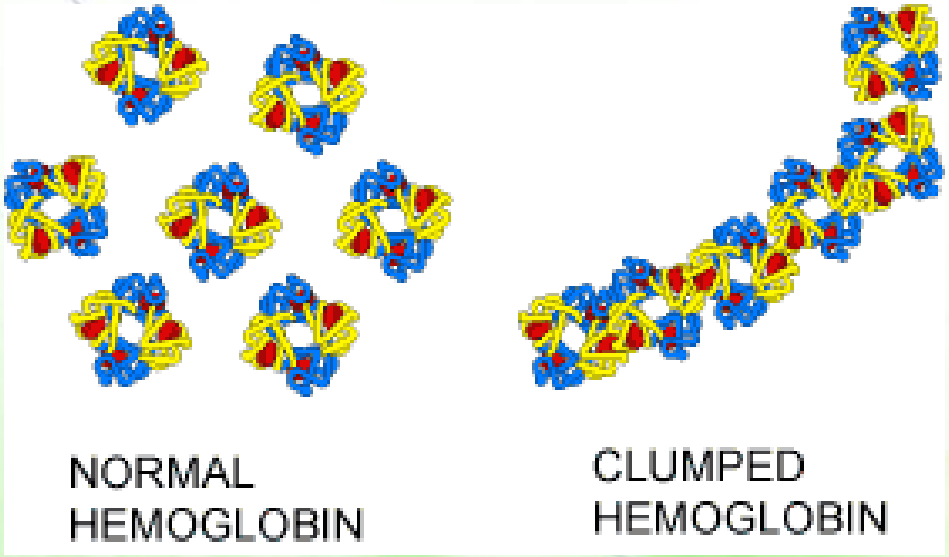
- The order in which the amino acids are covalently linked together.

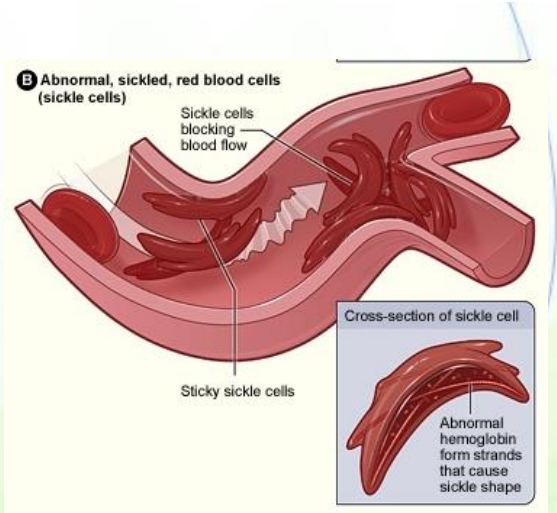
  Example: Leu—Gly—Thr—Val—Arg—Asp—His
- The primary structure of a protein determines the other levels of structure.
- A single amino acid substitution can give rise to a malfunctioning protein, as is the case with sickle-cell anemia.

Sickle cell hemoglobin (HbS)

- 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.



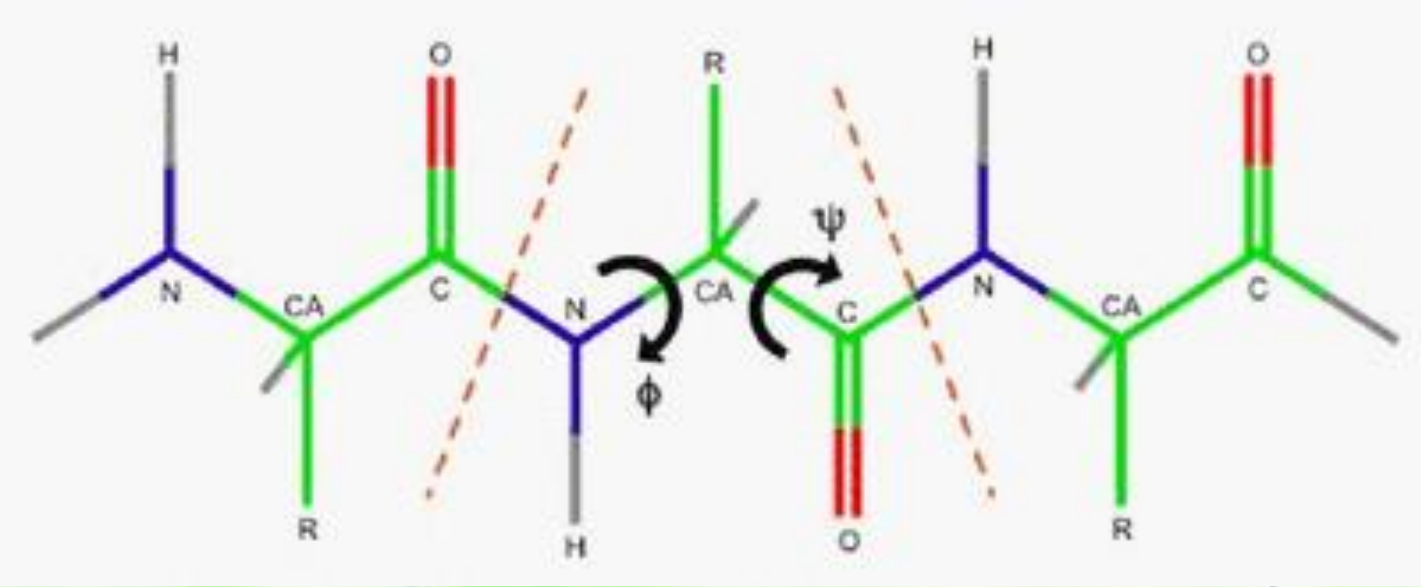




Secondary Structure

What is it? How is caused?

- The two bonds within each amino acid residue freely rotate
  - the bond between the α-carbon and the amino nitrogen
  - the bond between the α-carbon and the carboxyl carbon

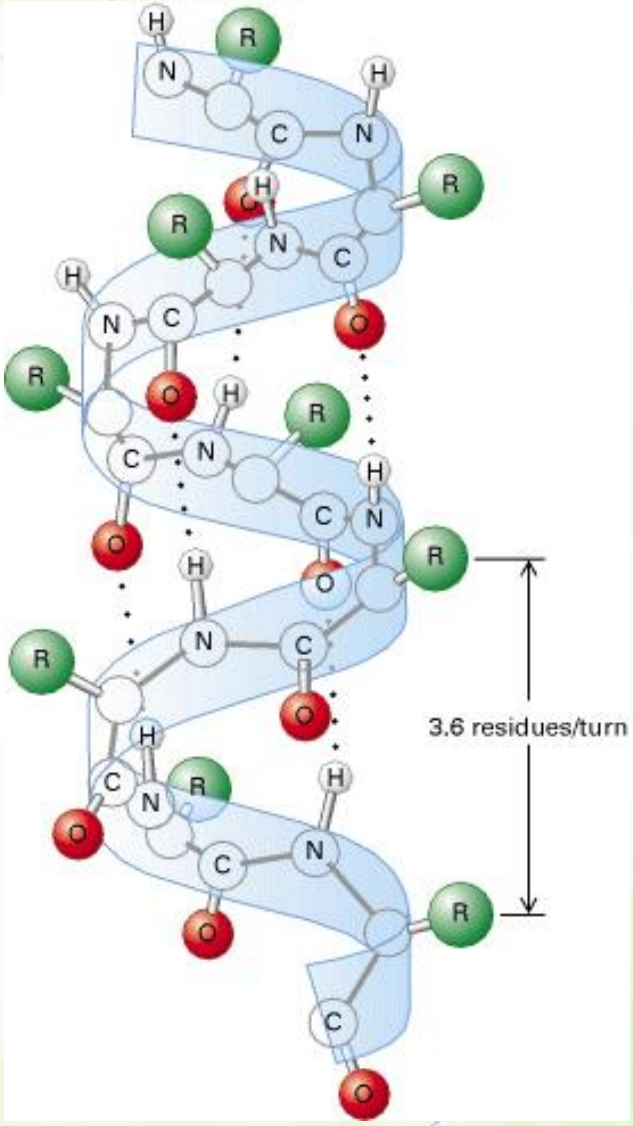


Common secondary structures

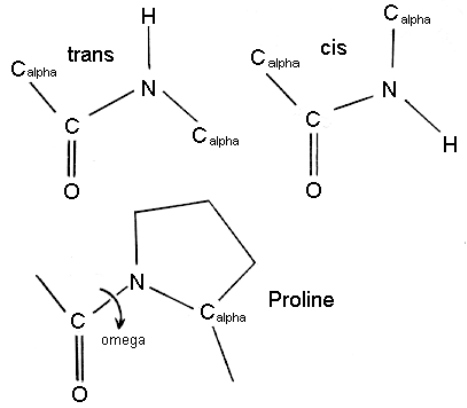
- A hydrogen-bonded, local arrangement of the backbone of a polypeptide chain.
- Polypeptide chains can fold into regular structures such as:
  - Alpha helix
  - Beta-pleated sheet
  - Turns
  - Loops

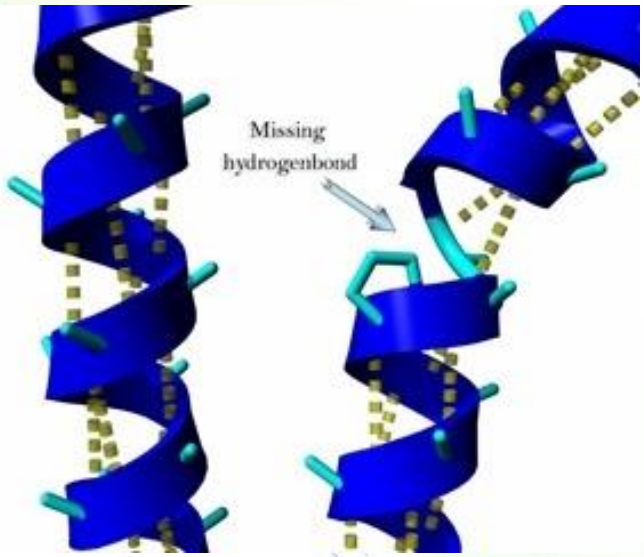
The α helix

- It looks like a helical rod.
- The helix has an average of 3.6 amino acids per turn.
- The pitch of the helix *(the linear distance between corresponding points on successive turns*) is 5.4 Å
  - 1 Å = 10^{-10} m
- It is very stable because of the linear hydrogen bondings.



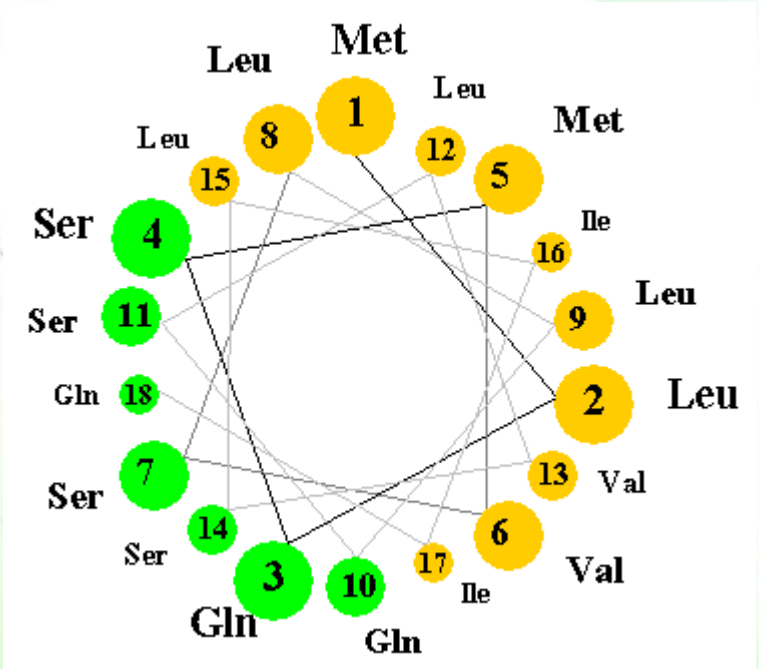
Amino acids NOT found in α-helix

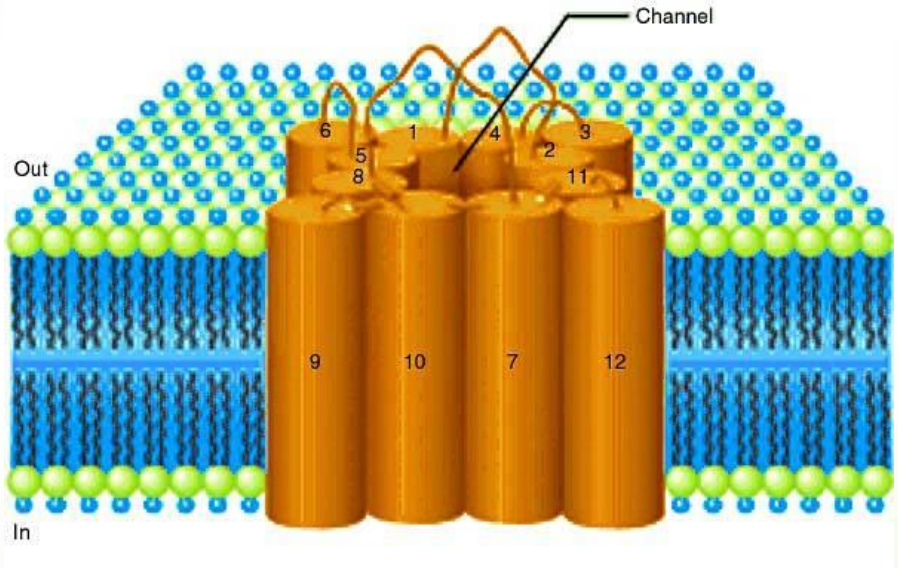




- Glycine: too small
- Proline
  - No rotation around N-Cα bond
  - No hydrogen bonding of α-amino group
- Close proximity of a pair of charged amino acids with similar charges
- Amino acids with branches at the β-carbon atom (valine, threonine, and isoleucine)

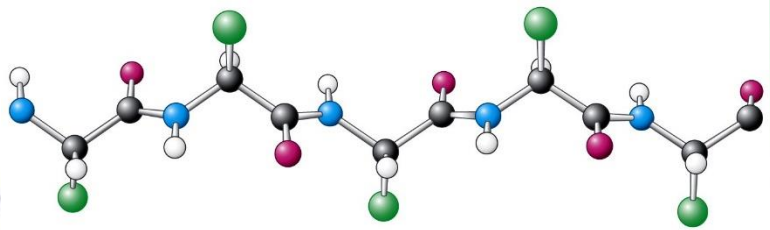
Amphipathic α helices



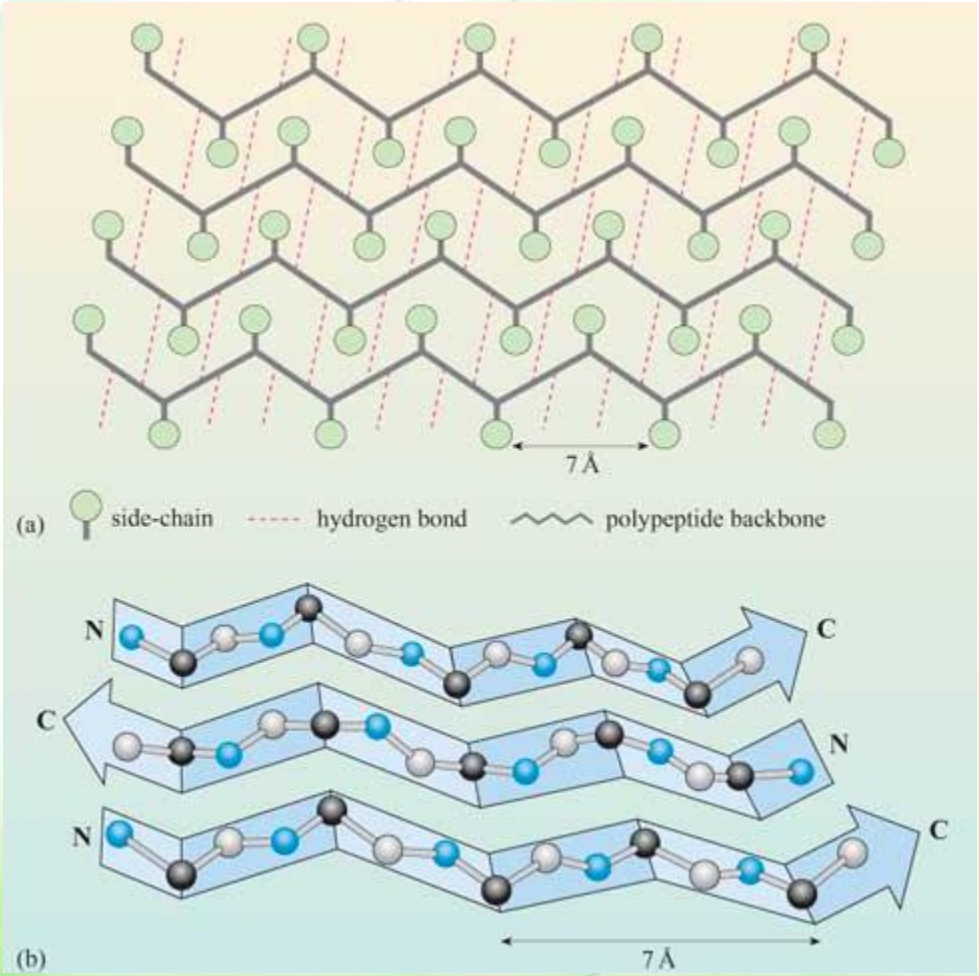


β pleated sheet (β sheet)

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

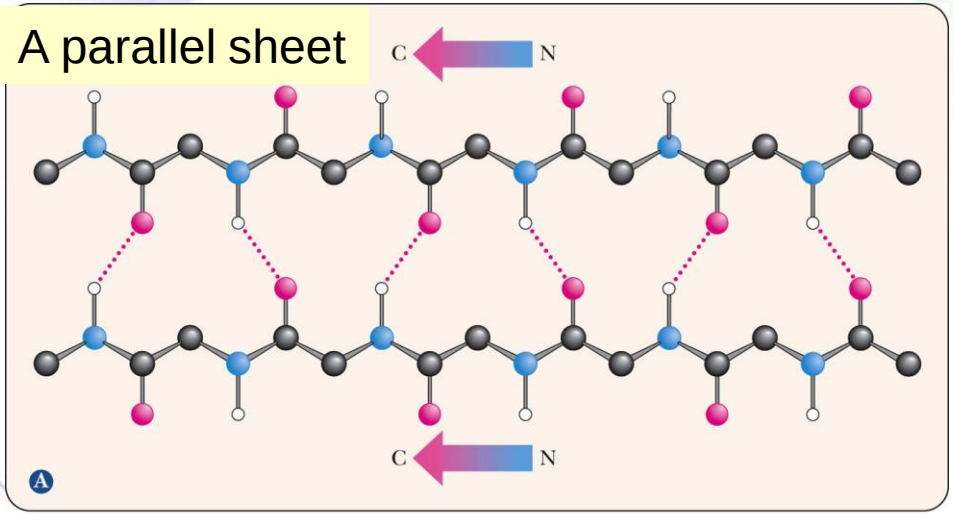


β strand

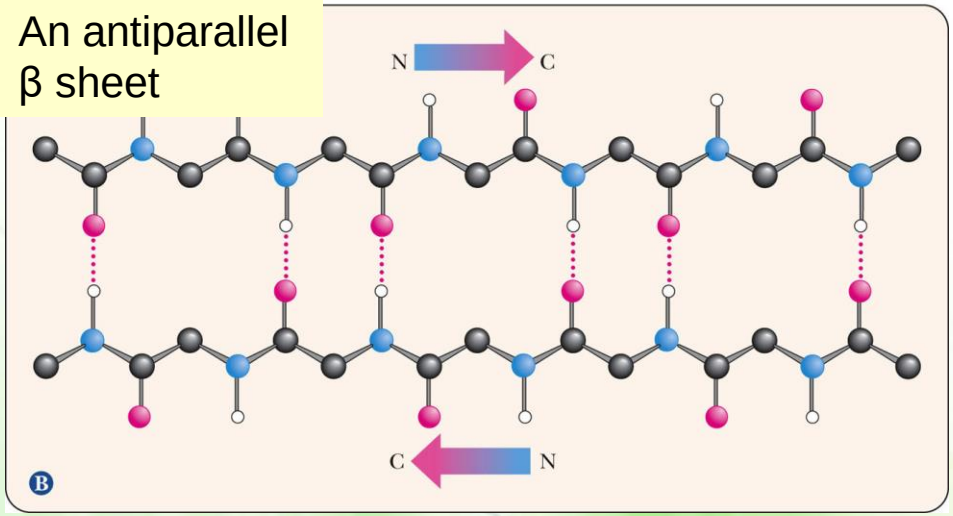


Parallel vs. antiparallel β sheets

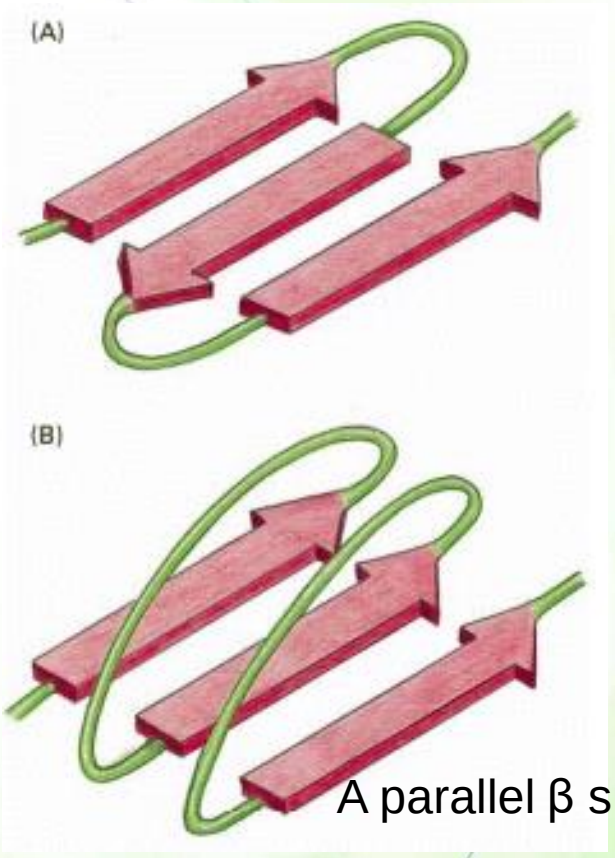
A parallel sheet



An antiparallel β sheet



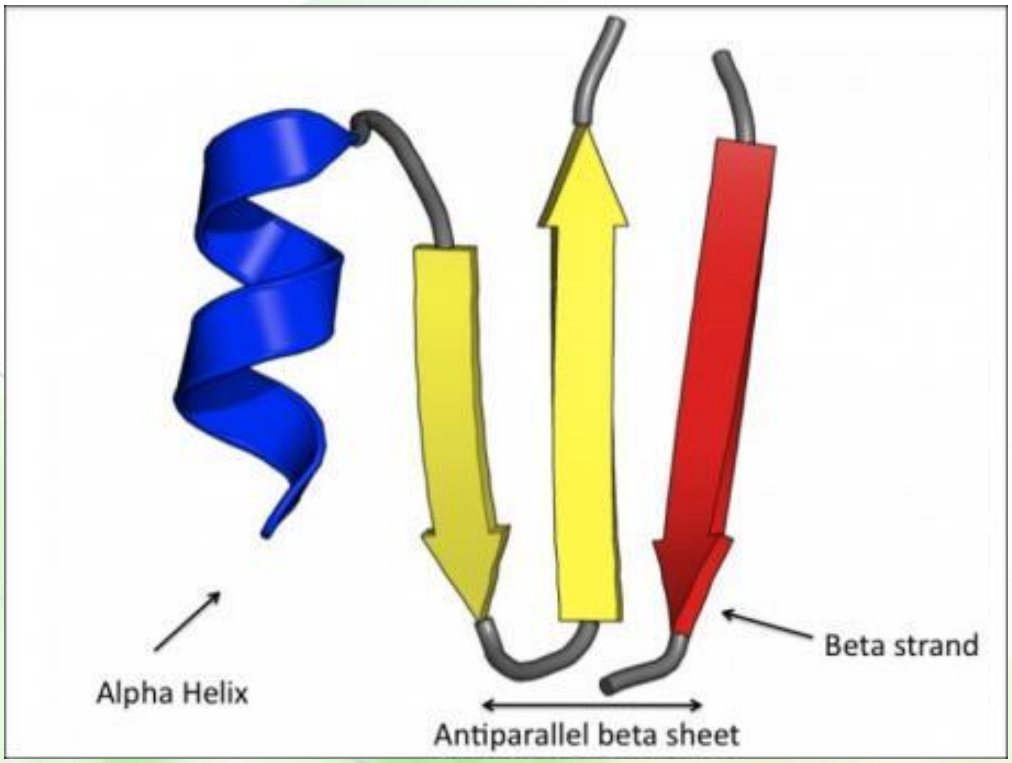
An antiparallel β sheet



A parallel β sheet

How many β strands can a β sheet have?

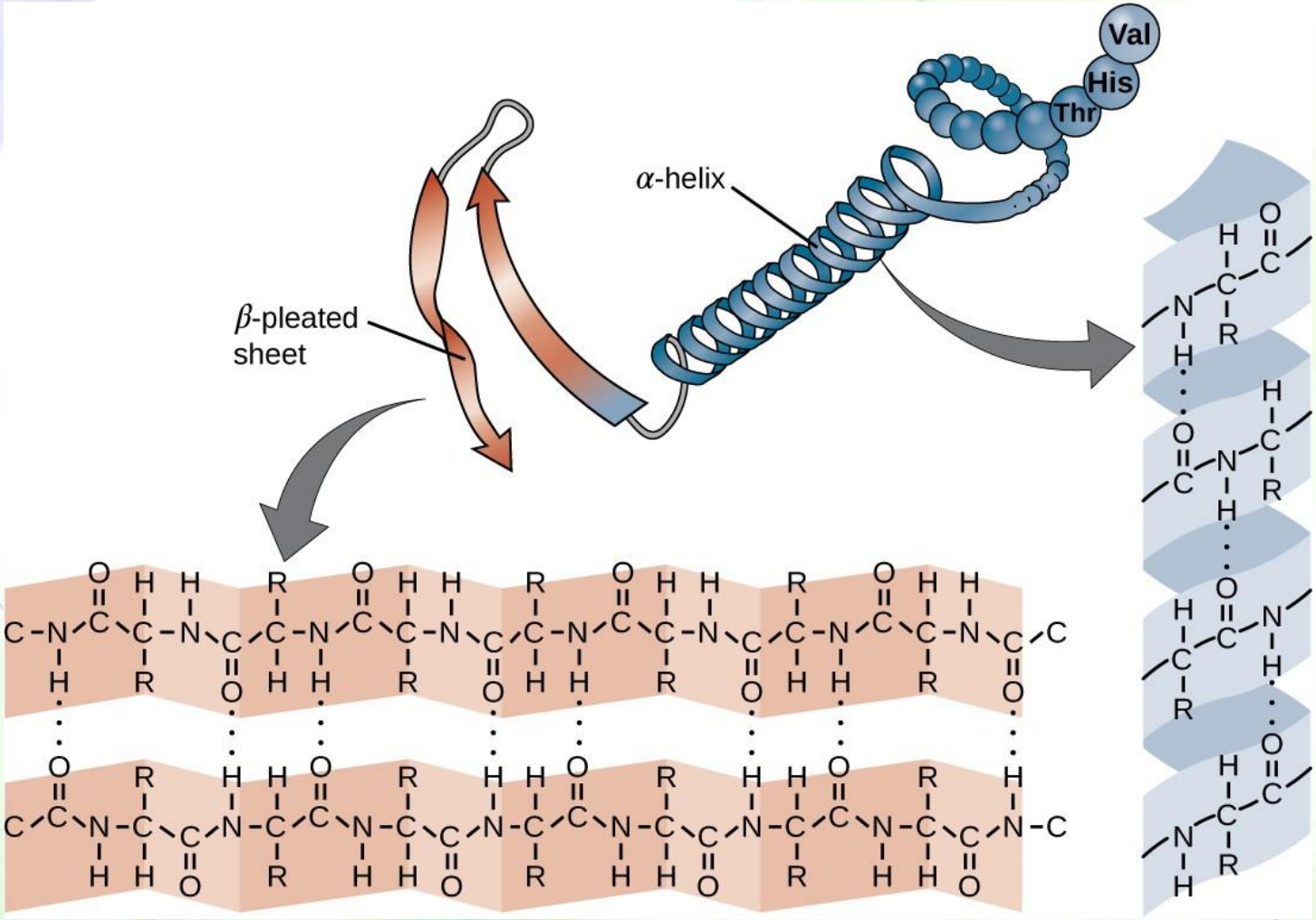
- β sheets can form between many strands, typically 4 or 5 but as many as 10 or more.
- Such β sheets can be purely antiparallel, purely parallel, or mixed.



Effect of amino acids in β-sheets

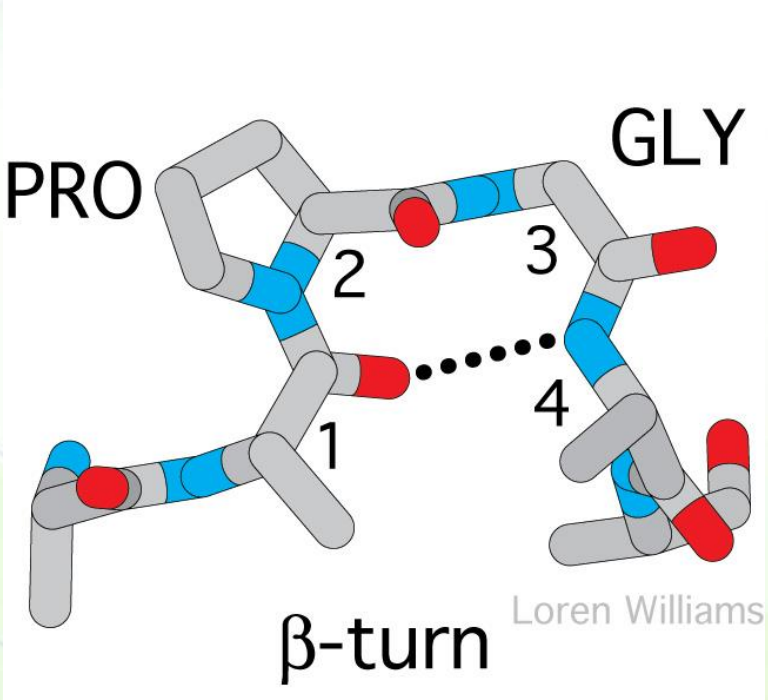
- Valine, threonine and Isoleucine with branched R groups at β-carbon and the large aromatic amino acids (phenylalanine, tryptophan, and tyrosine) tend to be present in β-sheets.
- Proline tends to disrupt β strands

Note how they are illustrated



β-turns

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

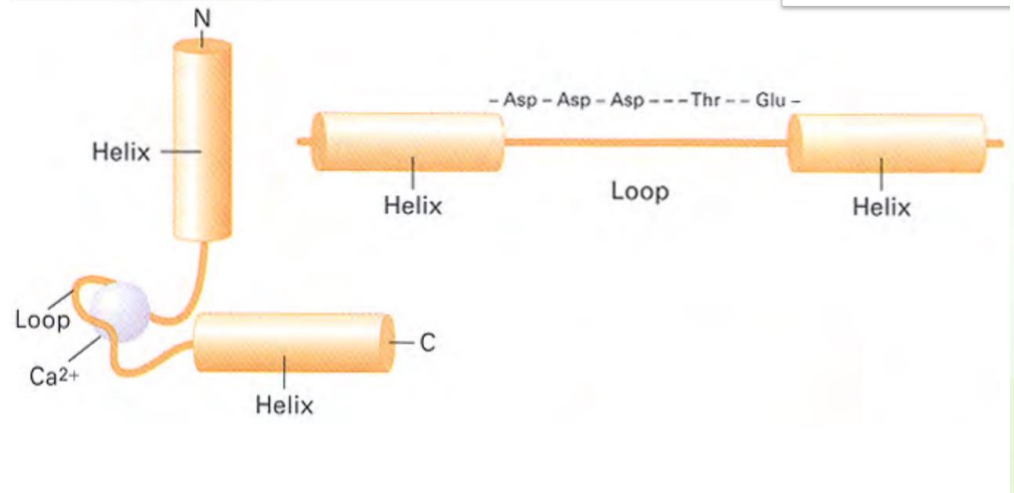


Super-secondary structures

- They are regions in proteins that contain an ordered organization of secondary structures.
- There are at least types:
  - Motifs
  - Domains

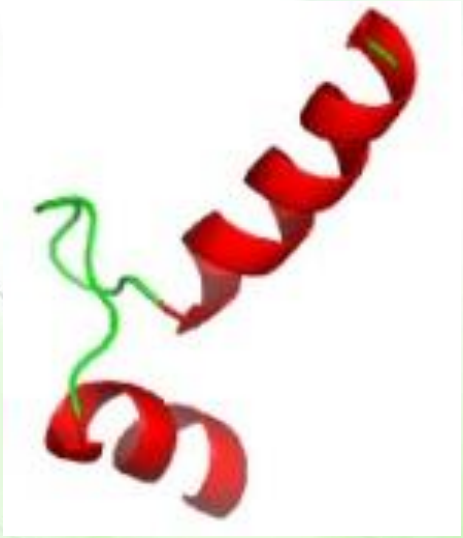
A motif (a module)

- A motif is a repetitive supersecondary structure, which can often be repeated and organized into larger motifs.
- It usually constitutes 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.

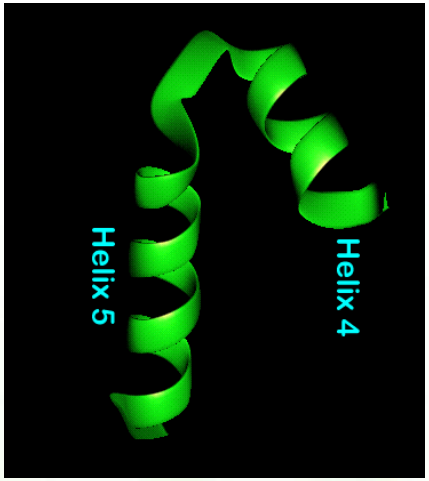


Examples of motifs

Helix-loop-helix is found in many proteins that bind DNA. It is characterized by two α-helices connected by a loop.

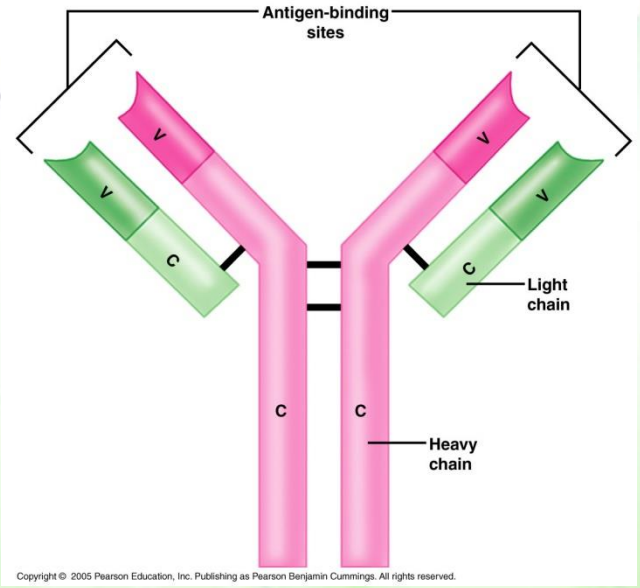


Helix-turn-helix is a structural motif capable of binding DNA. It is composed of two α helices joined by a short strand of amino acids

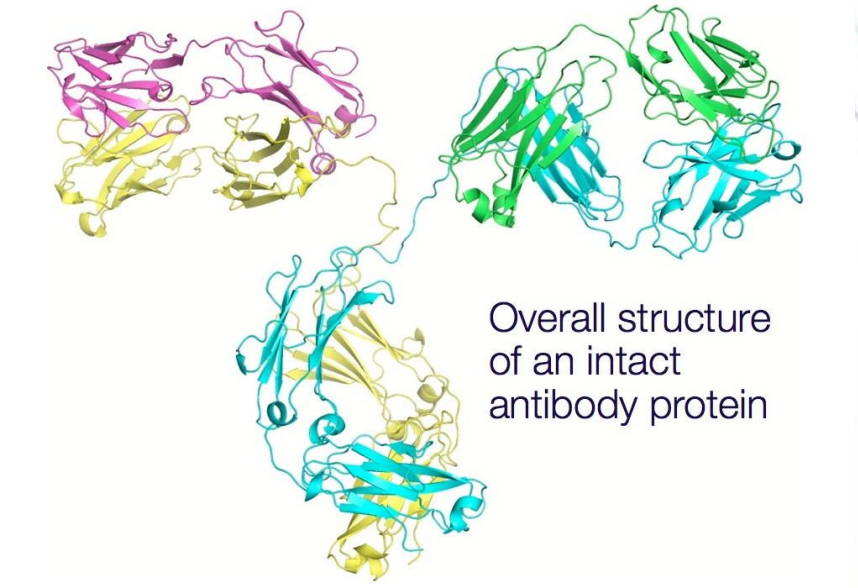


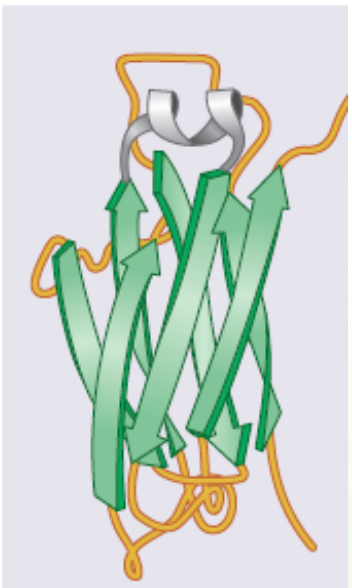
A more complex motif is...

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



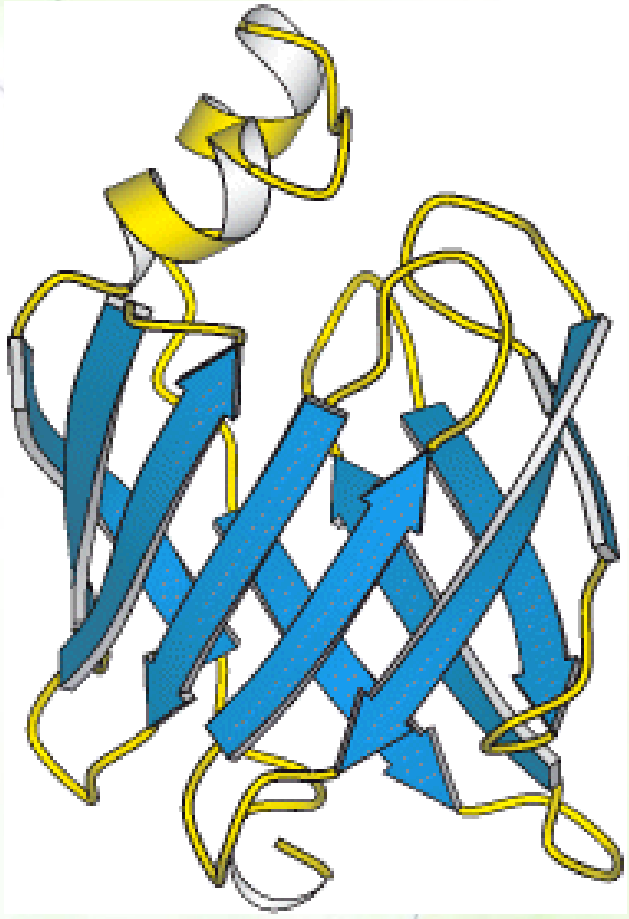
Copyright © 2005 Pearson Education, Inc. Publishing as Pearson Benjamin Cummings. All rights reserved.





A domain

- A domain is a compactly folded region of polypeptide found in proteins with similar function and/or structure.
- Domains with similar conformations are associated with the particular function.
- A structural domain may consist of 100–200 residues in various combinations of α helices, β sheets, turns, and random coils.
- They fold independently of the rest of the protein.
- Domains may also be defined in functional terms
  - enzymatic activity
  - binding ability (e.g., a DNA-binding domain)



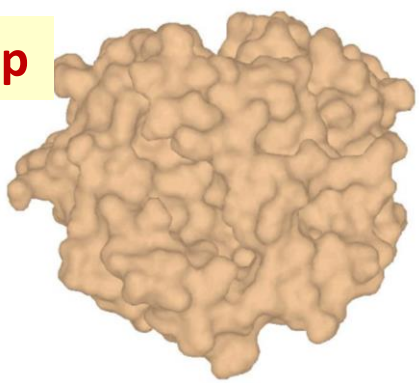
Tertiary structure

What is tertiary structure?

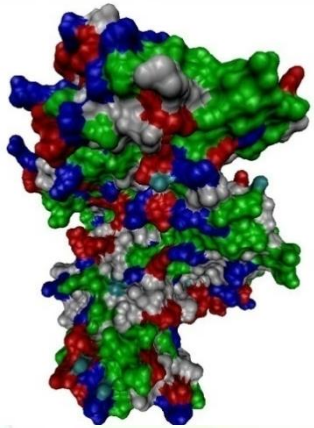
- The overall conformation of a polypeptide chain
- The three-dimensional arrangement of all the amino acids residues
- The spatial arrangement of amino acid residues that are far apart in the sequence

How to look at proteins…

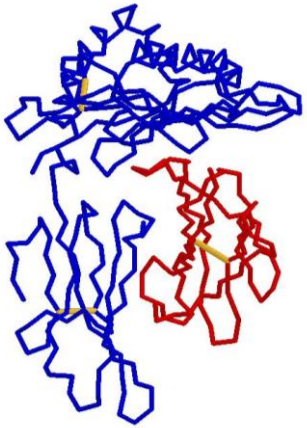
Protein surface map



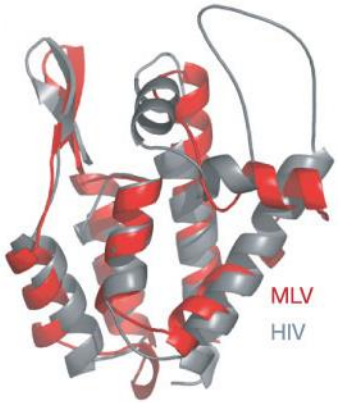
Space filling structure



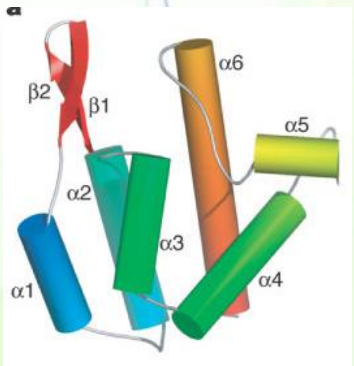
Trace structure



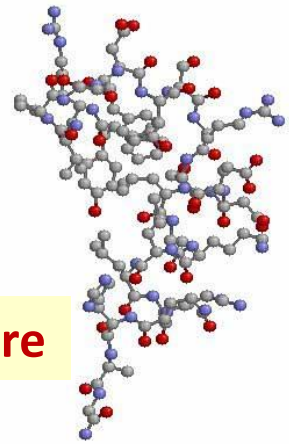
Ribbon structure



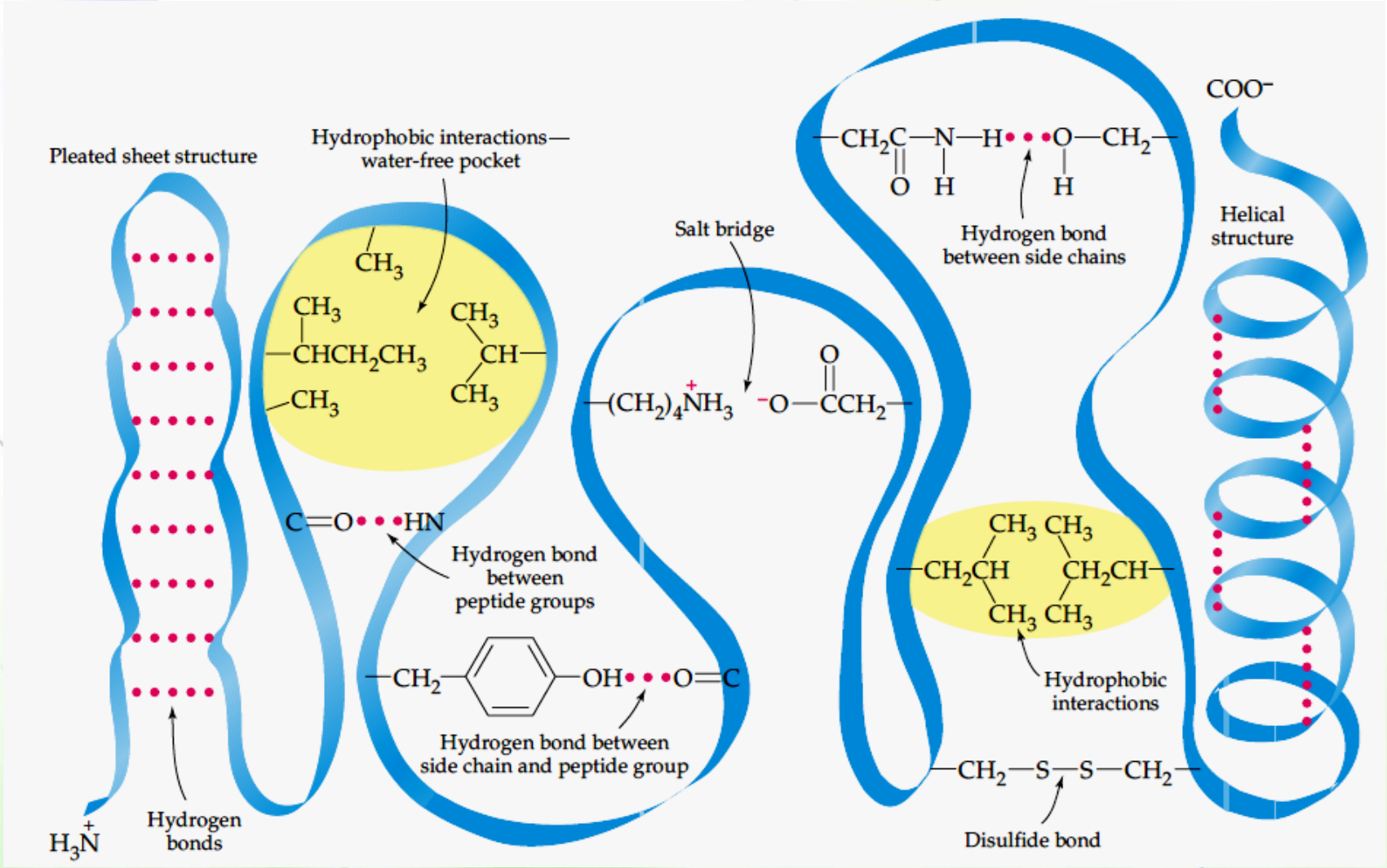
Cylinder structure



Ball and stick structure



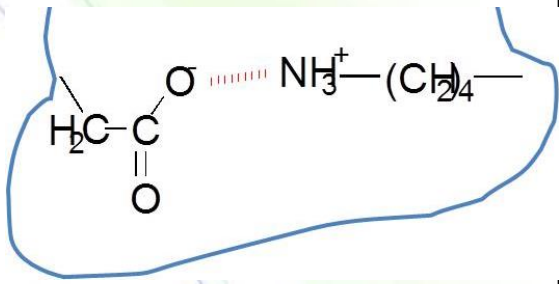
Shape determining forces



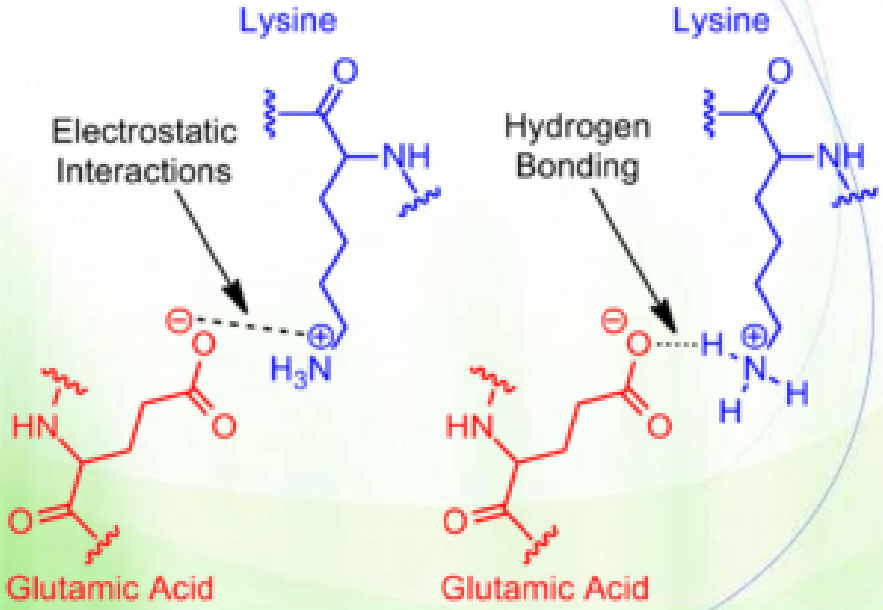
Non-covalent interactions

- 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.

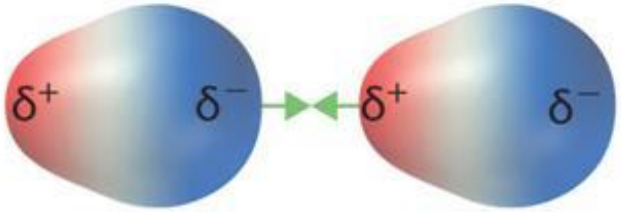


The same charged group can form either hydrogen bonding or electrostatic interactions

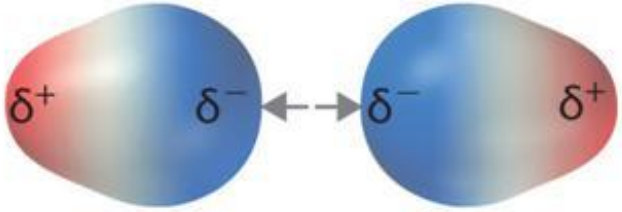


van der Waals attractions

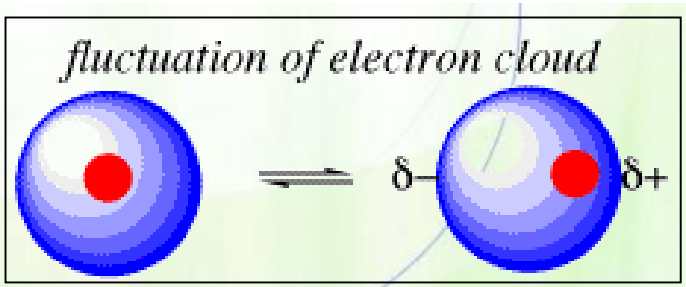
- There are both attractive and repulsive van der Waals forces that control protein folding.
- Although van der Waals forces are extremely weak, they are significant because there are so many of them in large protein molecules.



(b) Attraction

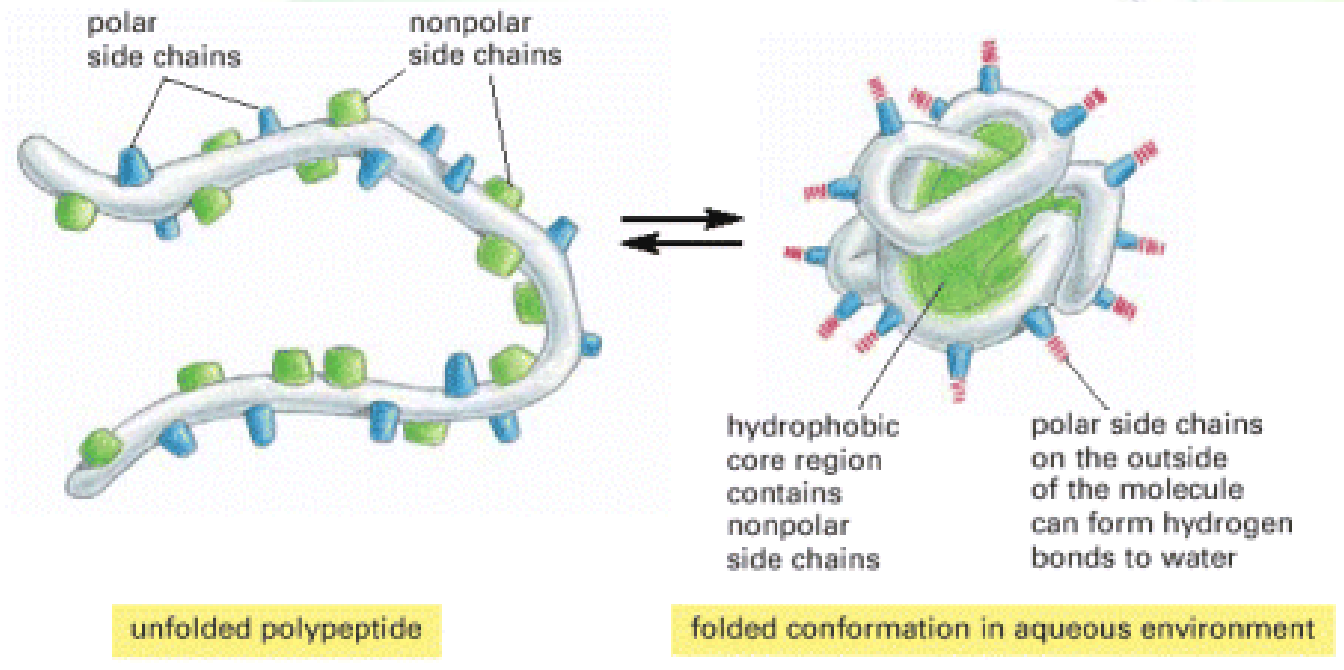


(d) Repulsion



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

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.

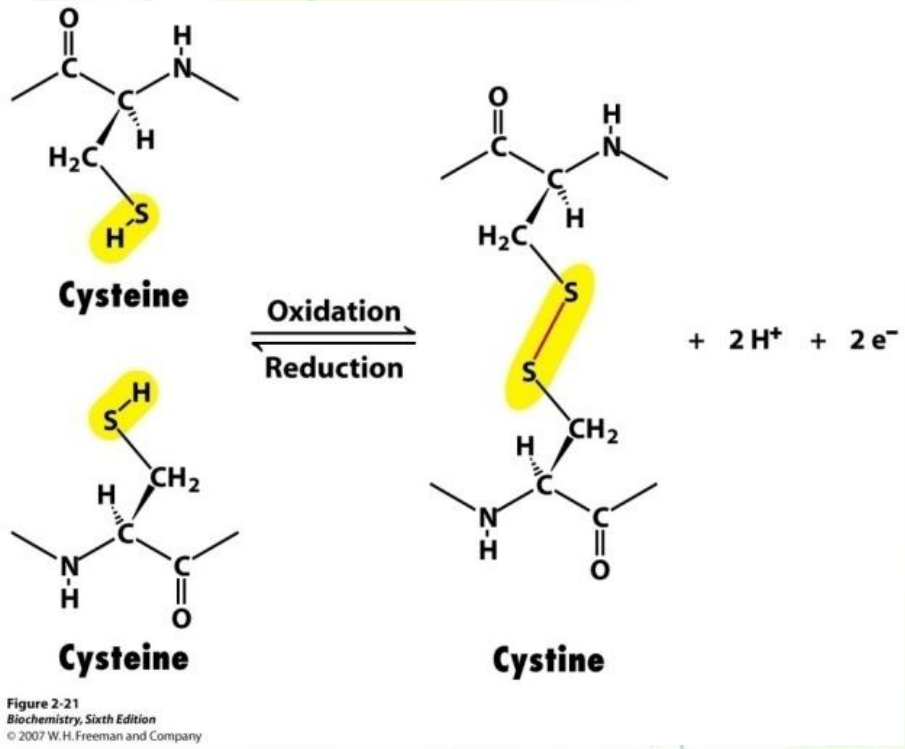
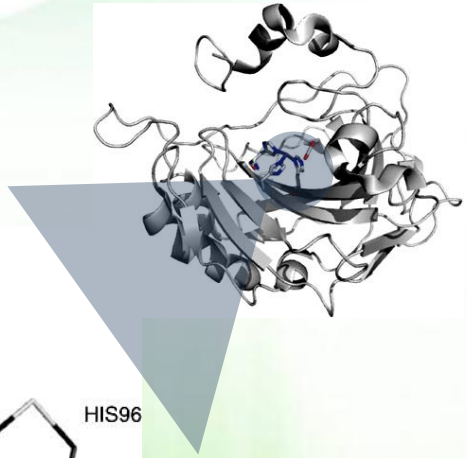


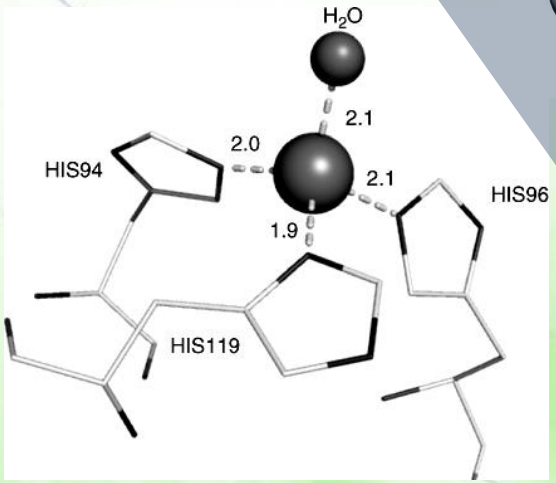
Figure 2-21
Biochemistry, Sixth Edition
© 2007 W. H. Freeman and Company

metal ions

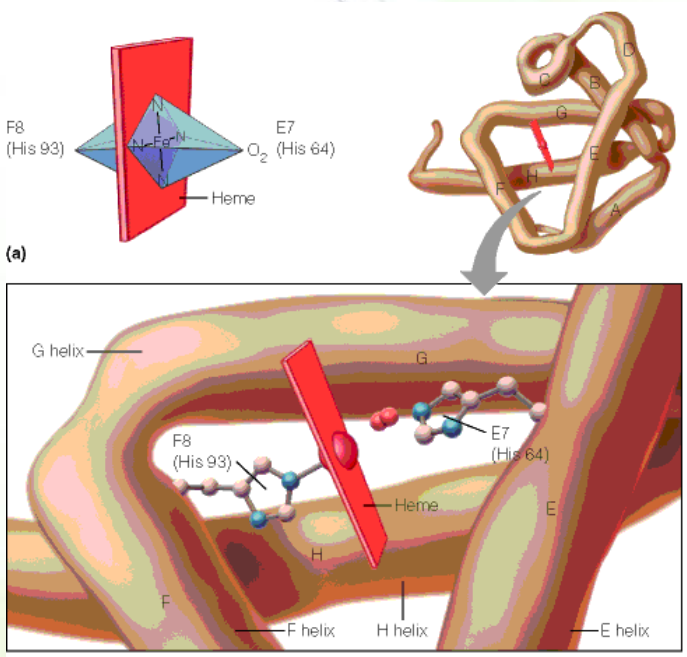
- Several proteins can be *complexed to a* single metal ion that can stabilize protein structure by forming:
  - Covalent interaction (myoglobin)
  - Salt bridges (carbonic anhydrase)

Carbonic anhydrase





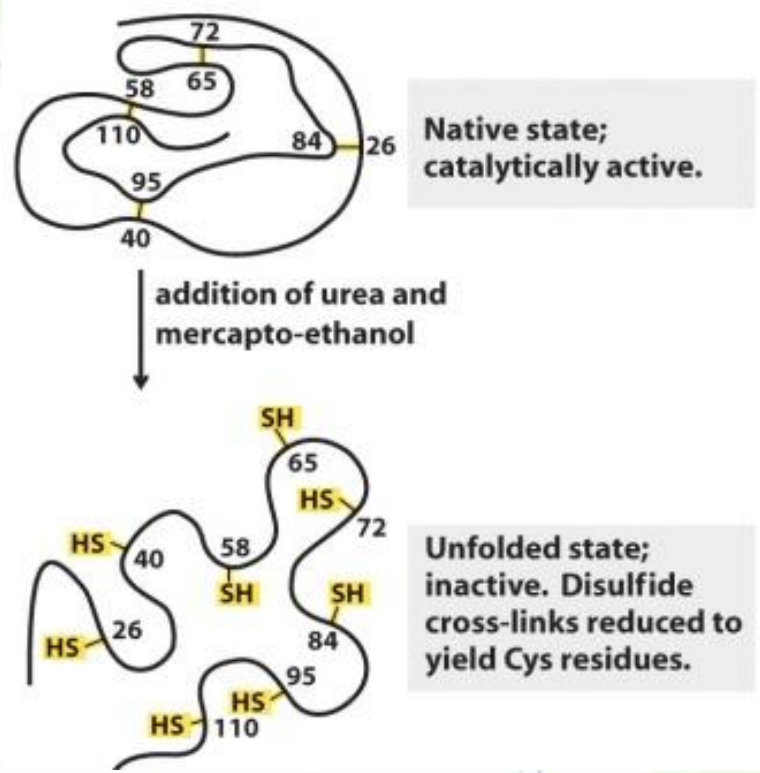
Myoglobin



Properties of Proteins:
Denaturation and Renaturation

Denaturation

- Denaturation is the disruption of the *native conformation* of a protein, the characteristic three-dimensional structure that it attains after synthesis
- Denaturation involves the breaking of the noncovalent bonds which determine the structure of a protein
- Complete disruption of tertiary structure is achieved by reduction of the disulfide bonds in a protein
- Generally, the denatured protein will lose its properties such as activity and become insoluble.



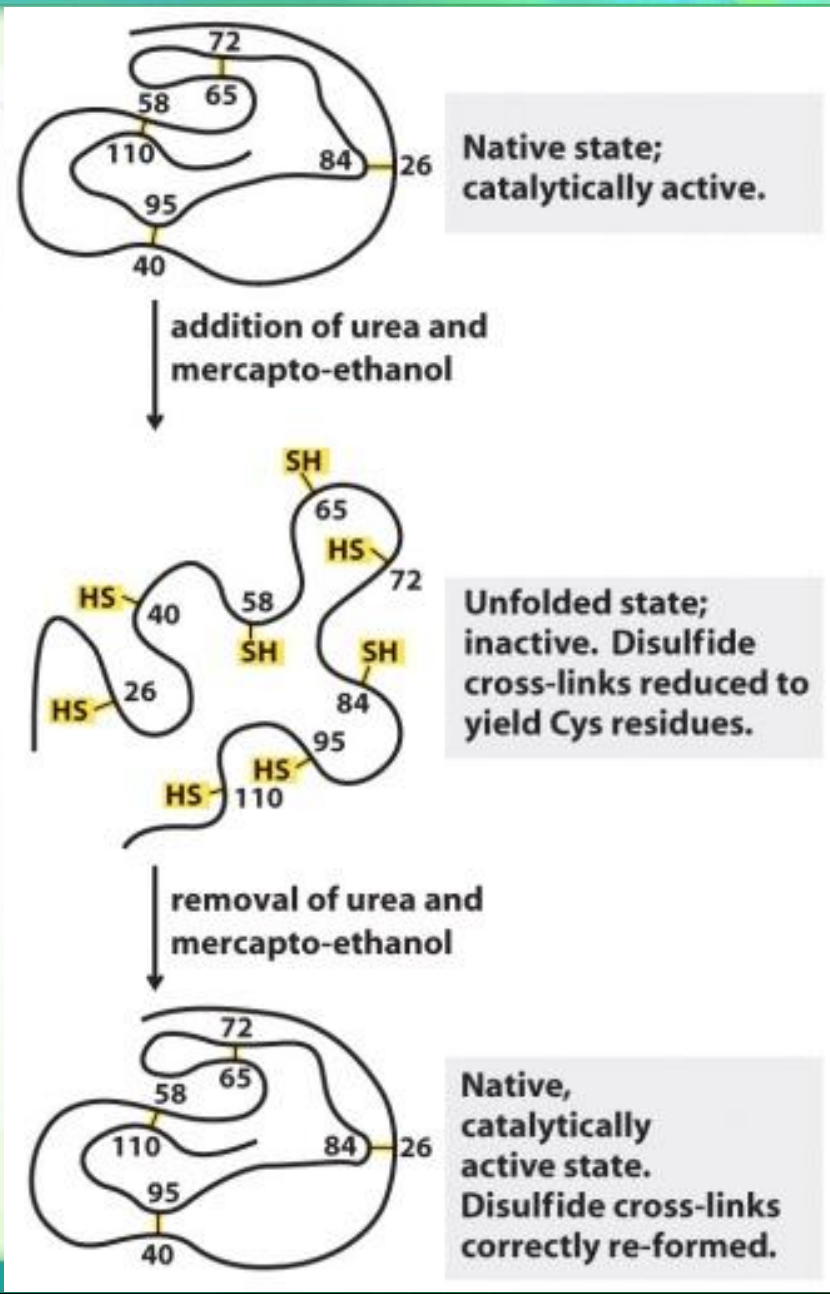
Denaturing agents

- 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 such as β-mercaptoethanol (βME) 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



Factors that determine protein structure

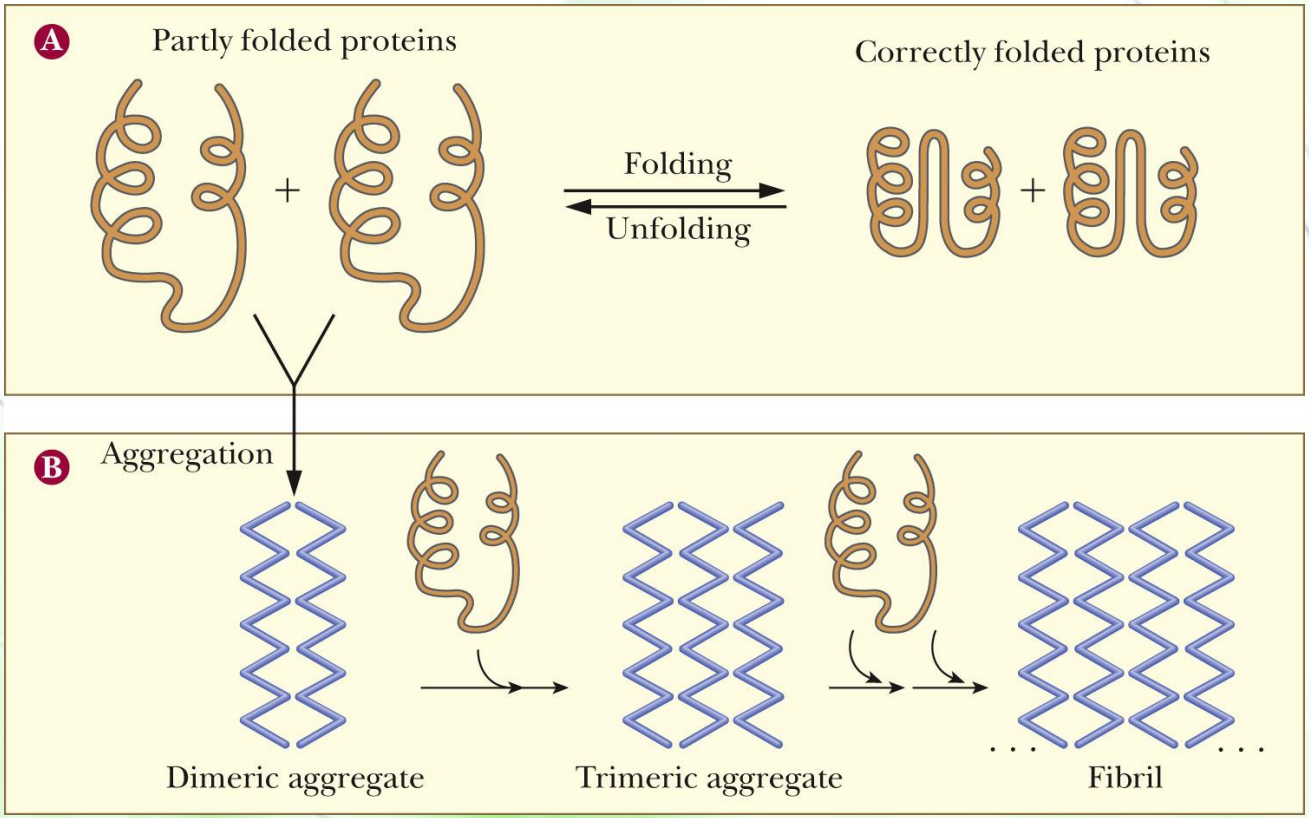
- The least amount of energy needed to stabilize the protein. This is determined by:
  - The amino acid sequence (the primary structure), mainly the internal residues.
  - The proper angles between the amino acids
    - The different sets of weak noncovalent bonds that form between the mainly the R groups.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

Can an unfolded protein re-fold?

- 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.

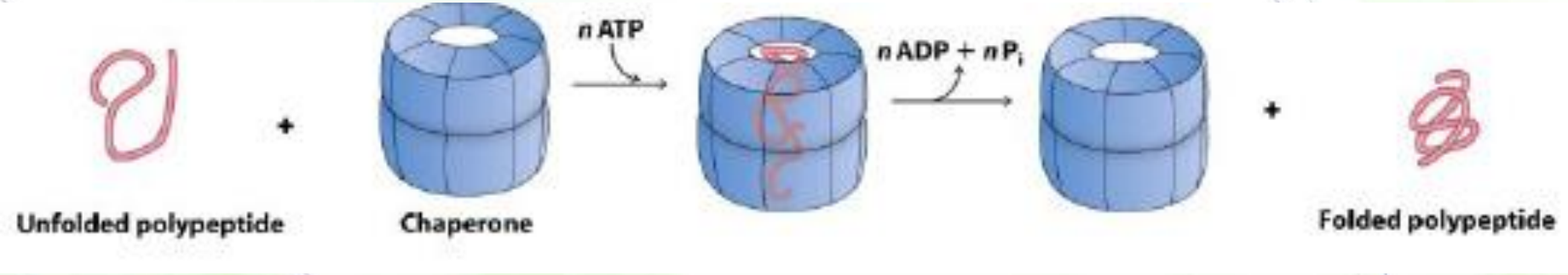
The problem of misfolding

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



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 .



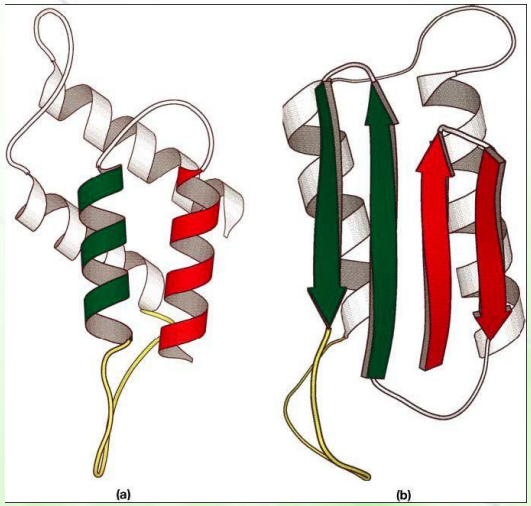
Many diseases are the result of defects in protein folding

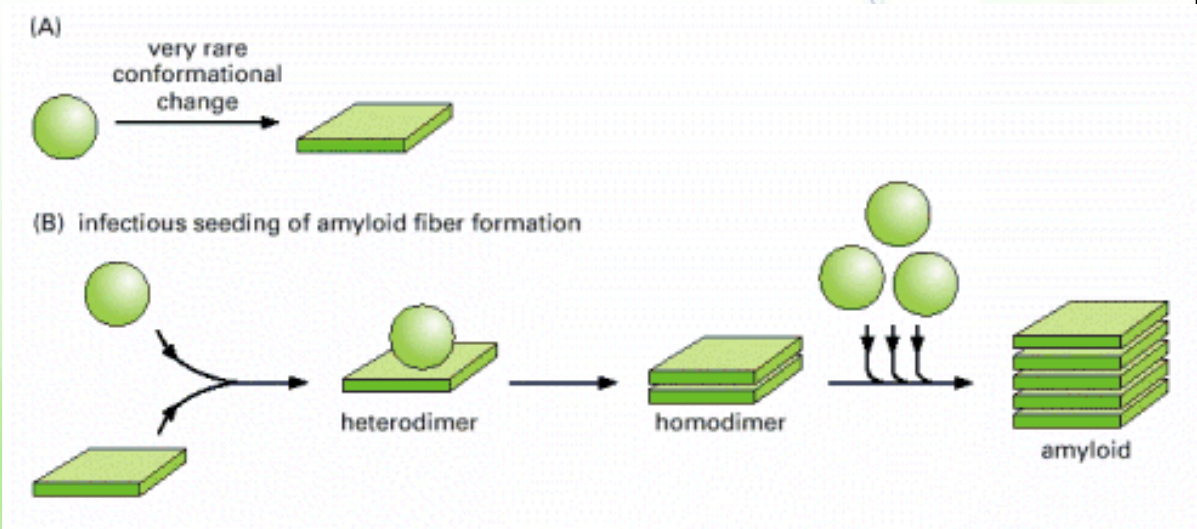
Outcome of protein misfolding

- 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 PrP^{sc}.
- PrP^{C} has a lot of α-helical conformation, but PrP^{sc} has more β strands forming aggregates.





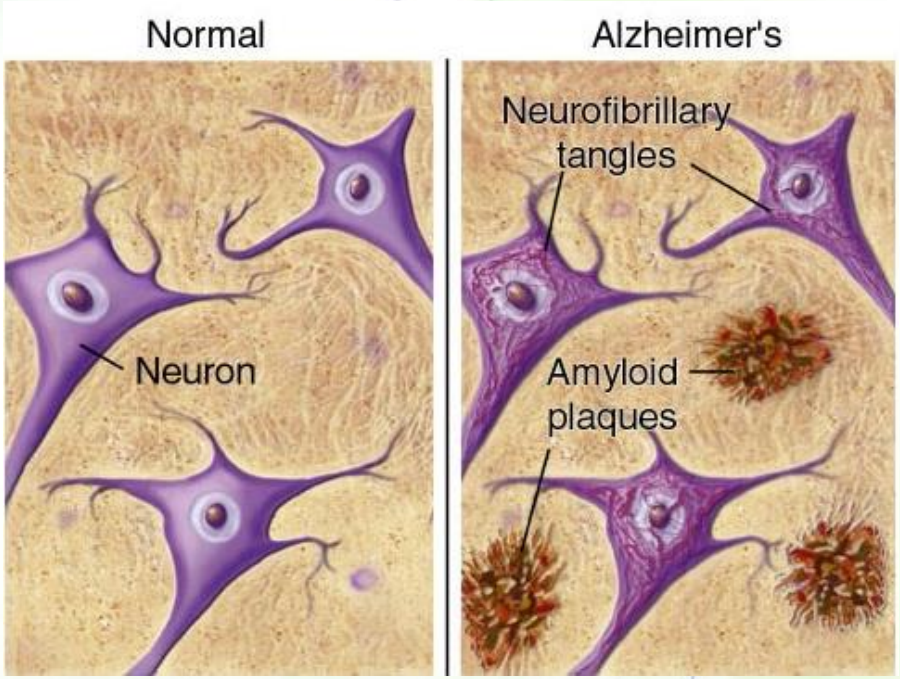
The prion protein

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

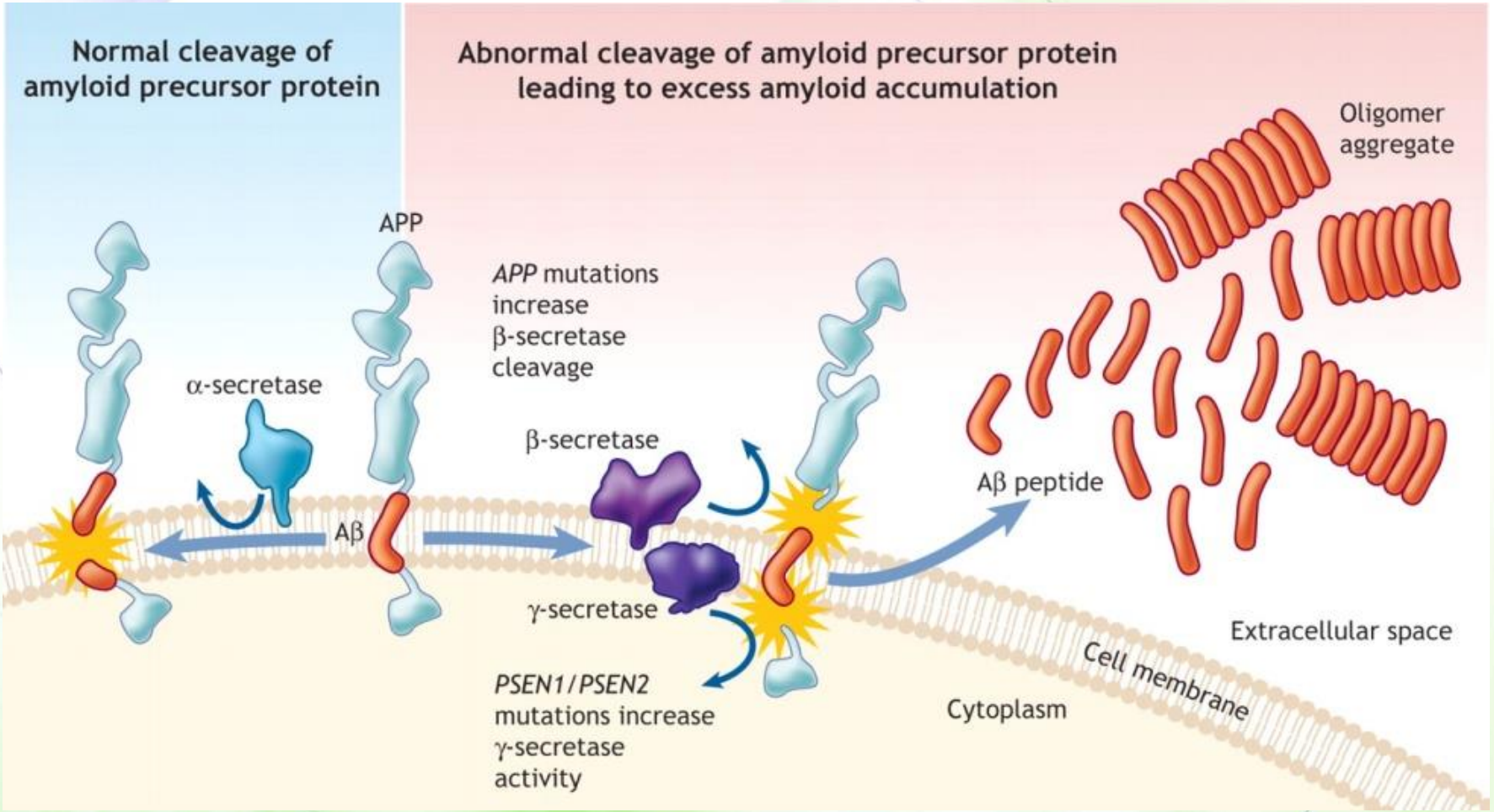
Alzheimer's Disease

- 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



Quaternary structure

What is it?

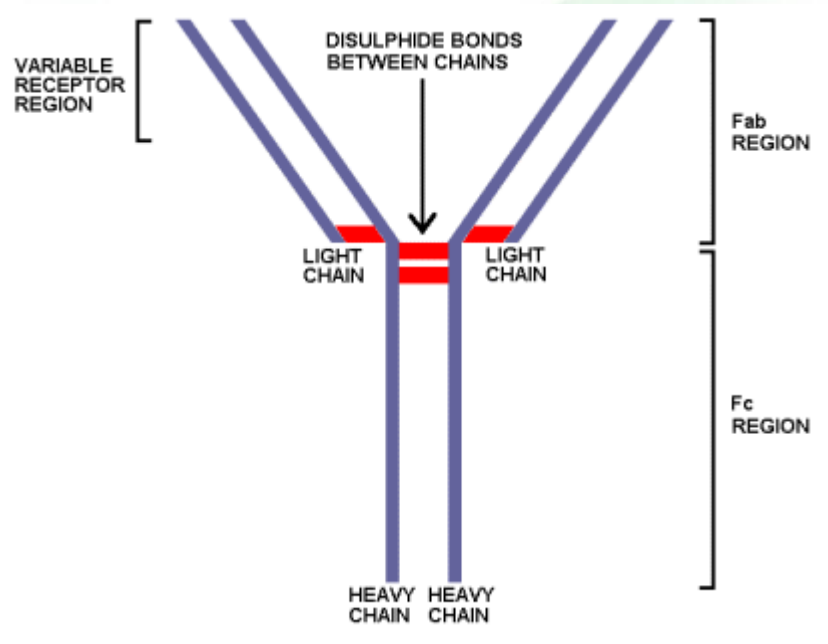
- Proteins are composed of more than one polypeptide chain.
  - They are oligomeric proteins (oligo = a few or small or short; mer = part or unit)
- The spatial arrangement of subunits and the nature of their interactions.
- Proteins made of
  - One subunit = monomer
  - Two subunits: dimer
  - Three subunts: trimer
  - Four subunit: tetramer
  - …etc

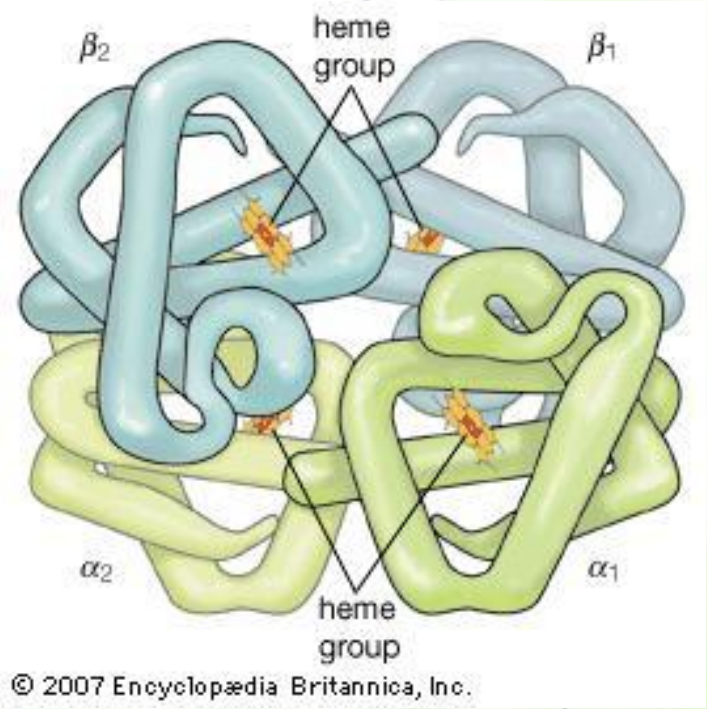
Naming of structures

- Each polypeptide chain in such a protein is called a subunit.
- Oligomeric proteins can be made of multiple polypeptides that are
  - identical → homooligomers (homo = same), or
  - different → heterooligomers (hetero = different)
- The simplest: a homodimer

How are the subunits connected?

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





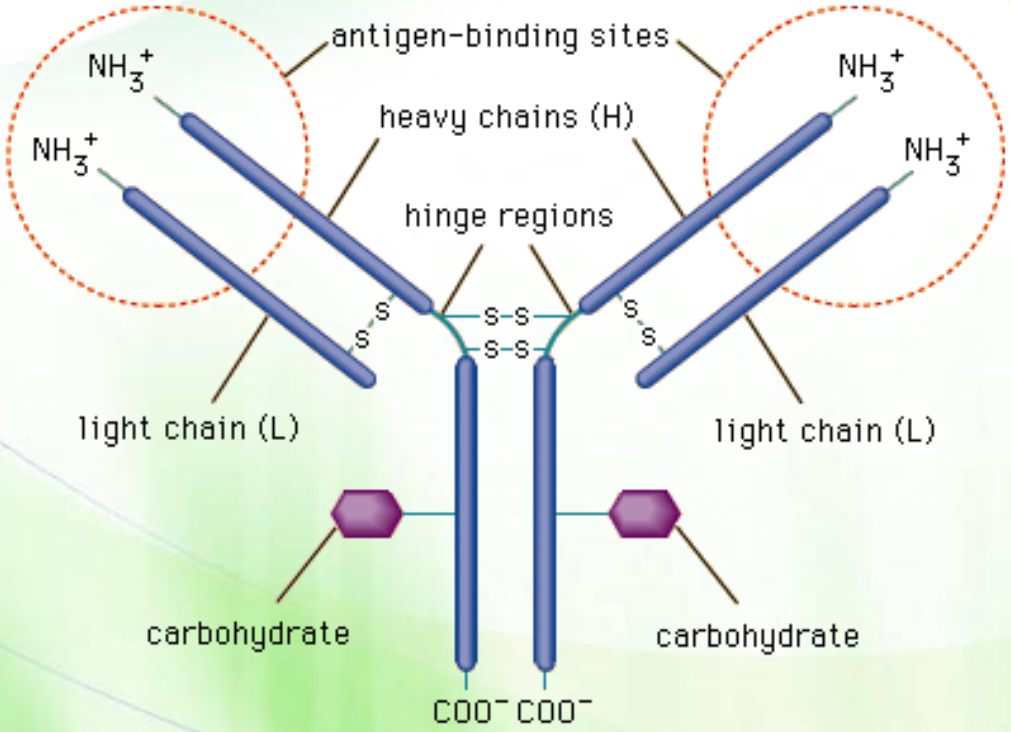
Complex protein structures

Holo- and apo-proteins

- Sometimes, proteins are linked (conjugated) to non-protein molecules. Proteins are known as holoproteins.
- If the non-protein component is removed, the protein is known as an apoprotein.

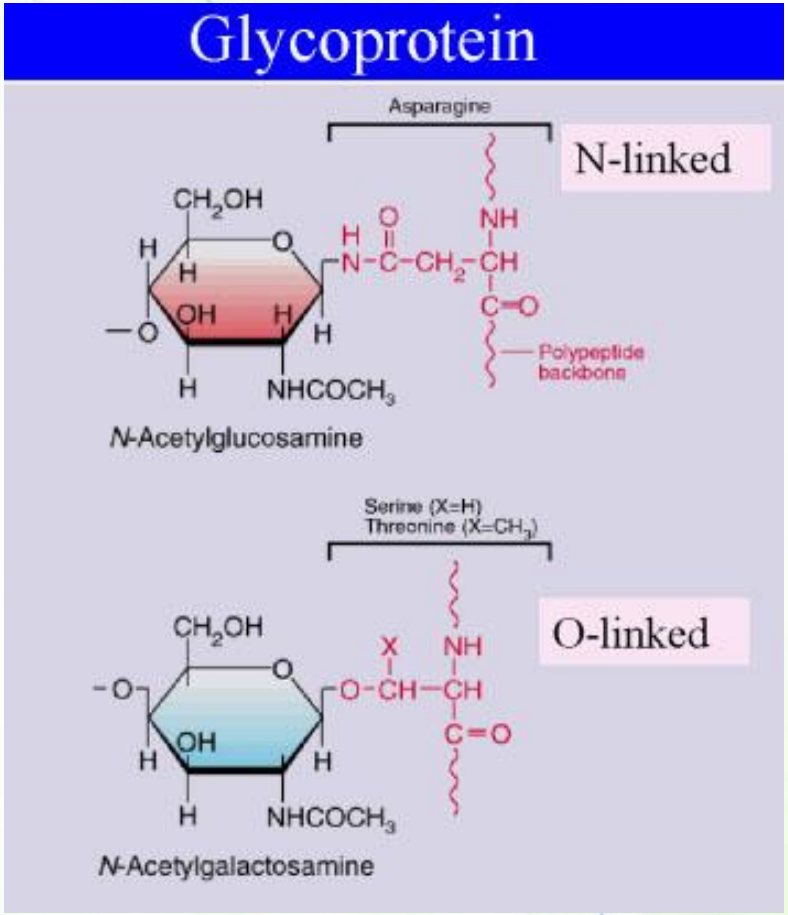
Glycoproteins

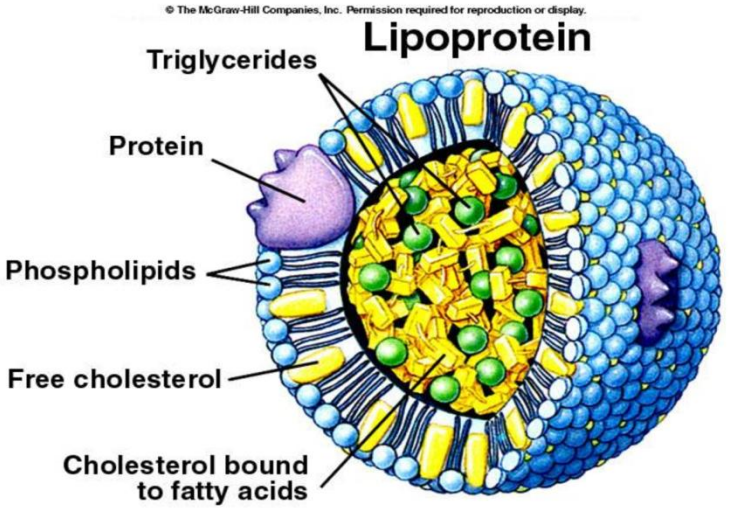
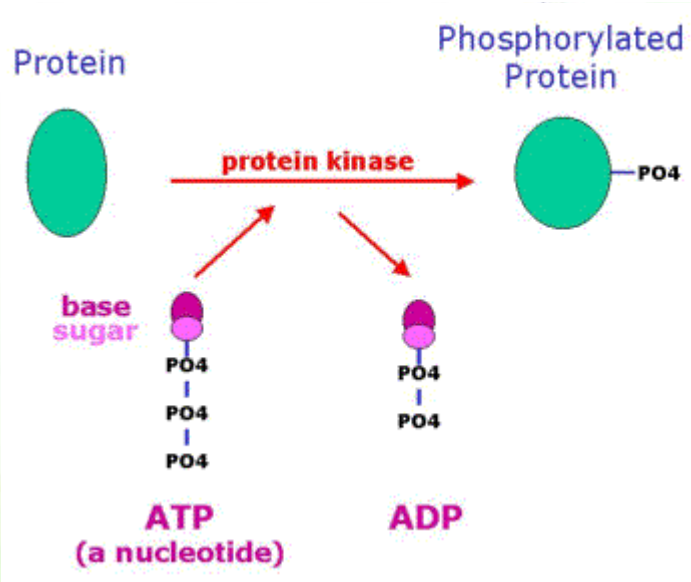
- Proteins also are found to be covalently conjugated with carbohydrates.
- Proteins covalently associated with carbohydrates are termed glycoproteins.



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



Others

- Proteins can also be associated with lipid and are termed lipoproteins.
- Other proteins are phsophorylated and these are known as phosphoproteins.