BME 2901-BIOCHEMISTRY

Three-Dimensional Structure of Proteins

by Assist. Prof. Görke Gürel Peközer

Yıldız Technical University
Biomedical Engineering Department
Fall 2019

Previously on Biochemistry Class…

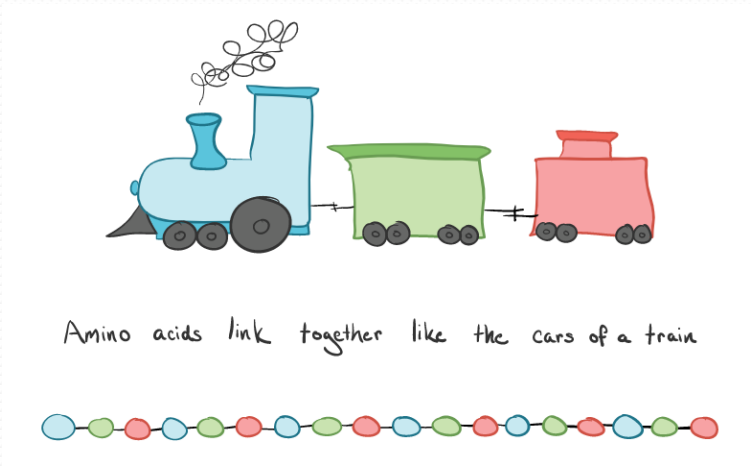
- Proteins are the end products of the decoding process that starts with the information in cellular DNA.

  DNA → RNA → Protein

- Proteins compose structural and motor elements in the cell, they aid in the storage, signaling, sensory functions and they serve as the catalysts for virtually every biochemical reaction that occurs in living things.
- The multiplicity of functions carried out by proteins, derived from a startingly simple code, arises from the huge number of different shapes they adopt.

Previously on Biochemistry Class…

- Each gene in cellular DNA contains the code for a unique protein structure.
- The building blocks of proteins are amino acids, which are small organic molecules that consist of an alpha (central) carbon atom linked to an amino group, a carboxyl group, a hydrogen atom, and a variable component called a radical group.
- Within a protein, multiple amino acids are linked together by **peptide bonds**, thereby forming a long chain.



Previously on Biochemistry Class…

- Different proteins are assembled with peptide bonding of different amino acid sequences.
- They are also held together by weak bonds and folded into a variety of three-dimensional structures or conformations.
- **Conformation** is a spatial arrangement of atoms that depends on the rotation of a bond or bonds. The *conformation* of a molecule, such as a protein, can change without breaking covalent bonds whereas the various **configurations** of a molecule can be changed only by breaking and re-forming covalent bonds.

Protein Conformation

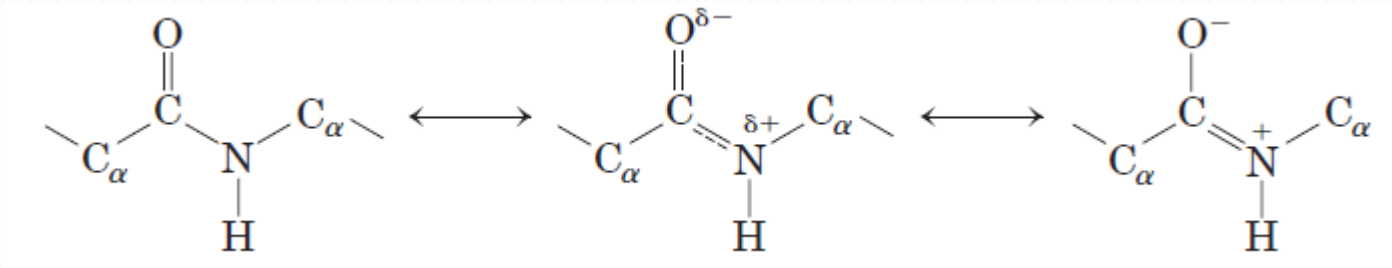
- Each type of protein has a particular three-dimensional structure, which is determined by the order of the amino acids in its polypeptide chain.
- The final folded structure, or **conformation**, adopted by any polypeptide chain is determined by energetic considerations:
  - A protein generally folds into the shape in which its Gibbs free energy (G) is minimized.
  - The conformations existing under a given set of conditions are usually the ones that are thermodynamically the most stable.
  - The folding process is thus energetically favorable, as it releases heat and increases the disorder of the universe.
- A protein's conformation is stabilized by disulfide bonds and weak interactions.

Constraints on possible conformations

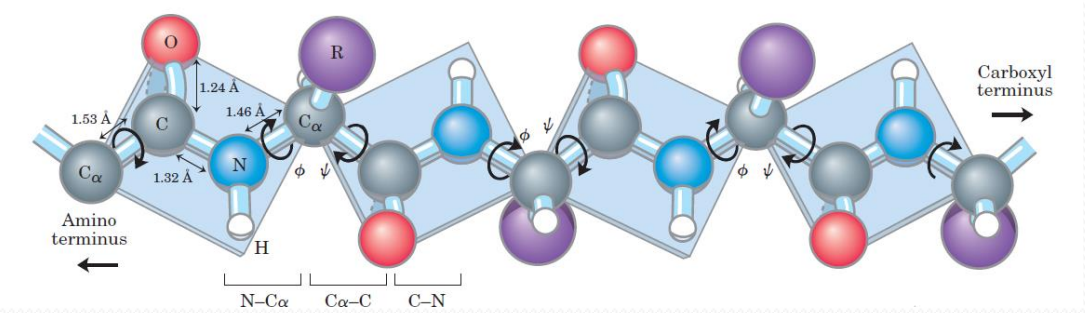
- Each protein can adopt huge number of potential conformations since every amino acid residue has a number of possible conformations and since there are many residues in a protein.
- However, under physiological conditions most proteins fold into a single stable shape known as its **native conformation**.
- The biological function of a protein depends on its native three-dimensional conformation.
- A number of factors constrain rotation around the covalent bonds in a polypeptide chain in its native conformation.
    - These include the allowed rotation about the certain covalent bonds, steric hinderence, presence of hydrogen bonds and other weak interactions between amino acid residues.

Constraints on possible conformations

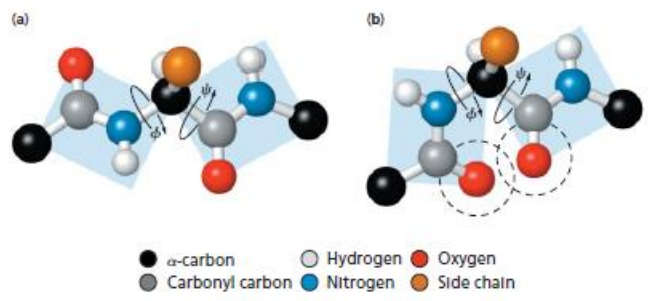
- The carbons of adjacent amino acid residues are separated by three covalent bonds, arranged as C_α—C—N—C_α.
- Peptide C—N bonds are unable to rotate freely because of their partial double-bond character.
- Rotation is permitted about the N—C_α and the C_α—C bonds.
- The backbone of a polypeptide chain can thus be pictured as a series of rigid planes with consecutive planes sharing a common point of rotation at C.
- The rigid peptide bonds limit the range of conformations that can be assumed by a polypeptide chain.

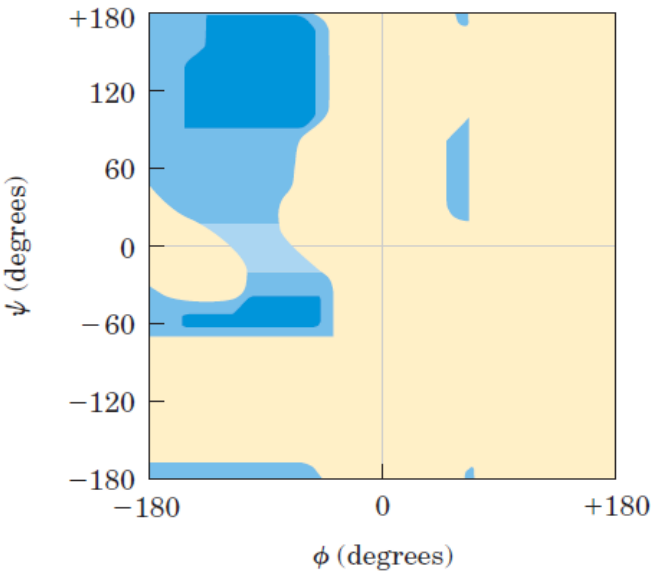


- By convention, the bond angles resulting from rotations at C_α are labeled *as* φ (phi) for the N—C_α bond and ψ (psi) for the C_α—C bond.
- Again by convention, both φ and ψ are defined as 180° when the polypeptide is in its fully extended conformation and all peptide bonds are in the same plane.
- In principle, and can have any value between -180° and 180°, but many values are prohibited by steric interference between atoms in the polypeptide backbone and amino acid side chains.
- Allowed values for and are graphically revealed when φ is ψ plotted versus in a **Ramachandran plot.**



- Sterically permited values of φ and ψ are plotted in Ramachandran plot.
- Conformations deemed possible are those that involve little or no steric interference, based on calculations using known van der Waals radii and bond angles.
- The areas shaded dark blue reflect conformations that involve no steric overlap and thus are fully allowed; medium blue indicates conformations allowed at the extreme limits for unfavorable atomic contacts; the lightest blue area reflects conformations that are permissible if a little flexibility is allowed in the bond angles.
- The asymmetry of the plot results from the L stereochemistry of the amino acid residues. The plots for other
- L-amino acid residues with unbranched side chains are nearly identical.
- The allowed ranges for branched amino acid residues such as Val, Ile, and Thr are somewhat smaller than for Ala.
- The Gly residue, which is less sterically hindered, exhibits a much broader range of allowed conformations.
- The range for Pro residues is greatly restricted because is limited by the cyclic side chain to the range of -35° to -85°.

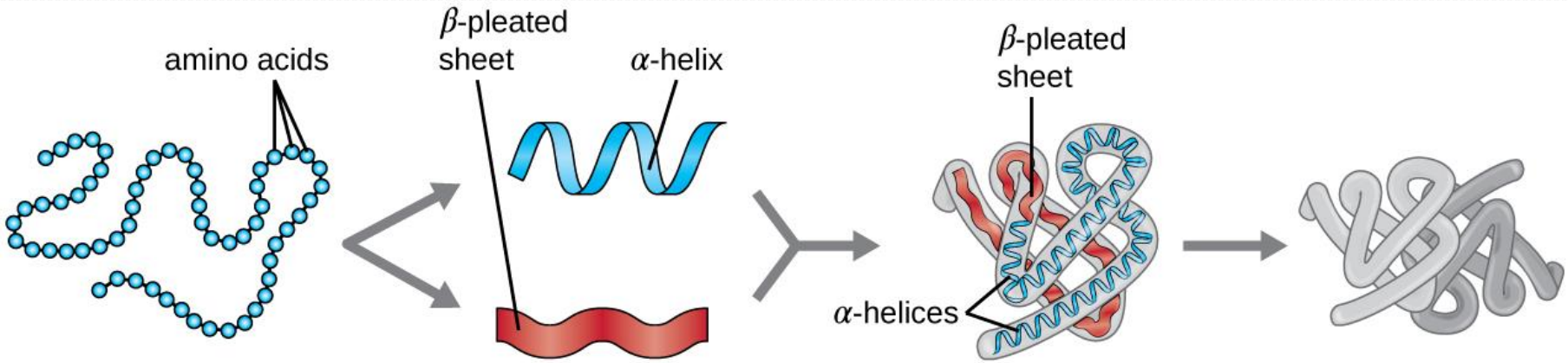




Levels of Protein Structure

- For proteins, the tasks of describing and understanding structure are approached at several levels of complexity.
- Four levels of protein structure are commonly defined:
  - The linear sequence of amino acids within a protein is considered the **primary structure** of the protein.
  - **Secondary structure** refers to particularly stable arrangements of amino acid residues giving rise to recurring structural patterns.
  - **Tertiary structure** describes all aspects of the three-dimensional folding of a polypeptide.
  - When a protein has two or more polypeptide subunits, their arrangement in space is referred to as **quaternary structure.**

Levels of Protein Structure



Primary Protein Structure
Sequence of a chain of amino acids

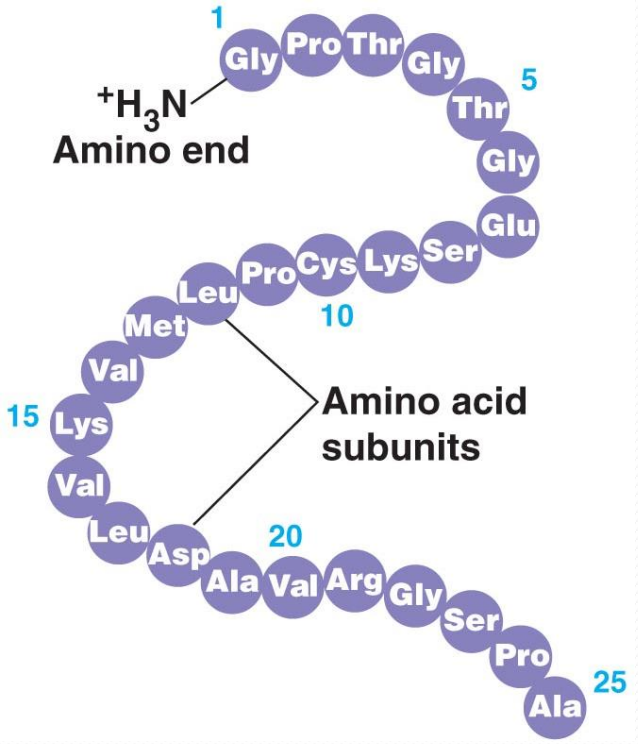
Secondary Protein Structure
Local folding of the polypeptide chain into helices or sheets

Tertiary Protein Structure
three-dimensional folding pattern of a protein due to side chain interactions

Quaternary Protein Structure
protein consisting of more than one amino acid chain

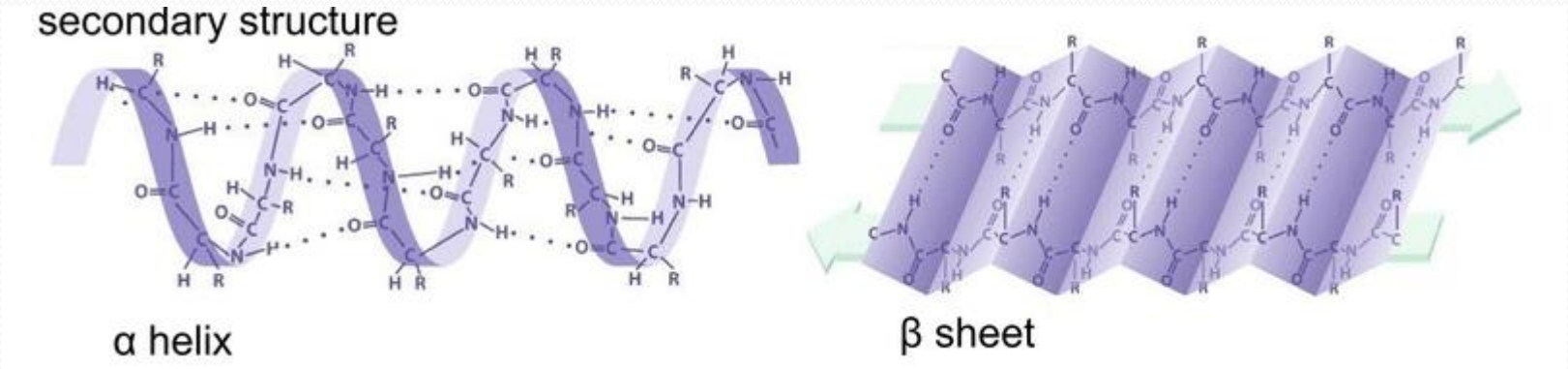
Primary Structure

- Protein's structure begins with its amino acid sequence, which is thus considered its primary structure.



Secondary Structure

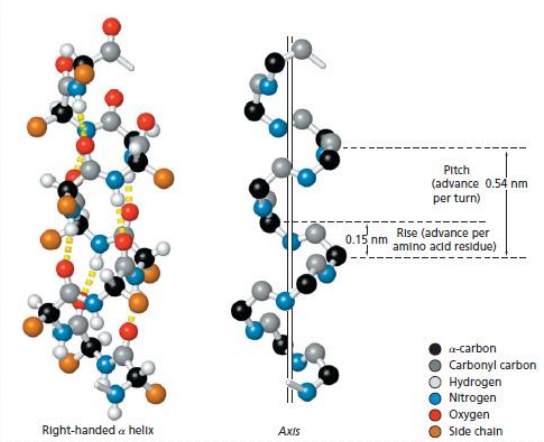
- Although the overall conformation of each protein is unique, some regular folding patterns can be detected.
- α helix and β sheets are the most stable and prominent secondary structures in proteins.

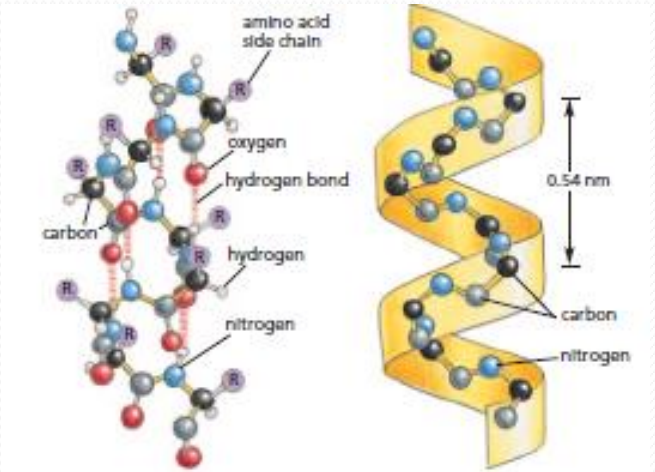


α-helix

- The simplest arrangement the polypeptide chain could assume with its rigid peptide bonds (but other single bonds free to rotate) is a helical structure, which is called the **α helix.**
- α helix is a regular structure that resembles a spiral staircase.

- In α helical structure, the polypeptide backbone is tightly wound around an imaginary axis drawn longitudinally through the middle of the helix, and the R groups of the amino acid residues protrude outward from the helical backbone.
- The repeating unit is a single turn of the helix, which extends about 5.4 Angstrom (Å) along the long axis.
- A hydrogen bond is made between every 4th amino acid, linking the C=O of one peptide bond to the N–H of another in the peptide backbone.
- This gives rise to a regular right handed helix with a complete turn every 3.6 amino acids.





Stability of α helix

The stability of the α helix depends on:

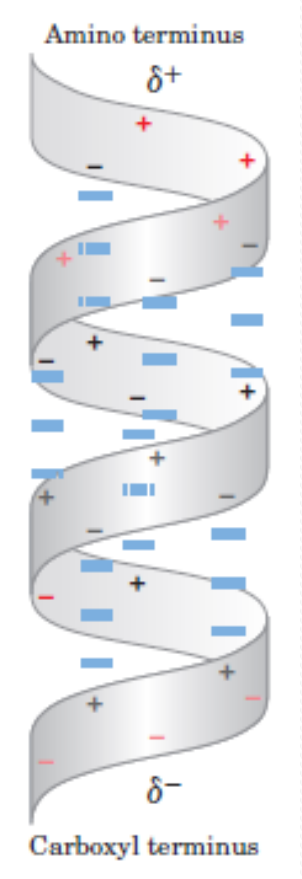
(1) the electrostatic repulsion (or attraction) between successive amino acid residues with charged R groups

(2) the bulkiness of adjacent R groups

(3) the interactions between R groups spaced three (or four) residues apart

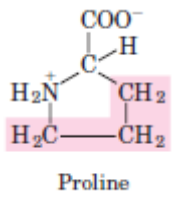
(4) the occurrence of Pro and Gly residues

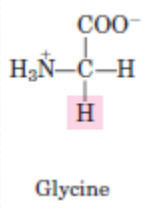
(5) the interaction between amino acid residues at the ends of the helical segment and the electric dipole inherent to the helix.

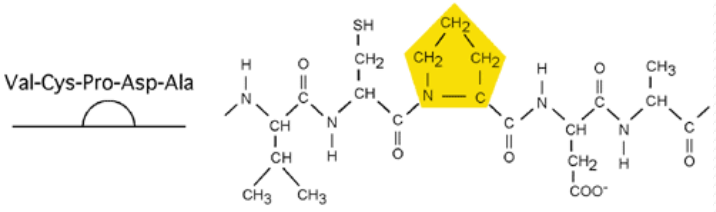


Proline and glycine are uncommon in α helices

- In proline, the nitrogen atom is part of a rigid ring, and rotation about the N—Cα bond is not possible. Thus, a Pro residue introduces a destabilizing kink in an α helix.
- In addition, the nitrogen atom of a Pro residue in peptide linkage has no substituent hydrogen to participate in hydrogen bonds with other residues.
- For these reasons, proline is only rarely found within an helix.
- Glycine occurs infrequently in helices for a different reason: it has more conformational flexibility than the other amino acid residues. Polymers of glycine tend to take up coiled structures quite different from an helix.

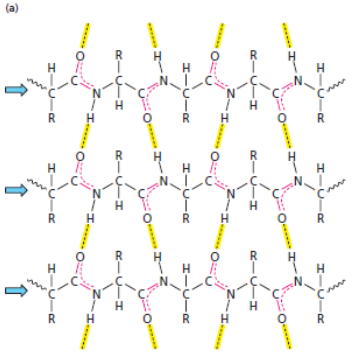


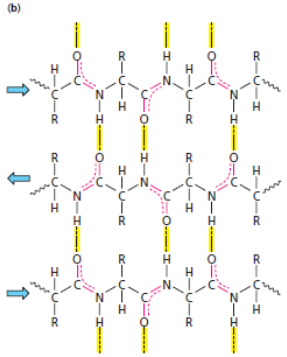


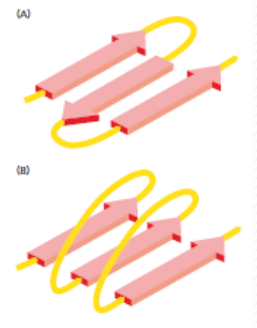


β sheets

- A β sheet is made when hydrogen bonds form between segments of a polypeptide chain that lie side by side.
- When the neighboring segments run in the same orientation (say, from the N-terminus to the C-terminus), the structure is a *parallel β sheet*; when they run in opposite directions, the structure is an *antiparallel β sheet*.
- Both types of β sheet produce a very rigid, pleated structure, and they form the core of many proteins.





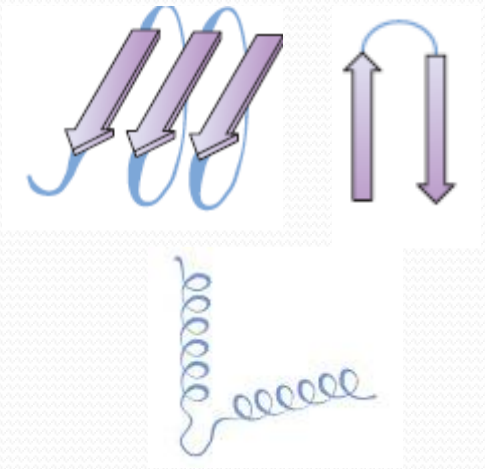


- β sheets have remarkable properties. They give **silk fibers** their extraordinary tensile strength. They also permit the formation of ***amyloid fibers***—insoluble protein aggregates that include those associated with neurodegenerative disorders, such as **Alzheimer's disease** and **prion diseases**.
- These structures, formed from abnormally folded proteins, are stabilized by β sheets that stack together tightly, with their amino acid side chains interdigitated like the teeth of a zipper.

Loops and Turns

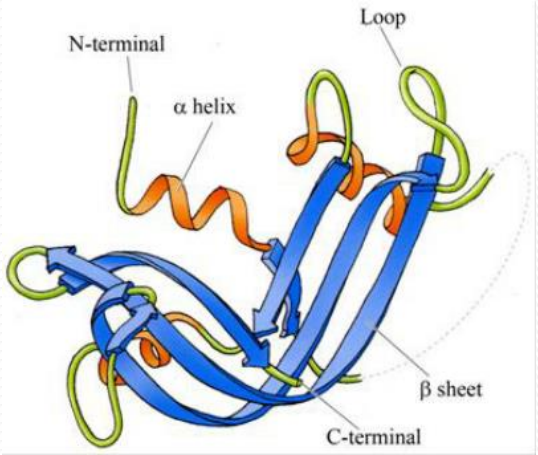
- In proteins, which have a compact folded structure, nearly one-third of the amino acid residues are in turns or loops where the polypeptide chain reverses direction.
- These are the connecting elements that link successive runs of α helix or β conformation.
- Loops and turns connect a helices and b strands and allow the polypeptide chain to fold back on itself producing the compact three-dimensional shape seen in the native structure.

Loops often contain hydrophilic residues and are usually found on the surfaces of proteins where they are exposed to solvent and form hydrogen bonds with water.
Loops containing only a few (up to five) residues are referred to as **turns** if they cause an abrupt change in the direction of a polypeptide chain.
The most common types of turns are β **turns. Gly and Pro are the most abundant aminoacids in the turns.**
Of the peptide bonds between amino acid residues other than Pro, over **99.95% are in the trans configuration**. However, about 6% of proline are in the cis configuration; many of these occur at turns.



Tertiary Structure

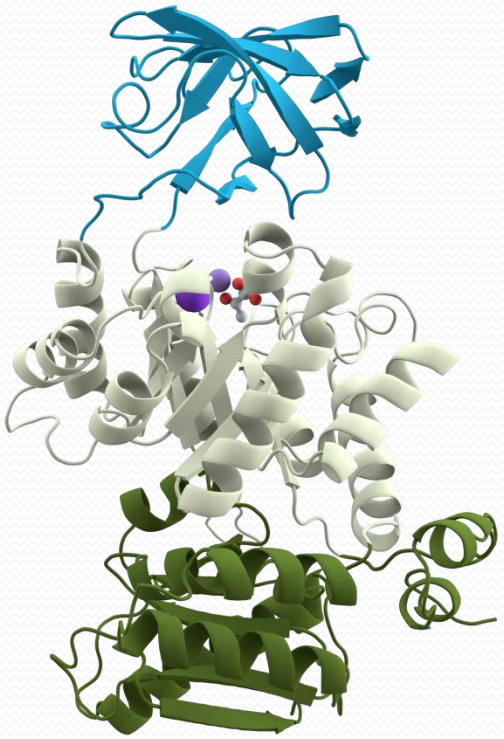
- The overall three-dimensional arrangement of all atoms in a protein is referred to as the protein's **tertiary structure.**
- Tertiary structure describes the completely folded and compacted polypeptide chain.



- Secondary structure refers to the spatial arrangement of amino acid residues that are adjacent in the primary structure, however, tertiary structure includes *longer-range* aspects of amino acid sequence.
- Tertiary structures are stabilized by the interactions of amino acid side chains in nonneighboring regions of the polypeptide chain. The formation of tertiary structure brings distant portions of the primary and secondary structures close together.
- Amino acids that are far apart in the polypeptide sequence and that reside in different types of secondary structures may interact within the completely folded structure of a protein.

- Many folded polypeptides consist of several distinct units, called **domains,** linked by a short stretch of amino acid residues.
- Domain is defined as any segment of a polypeptide chain that can fold independently into a compact, stable structure.
- Domain is the modular unit from which many larger proteins are constructed.
- The different domains of a protein are often associated with different functions such as catalytic domain, DNA binding domain etc.

Pyruvate kinase



Substrate binding domain

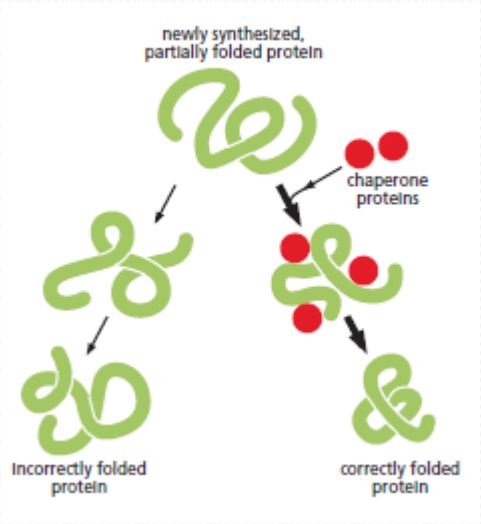
Ligand binding domain

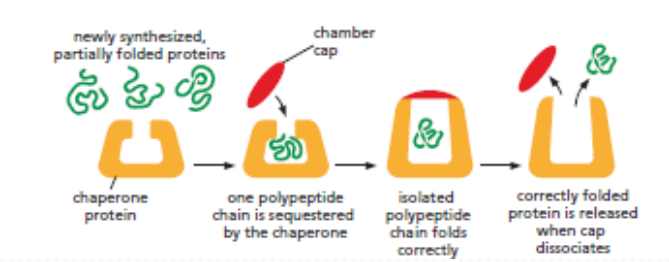
Regulatory domain

Folding Event

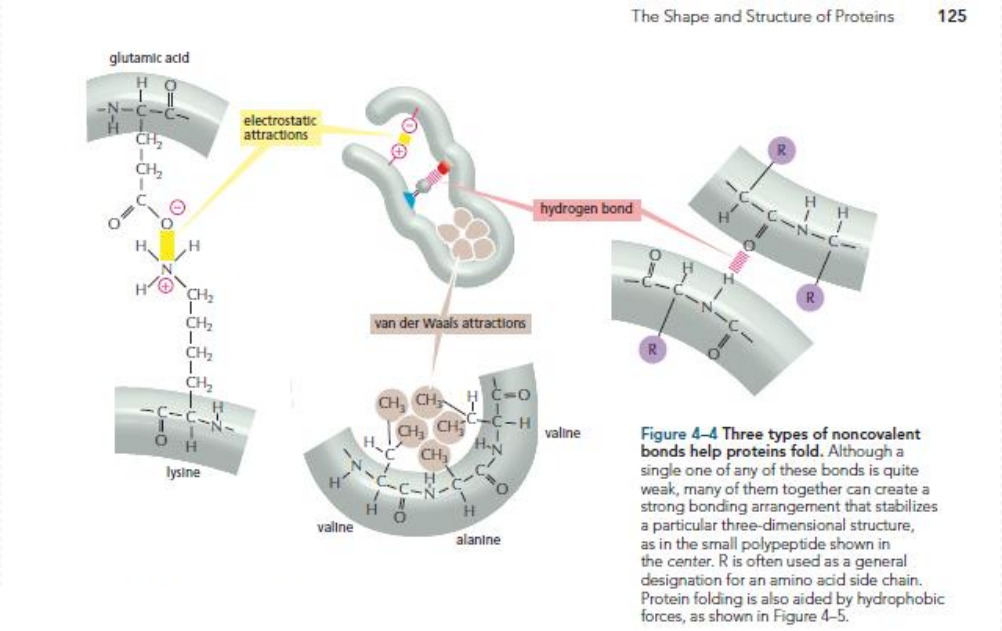
- As the newly synthesized polypeptide emerges from the ribosome, it folds into its characteristic three-dimensional shape.
- Folded proteins occupy a low-energy well that makes the native structure much more stable than alternative conformations.
- Folding is extremely rapid—in most cases the native conformation is reached in less than a second.
- As a protein folds the first few interactions trigger subsequent interactions. This is an example of cooperative effects in protein where the formation of one part of a structure leads to the formation of the remaining parts of the structure.
- As the protein begins to fold, it adopts lower and lower energies.
- In its final, stable, conformation, the native protein is much less sensitive to degradation than an extended, unfolded polypeptide chain. Thus, native proteins can have very long half-lives of many generations.

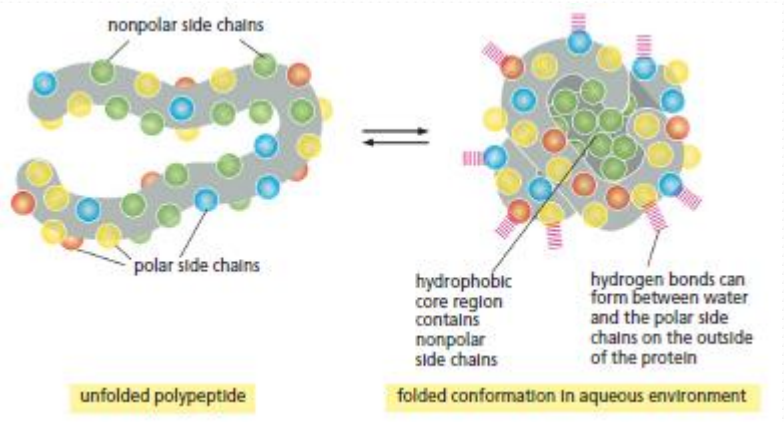
- Thus, folding event is rapid and spontaneous and dictated by the aminoacid sequence of the protein.
- However, chaperons generally assist this folding event.
- Some of these chaperones bind to partly folded chains and help them to fold along the most energetically favorable pathway.
- Others form “isolation chambers” in which single polypeptide chains can fold without the risk of forming aggregates in the crowded conditions of the cytoplasm.
- Chaperones make the folding process more efficient and reliable.





- Protein folding and stabilization depend on several noncovalent forces including the hydrophobic effect, hydrogen bonding, van der Waals interactions, and charge–charge interactions.
- Although noncovalent interactions are weak individually, collectively they account for the stability of the native conformations of proteins.
- The weakness of each noncovalent interaction gives proteins the resilience and flexibility to undergo small conformational changes.

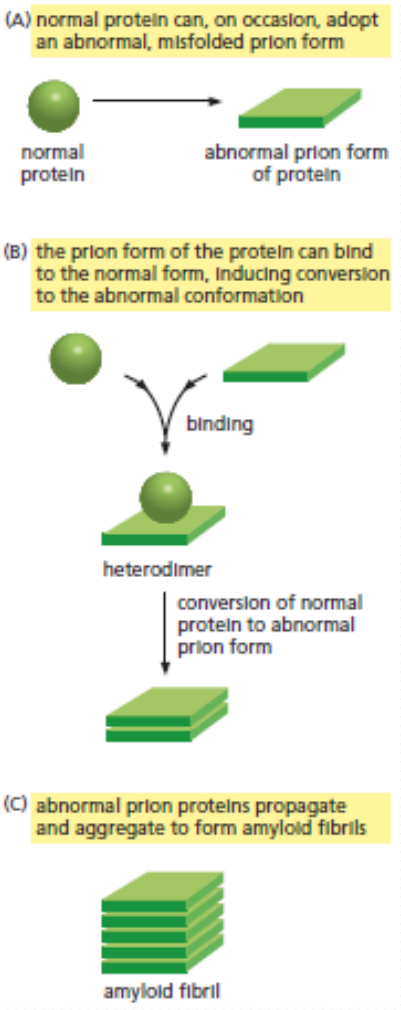




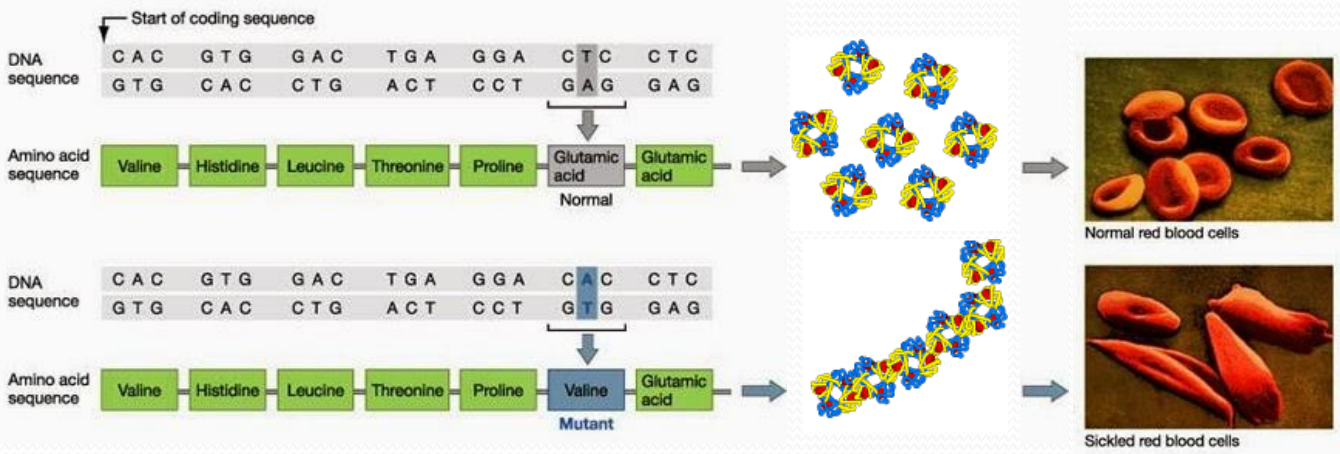
Errors in protein folding

- Changing the structure of the machine can alter its function. A plane can't fly unless all its parts are put together in the right way.
- The same is true for proteins. Many of the diseases are the consequences of errors in protein structure: something happens in the body that causes a protein to lose an aspect of its native conformation, and this change of structure causes a loss or a detrimental change of function.

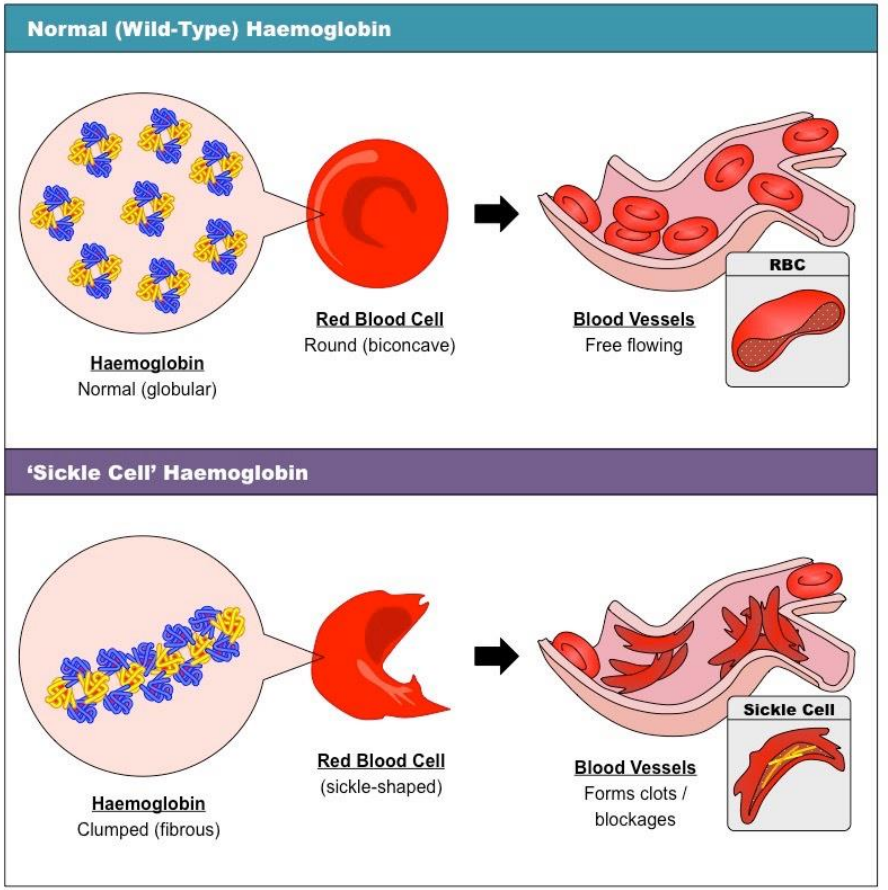
- Each protein normally folds into a single stable conformation. This conformation, however, sometimes can change slightly when the protein interacts with other molecules in the cell. This change in shape is crucial to the function of the protein.
- When proteins fold incorrectly, they sometimes form aggregates that can damage cells and even whole tissues.
- Misfolded proteins are thought to contribute to a number of neurodegenerative disorders, such as Alzheimer's disease and Huntington's disease. Some infectious neurodegenerative diseases—including mad cow disease in cattle, and Creutzfeldt– Jakob disease (CJD) in humans—are caused by misfolded proteins called **prions**.
- The misfolded prion form of a protein can convert the properly folded version of the protein in an infected brain into the abnormal conformation.
- This allows the misfolded prions, which tend to form aggregates, to spread rapidly from cell to cell, eventually causing the death of the affected animal or human. Prions are considered "infectious" because they can also spread from an affected individual to a normal individual via contaminated food, blood, or surgical instruments.



- Because the structure and thus activity of each protein depend on its amino acid sequence, a protein with an altered sequence may function poorly or not at all.
- A mutation that affects just a single nucleotide pair can severely compromise an organism's fitness if the change occurs in a vital position in the DNA sequence.
- For example, a change in a single nucleotide in a hemoglobin gene can cause cells to make hemoglobin with an incorrect sequence of amino acids which causes the disease *sickle-cell anemia*.
- The sickle-cell hemoglobin is less soluble than normal hemoglobin and forms fibrous intracellular precipitates, which produce the characteristic sickle shape of affected red blood cells.

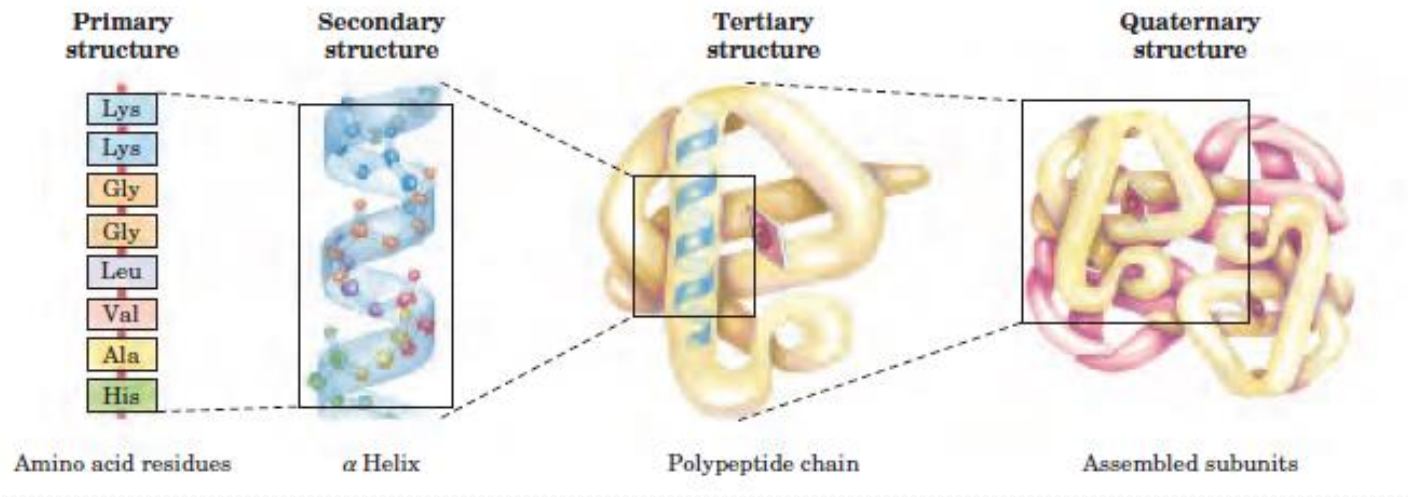


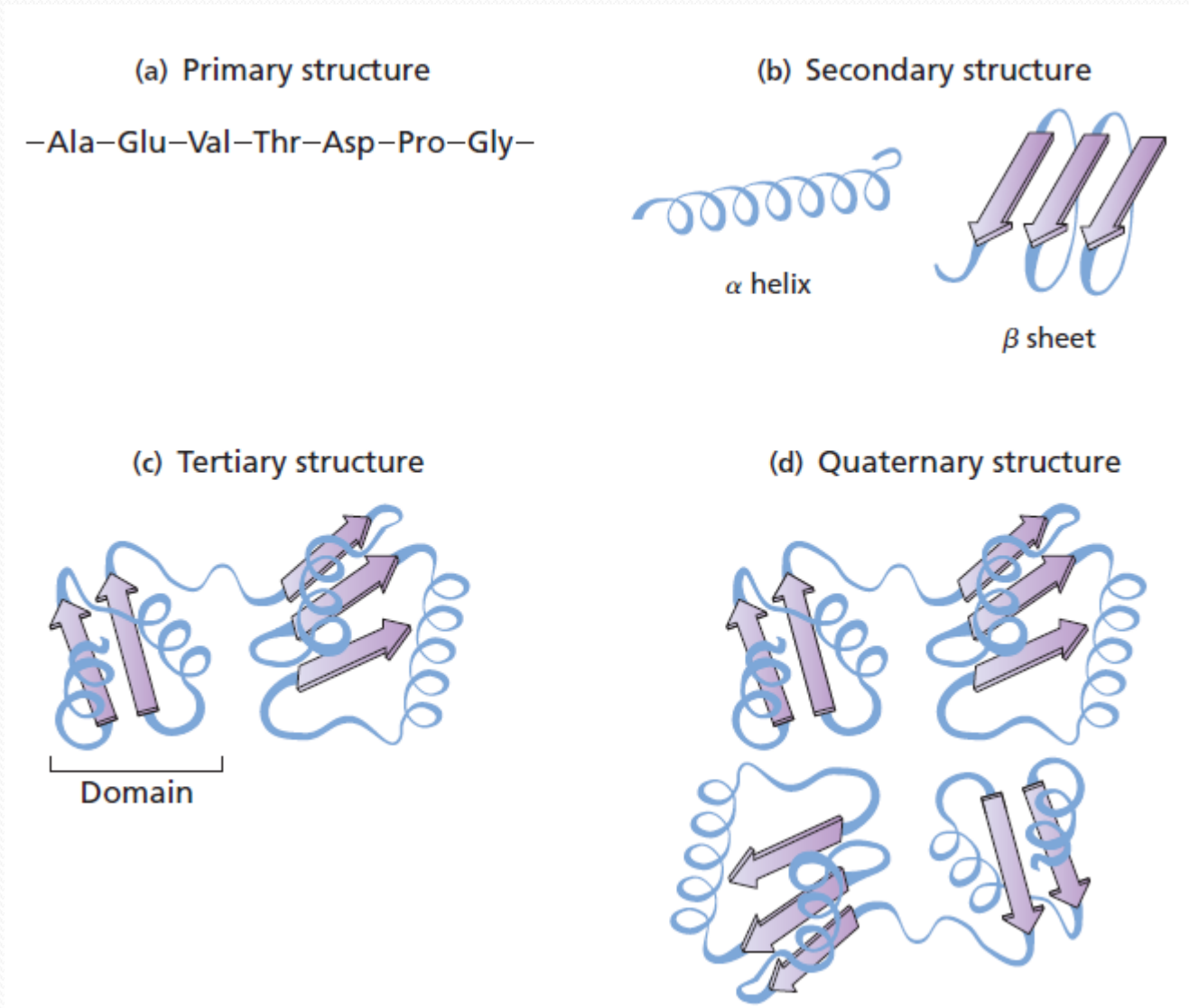
- Because these cells are more fragile and frequently tear as they travel through the bloodstream, patients with this potentially life-threatening disease have fewer red blood cells than usual—that is, they are anemic.
- This anemia can cause weakness, dizziness, headaches, and breathlessness. Moreover, the abnormal red blood cells can aggregate and block small vessels, causing pain and organ failure.



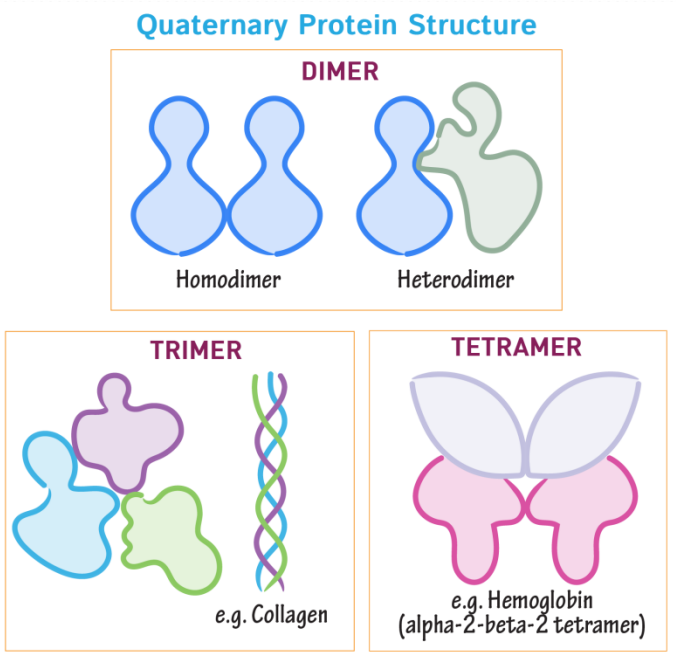
Quaternary Structure

- Some proteins contain two or more separate polypeptide chains, or subunits, which may be identical or different.
- The arrangement of these protein subunits in three-dimensional complexes constitutes **quaternary structure.**




(a) Primary structure
–Ala–Glu–Val–Thr–Asp–Pro–Gly–
(b) Secondary structure
α helix
β sheet
(c) Tertiary structure
Domain
(d) Quaternary structure

- If the final protein is made of two subunits, the protein is said to be a dimer.
- If three subunits must come together, the protein is said to be a trimer, four subunits make up a tetramer, etc.
- If the subunits are identical, the prefix “homo” is used, as in “homodimer.”
- If the subunits are different, we use “hetero,” as in “heterodimer.”

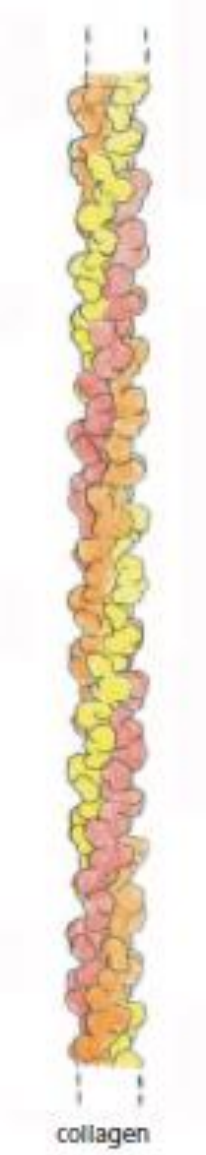


Higher Levels of Protein Structure

- In considering these higher levels of structure, it is useful to classify proteins into two major groups: **fibrous proteins,** having polypeptide chains arranged in long strands or sheets, and **globular proteins,** having polypeptide chains folded into a spherical or globular shape.
- The two groups are structurally distinct: fibrous proteins usually consist largely of a single type of secondary structure; globular proteins often contain several types of secondary structure.
- The two groups differ also functionally in that the structures that provide support, shape, and external protection to vertebrates are made of fibrous proteins, whereas most enzymes and regulatory proteins are globular proteins.

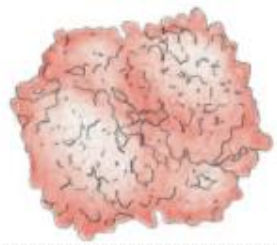
Fibrous proteins

- Fibrous proteins are made up of polypeptide chains that are elongated and fibrous in nature or have a sheet like structure.
- Fibrous proteins share properties that give mechanical strength and/or flexibility to the structures in which they occur.
- In each case, the fundamental structural unit is a simple repeating element of secondary structure (α helix or β sheet).
- All fibrous proteins are insoluble in water, a property conferred by a high concentration of hydrophobic amino acid residues both in the interior of the protein and on its surface.
- These hydrophobic surfaces are largely buried by packing many similar polypeptide chains together to form elaborate supramolecular complexes.
- Collagen, keratin and silk are among those.



Globular proteins

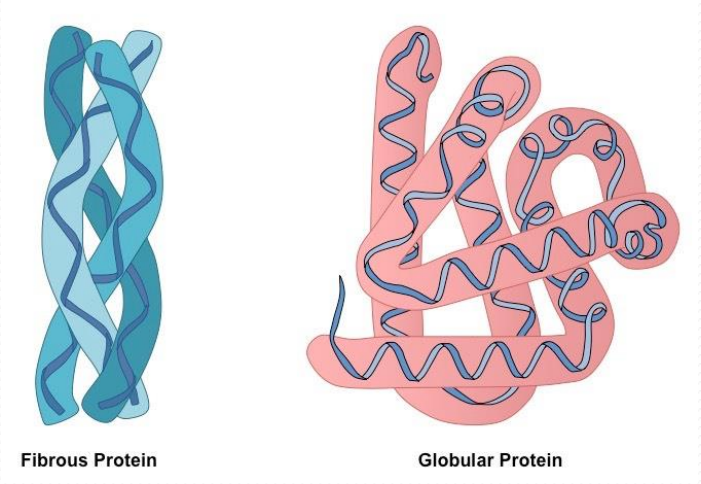
- They are generally water-soluble, compact and roughly spherical macromolecules whose polypeptide chains are tightly folded. This folding generates a compact form.
- This folding provides the structural diversity necessary for proteins to carry out a wide array of biological functions.
- Characteristically, they have a hydrophobic interior and a hydrophilic surface.
- They possess indentations or clefts that specifically recognize and transiently bind other compounds. By selectively binding other molecules these proteins serve as dynamic agents of biological action.
- Globular proteins include enzymes, transport proteins, motor proteins, regulatory proteins, immunoglobulins, and proteins with many other functions.



Hemoglobin

Comparison of Fibrous and Globular Proteins

	Fibrous	Globular
Shape	Long and narrow	Round / spherical
Purpose	Structural	Functional
Acid Sequence	Repetitive amino acid sequence	Irregular amino acid sequence
Durability	Less sensitive to changes in pH, temperature, etc.	More sensitive to changes in pH, temperature, etc.
Examples	Collagen, myosin, fibrin, actin, keratin, elastin	Enzymes, haemoglobin, insulin, immunoglobulin
Solubility	(Generally) insoluble in water	(Generally) soluble in water



Fibrous Protein

Globular Protein

Denaturation

- Protein structures have evolved to function in particular cellular environments.
- Conditions different from thosemin the cell can result in protein structural changes, large and small.
- A loss of three-dimensional structure sufficient to cause loss of function is called **denaturation.**
- Proteins can be denatured by heat, extremes of pH, by certain organic solvents such as alcohol or acetone, by certain solutes such as urea and by detergents.
- Organic solvents, urea, and detergents act primarily by disrupting the hydrophobic interactions that make up the stable core of globular proteins; extremes of pH alter the net charge on the protein, causing electrostatic repulsion and the disruption of some hydrogen bonding. Heat also affects hydrogen bonding.

Renaturation

- The tertiary structure of a globular protein is determined by its amino acid sequence.
- The most important proof of this came from experiments showing that denaturation of some proteins is reversible.
- Certain globular proteins denatured by heat, extremes of pH, or denaturing reagents will regain their native structure and their biological activity if returned to conditions in which the native conformation is stable.

Studying 3D Structure of Proteins

- Studying 3D structure of proteins is important for the identification of:
  - Shape and domain structure
  - Protein classification
  - Function for uncharacterized proteins
  - Interactions with other macromolecules
  - Interactions with small ligands: metal ions, nucleotides, substrates, cofactors and inhibitors
  - Enzyme mechanism
  - Structure-based drug development

Assumptions and Deductions

- Proteins with similar sequences have similar 3D-structures .
- Proteins with similar 3D-structure are likely to have similar function (generally true, but exceptions exist).
- Proteins with similar function can have entirely different sequences.

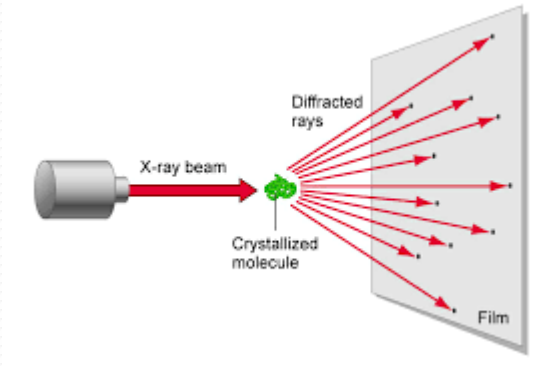
Studying 3D Structure of Proteins

- Protein structure is acquired using both experimental and computational methods.
- 3D structure of proteins can be studied by:
  - X-Ray Crystallography analysis
  - Nuclear Magnetic Resonance (NMR) analysis
  - Protein Structure Prediction

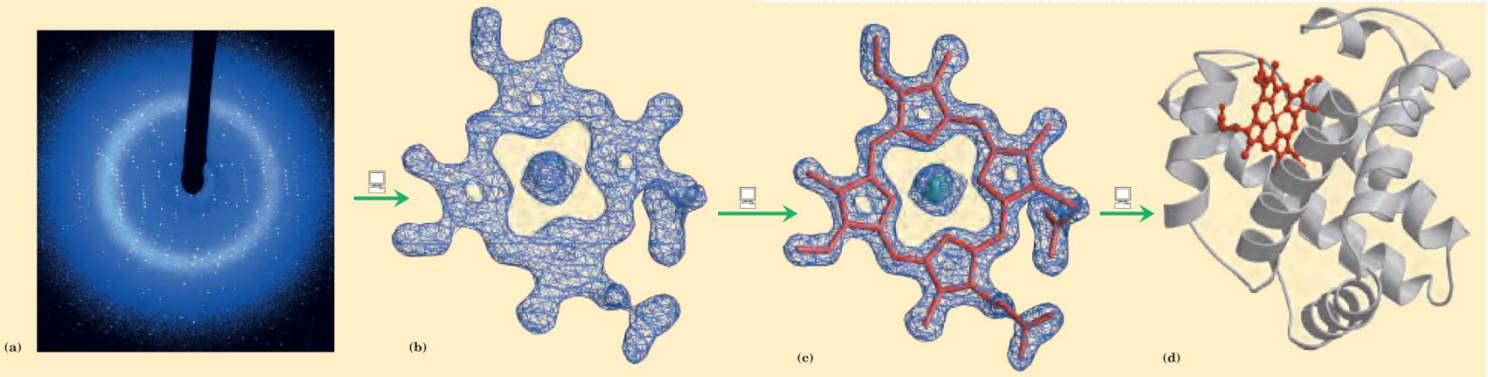
    https://www.rcsb.org/

X-Ray Crystallography

- X-ray diffraction allows measurement of the short distances between atoms and yields a three-dimensional electron density map, which can be used to build a model of the protein structure.
- To determine the three-dimensional structure of a protein by X-ray diffraction, a large, well-ordered single crystal is required.
- The structures of many important proteins are not yet known, simply because they have proved difficult to crystallize.



- Once a crystal is obtained, it is placed in an x-ray beam between the x-ray source and a detector, and a regular array of spots called reflections is generated.
- The spots are created by the diffracted x-ray beam, and each atom in a molecule makes a contribution to each spot.
- An electron-density map of the protein is reconstructed from the overall diffraction pattern of spots by using a mathematical technique called a Fourier transform.
- A model for the structure is then built that is consistent with the electron-density map.

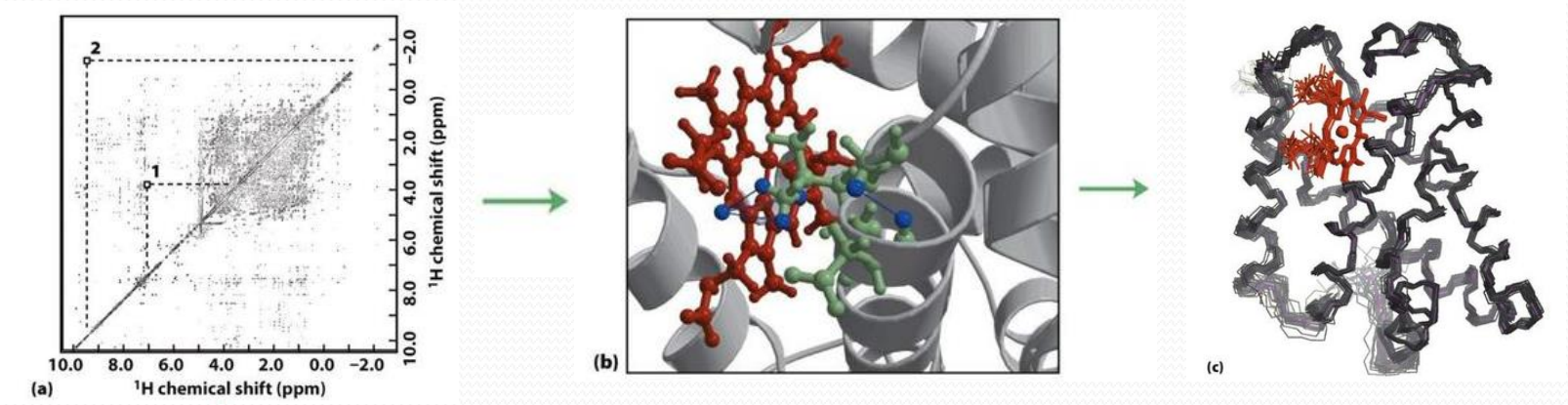


NMR

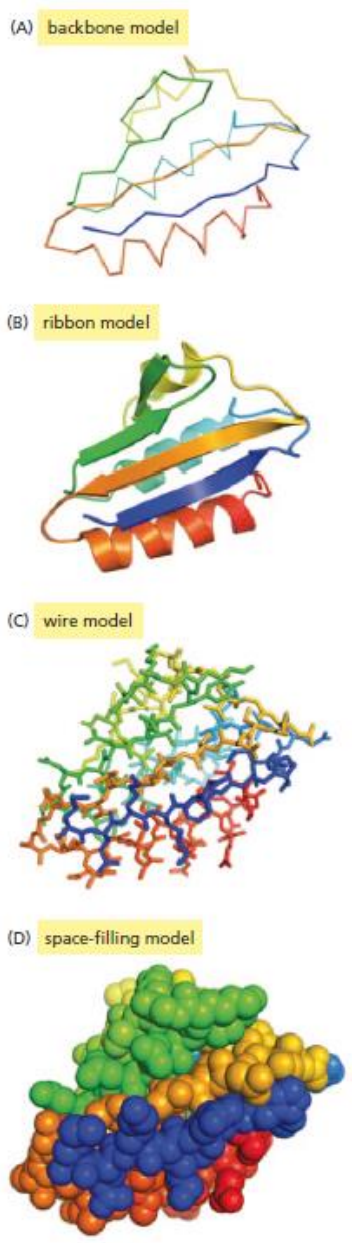
- NMR does not require protein crystals.
- It determines the structure of small proteins, less than 50,000 Daltons,
- A concentrated solution of pure protein is placed in a strong magnetic field and then bombarded with radio waves of different frequencies.
- Depending on the environment of atoms within the protein, the nuclei of individual atoms will absorb different frequencies of radio signals.
- Besides, the absorption signals of different nuclei may be perturbed by adjacent nuclei.
- Hydrogen, in particular, will generate an NMR signal that can be used to determine the distances between these atoms in different parts of the protein.
- This information is then used to build a model of how the hydrogens are arranged in space and the overall structure of the protein.

NMR

- However, even a small protein has hundreds of H atoms thus an NMR spectrum can be too complex for analysis.
- Combined with the known amino acid sequence, an NMR spectrum can allow the computation of the 3D structure of the protein.
- NMR can also illuminate the dynamic side of protein structure, including conformational changes, protein folding, and interactions with other molecules.



- The backbone model shows the overall organization of the polypeptide chain and provides a straightforward way to compare the structures of related proteins.
- The ribbon model shows the polypeptide backbone in a way that emphasizes its various folds.
- The wire model includes the positions of all the amino acid side chains; this view is especially useful for predicting which amino acids might be involved in the protein's activity.
- The space-filling model provides a contour map of the protein surface, which reveals which amino acids are exposed on the surface and shows how the protein might look to a small molecule such as water or to another macromolecule in the cell.



Structure – Function Relationship

- Comparative structure-function analysis of proteins are done at three levels: aminoacid sequence, three dimensional conformation and spatial configuration.
- Sequence similarity corresponds to structure similarity. However, alignment of three-dimensional structures is crucial for studying structure-function relationships in enzymes, which would not be possible by sequence alignment alone. It is usually assumed that orientation of a side chain of amino acid residues around an active site, and packing of a polypeptide chain of enzymes contribute to recognition, binding, and catalysis of reaction
- Computer programs can use the information of a known protein from a database to predict the active site of a new protein.
- Once the program has identified the potential ligand-binding sites, other programs can test the fit and the binding ability of thousands of possible ligand molecules - even theoretical ligands that may not yet exist.
- This has tremendous possibilities for the design of new drugs, particularly for cancer therapy.