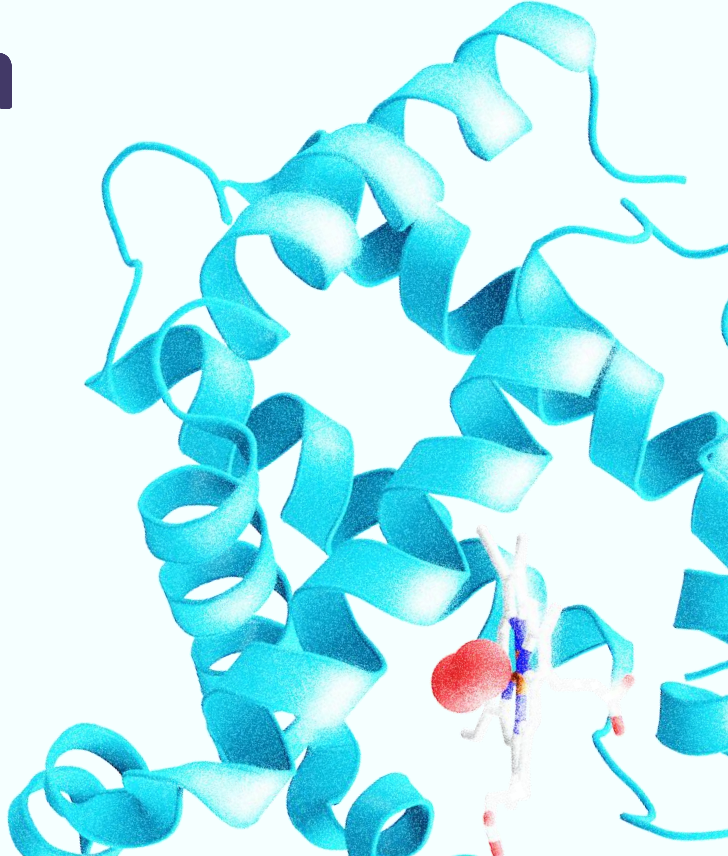
Proteins
Structure and Function

Introduction to Bioinformatics Course

Niloufar Razani
Department of Computer Engineering
Sharif University of Technology

Fall 2023



The Shape and Structure of Proteins

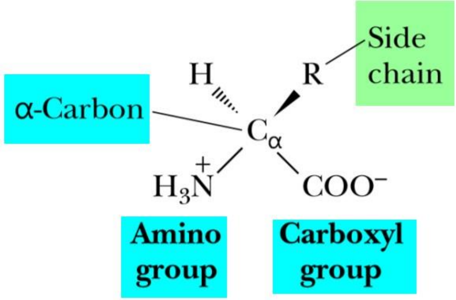
- Not only do proteins give the cell its shape and structure, they also carry out nearly all of its countless functions. For example:
- **Enzymes** promote intracellular chemical reactions by providing intricate molecular surfaces, contoured with particular bumps and crevices, that can cradle or exclude specific molecules.
- **Proteins embedded in the plasma membrane** form the channels and pumps that control the passage of nutrients and other small molecules into and out of the cell.
- **Gene Regulatory Proteins** bind to DNA to switch genes on or off.
- **Specialized proteins** also act as antibodies, toxins, hormones, antifreeze molecules, elastic fibers, or luminescence generators. And so on …

The multiplicity of functions carried out by proteins arises from the huge number of different shapes they adopt.

Amino Acids are the subunits of Proteins

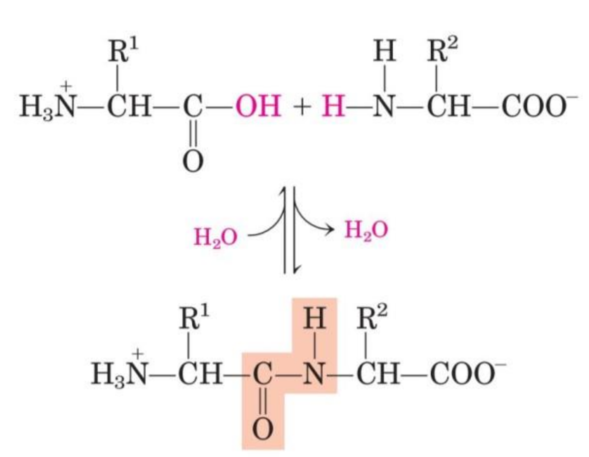
- Proteins are assembled from a set of **20 different Amino Acids**, each with different chemical properties.

- **Amino acids** are a varied class of molecules with one defining property: they all possess a **carboxylic acid group** and an **amino group**, both linked to the same carbon atom called the **α-carbon.**



- Their chemical variety comes from the **Side chain** that is also attached to the **α-carbon.**

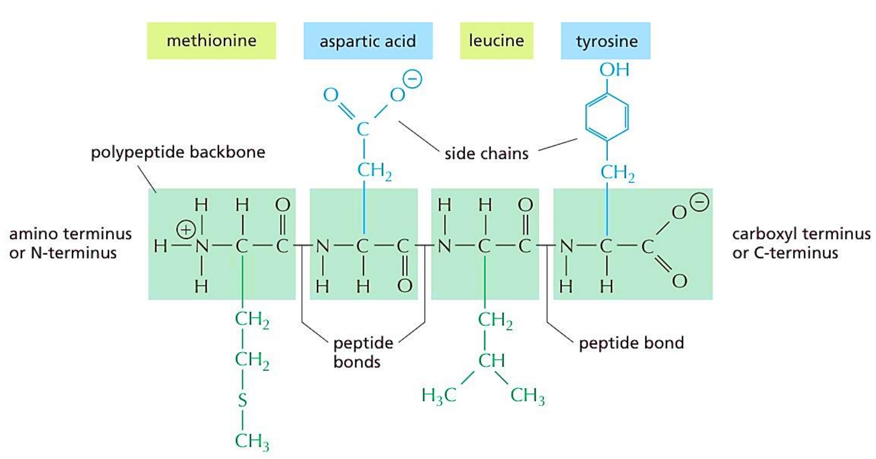
- A protein molecule is made from a long chain of these amino acids, each linked to its neighbor through a **covalent peptide bond.**



- Proteins are therefore referred to as **polypeptides or polypeptide chains.**

Amino Acids are the subunits of Proteins

- In each type of protein, the amino acids are present in a unique order, called the **amino acid sequence**, which is exactly the same from one molecule of that protein to the next.
- Each polypeptide chain consists of a **Backbone** that supports the different amino acid side chains.
- Projecting from this repetitive backbone are any of the 20 different amino acid **side chains**.
- These side chains give each amino acid its unique properties.



- Some of these side chains are **nonpolar and hydrophobic** ("water-fearing"), some are **negatively or positively charged (hydrophilic)**, some are **chemically reactive**, and so on..

Twenty different amino acids

AMINO ACID			SIDE CHAIN
Aspartic acid	Asp	D	negative
Glutamic acid	Glu	E	negative
Arginine	Arg	R	positive
Lysine	Lys	K	positive
Histidine	His	H	positive
Asparagine	Asn	N	uncharged polar
Glutamine	Gln	Q	uncharged polar
Serine	Ser	S	uncharged polar
Threonine	Thr	T	uncharged polar
Tyrosine	Tyr	Y	uncharged polar

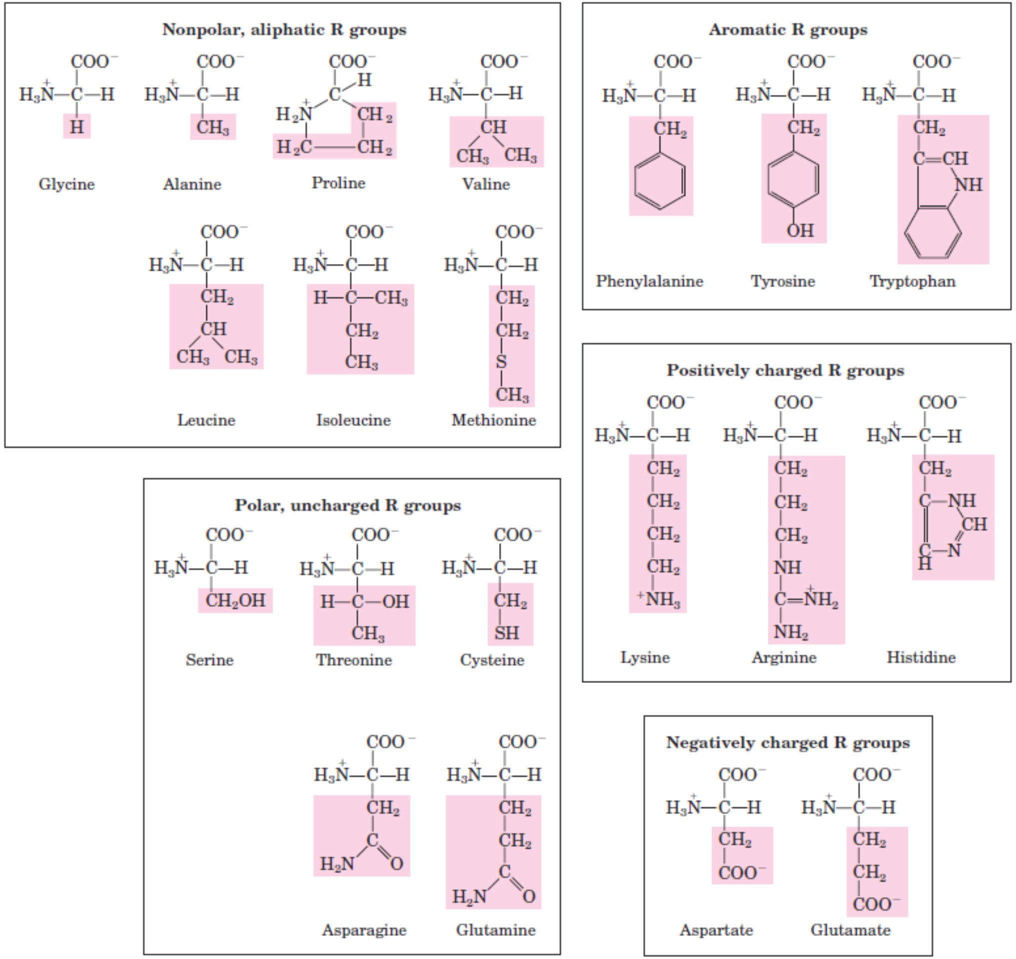
POLAR AMINO ACIDS
(hydrophilic)

AMINO ACID			SIDE CHAIN
Alanine	Ala	A	nonpolar
Glycine	Gly	G	nonpolar
Valine	Val	V	nonpolar
Leucine	Leu	L	nonpolar
Isoleucine	Ile	I	nonpolar
Proline	Pro	P	nonpolar
Phenylalanine	Phe	F	nonpolar
Methionine	Met	M	nonpolar
Tryptophan	Trp	W	nonpolar
Cysteine	Cys	C	nonpolar

NONPOLAR AMINO ACIDS
(hydrophobic)

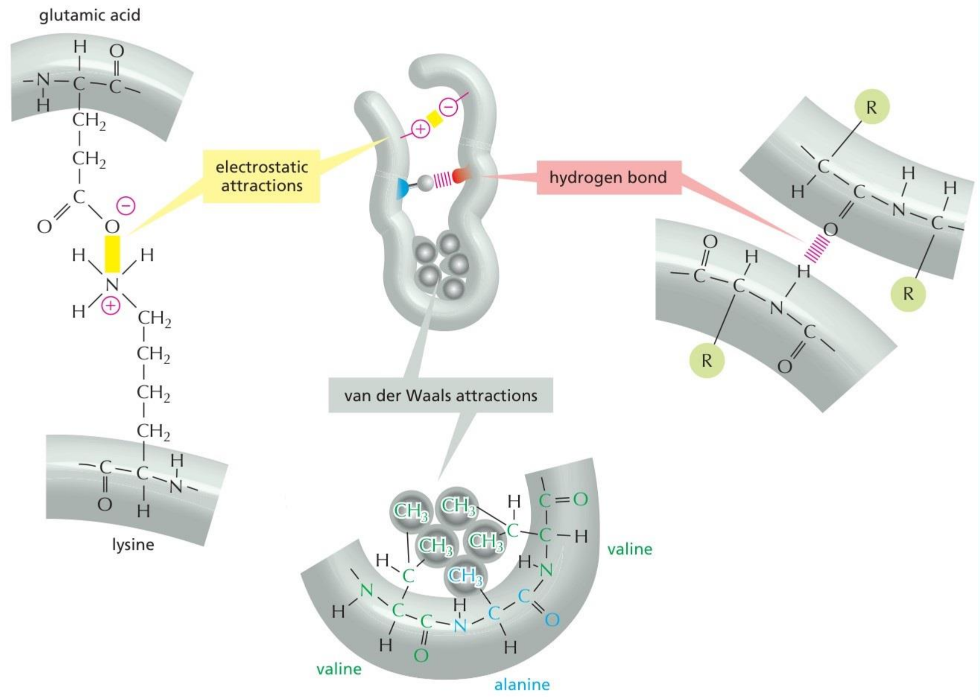
Twenty different amino acids

- **Nonpolar/Aliphatic amino acids** contain R groups which are nonpolar and hydrophobic. Aliphatic R groups can be linear or branched.
- **Nonpolar/Aromatic** R groups contain ring structures.
- **Polar/Uncharged amino acids** contain more electronegative atoms like oxygen or nitrogen.
- **Positively charged amino acids** have R groups that are protonated at physiological pH.
- **Negatively charged amino acids** contain R groups with a deprotonated carboxyl group, giving them a negative charge.



The Shape of a Protein is Specified by its amino acid Sequence

- Long polypeptide chains are very flexible and therefore proteins can in principle **fold** in an enormous number of ways.
- Each folded chain is constrained by many different sets of weak noncovalent bonds that form within proteins.
- These bonds include **hydrogen bonds, electrostatic attractions**, and **van der Waals attractions.**

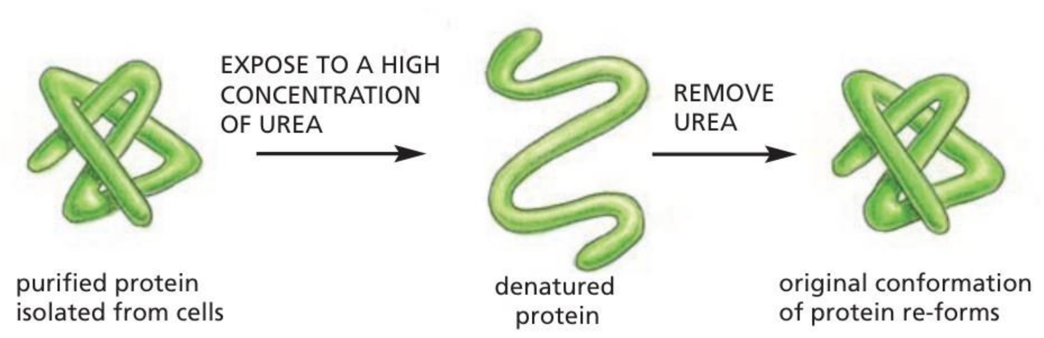


Proteins fold into a conformation of lowest energy

- Each type of protein has a particular 3D structure, which is determined by the order of the amino acids in its chain.

- A protein generally folds into the shape in which **the free energy is minimized**.

- When proteins fold incorrectly, they sometimes form aggregates that can damage cells and even whole tissues.



The fact that a renatured protein can, on its own, regain the correct conformation indicates that all the information necessary to specify the three-dimensional shape of a protein is contained in its amino acid sequence.

Proteins come in a wide Variety of complicated Shapes

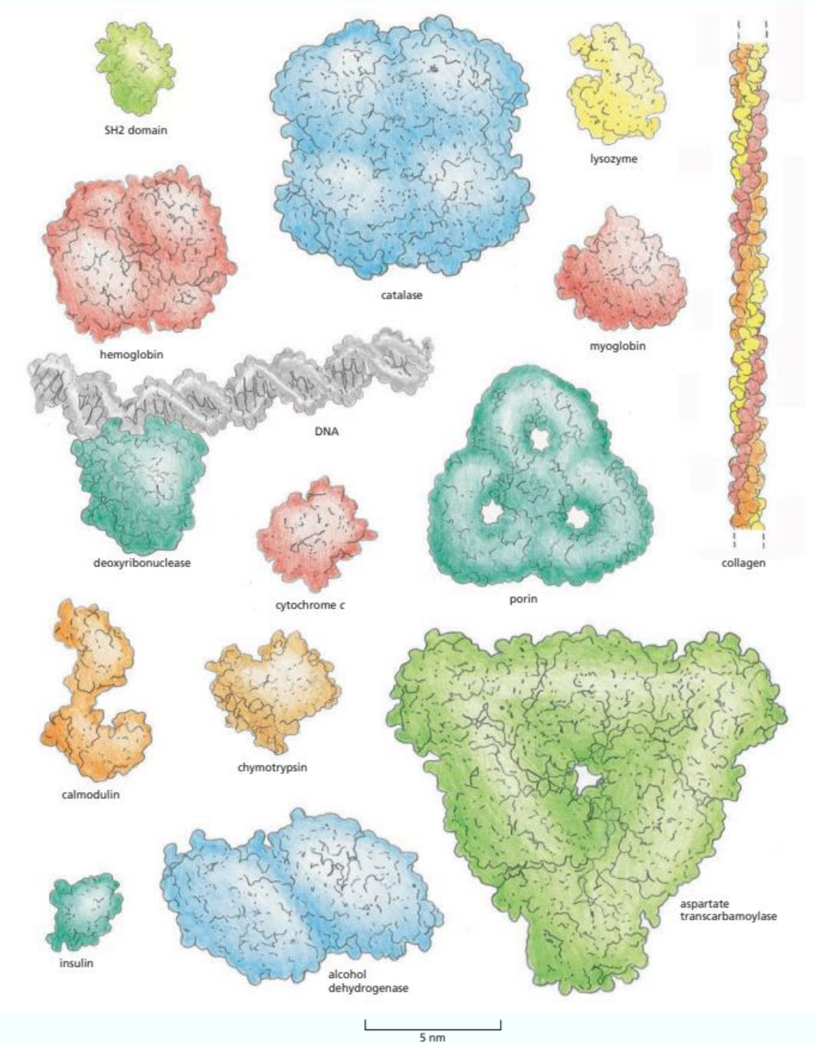
- Proteins are the most structurally diverse macromolecules in the cell.

- The vast majority of proteins are between 50 and 2000 amino acids long.

- Proteins can be **Globular**, **Membrane** or **Fibrous**; they can form filaments, sheets, rings, or spheres.

- To discover the precise folding pattern of any protein is by experiment, using either X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy.

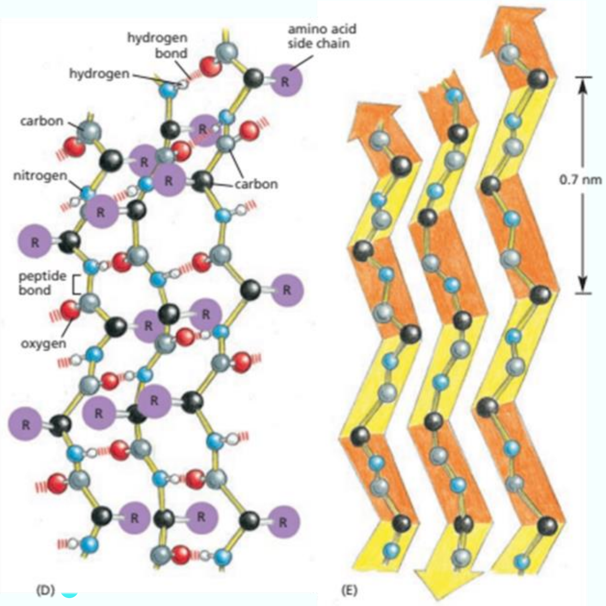
AlphaFold is an artificial intelligence program developed by DeepMind which performs predictions of protein structure.



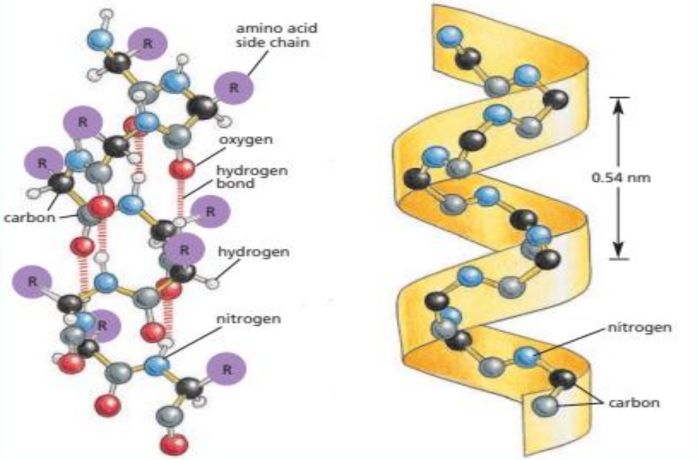
The α helix and the β Sheet are common folding patterns

- Although the overall conformation of each protein is unique, two regular folding patterns are often present:

  α helix and β Sheet



These two folding patterns are particularly common because they result from hydrogen bonds that form between the N–H and C=O groups in the **polypeptide backbone.**



- Because the amino acid side chains are not involved in forming these hydrogen bonds, **α** helices and **β** sheets can be generated by many different amino acid sequences.

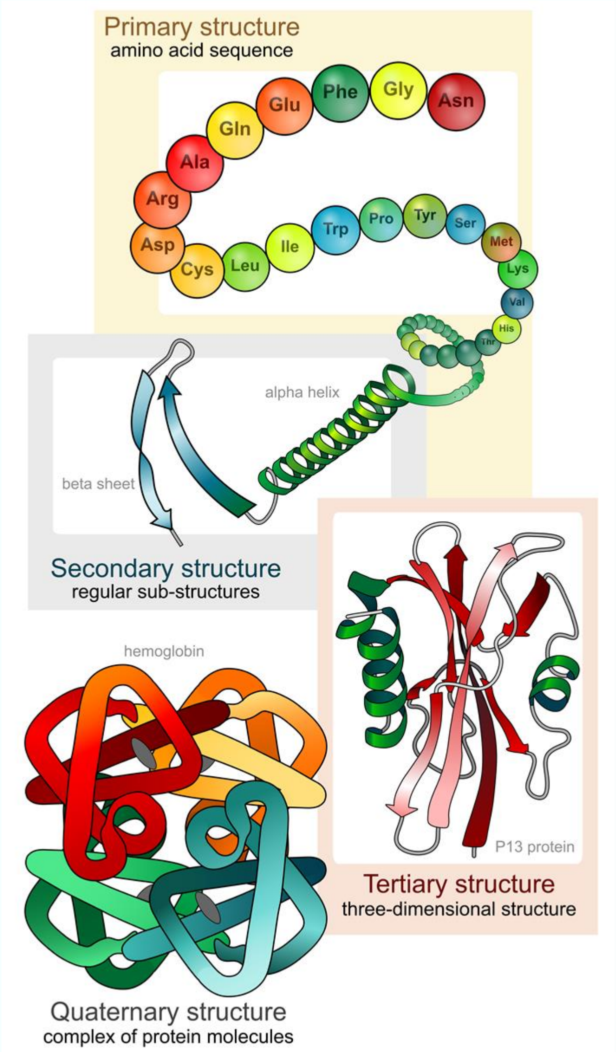
Proteins have Several levels of organization

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

- The next level of organization includes the α helices and β sheets that form within certain segments of a polypeptide chain; these folds are elements of the protein's **secondary structure.**

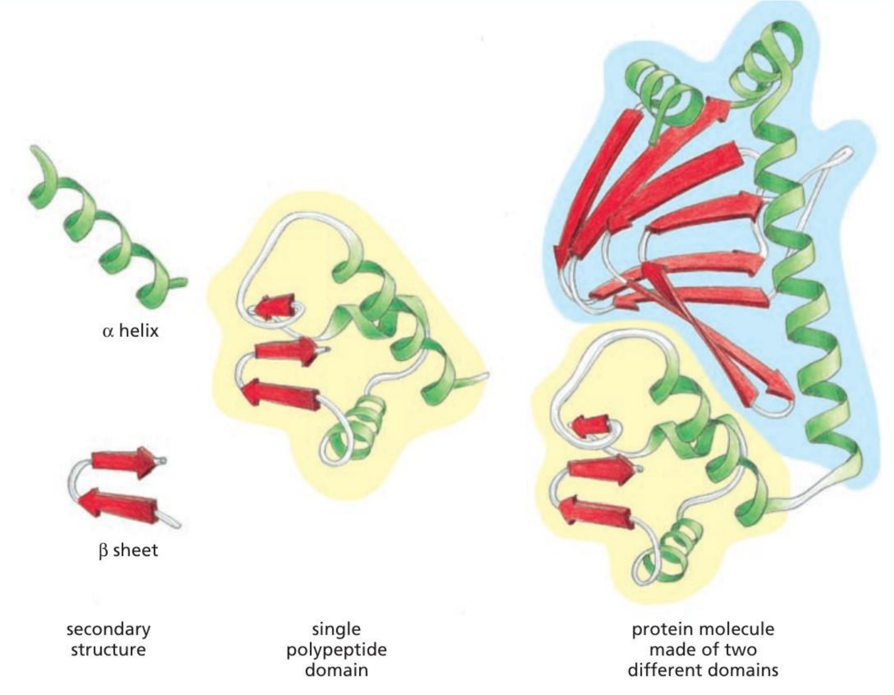
- The full, three-dimensional conformation formed by an entire polypeptide chain—including the α helices, β sheets, random coils, and any other loops and folds that form between the N- and C-termini—is referred to as the **tertiary structure.**

- Finally, if a particular protein molecule is formed as a complex of more than one polypeptide chain, then the complete structure is designated its **quaternary structure**.



Proteins have Several levels of organization

- A **Domain** is a part of a polypeptide chain that is independently stable or could undergo movements as a single entity with respect to the entire protein.
- A domain usually consists of **100–250 amino acids**, and it is the modular unit from which many larger proteins are constructed.
- The different domains of a protein are often associated with different functions.
- Small protein molecules, such as the oxygen-carrying muscle protein myoglobin, contain only a single domain



Few of the Many possible polypeptide chains will Be useful

- In theory, a vast number of different polypeptide chains could be made. Because each of the 20 amino acids is chemically distinct and each can, in principle, occur at any position in a protein chain.

- However, only a very small fraction of this unimaginably large number of polypeptide chains would fold into a stable, well-defined three-dimensional conformation.

- **why do most proteins present in cells adopt unique and stable conformations?**

- a protein with many different conformations and variable properties would not be biologically useful, for it would be like a tool that unexpectedly changes its function.

RESOURCES

- Alberts, Bruce, et al. *Essential cell biology*, Chapter2, Chemical Components of Cells.
- Alberts, Bruce, et al. *Essential cell biology*, Chapter4, Protein Structure and Function.