Bio-organic chemistry

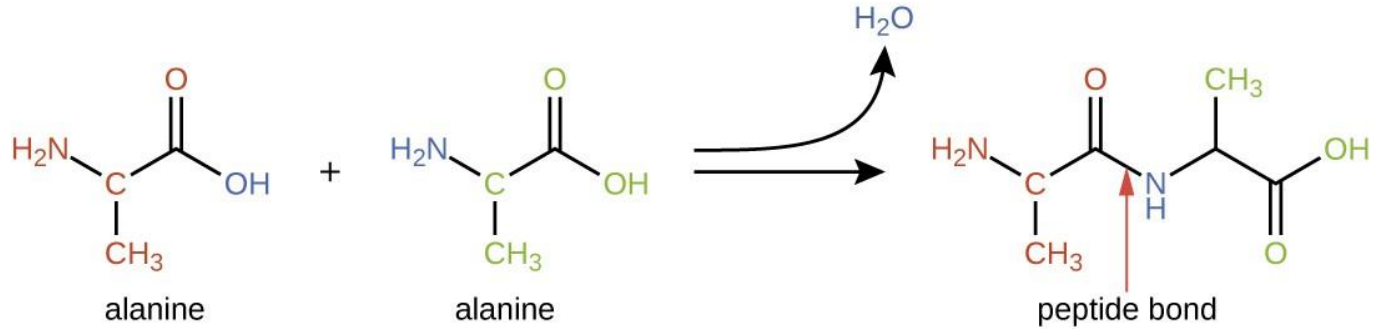
Lecture #4

Proteins and polypeptides, their structures, biological significance. Methodology for establishing their amino acid sequence.

Lecturer:
Dr. Gulnaz Seitimova
Associate Professor

Amino Acids and Peptide Bonds

Amino acids may chemically bond together by reaction of the **carboxylic acid** group of one molecule with the **amine group** of another. This reaction forms a **peptide bond** and a water molecule and is another example of **dehydration synthesis** (Figure 2). Molecules formed by chemically linking relatively modest numbers of **amino acids** (approximately 50 or fewer) are called **peptides**, and prefixes are often used to specify these numbers: **dipeptides** (two amino acids), **tripeptides** (three amino acids), and so forth. More generally, the approximate number of amino acids is designated: **oligopeptides** are formed by joining up to approximately 20 amino acids, whereas **polypeptides** are synthesized from up to approximately 50 amino acids. When the number of amino acids linked together becomes very large, or when multiple polypeptides are used as building subunits, the macromolecules that result are called **proteins**. The continuously variable length (the number of monomers) of these **biopolymers**, along with the variety of possible ***R* groups** on each amino acid, allows for a nearly unlimited diversity in the types of proteins that may be formed.



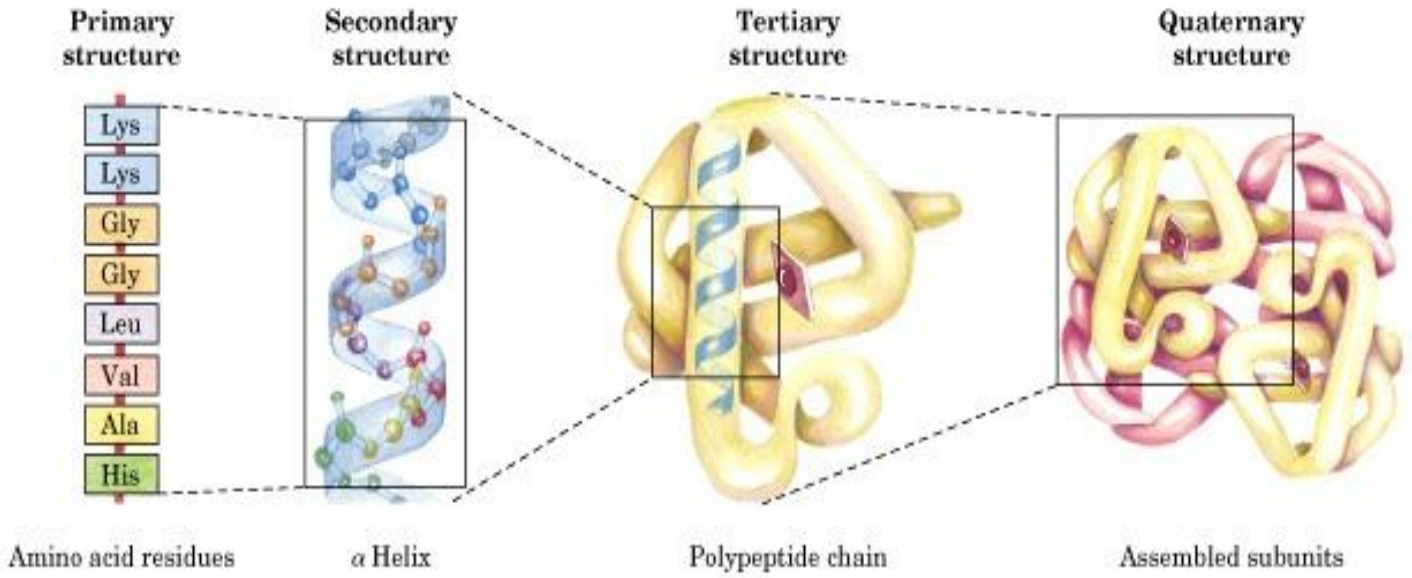
Peptide bond formation is a dehydration synthesis reaction. The carboxyl group of the first amino acid (alanine) is linked to the amino group of the incoming second amino acid (alanine). In the process, a molecule of water is released.

The four levels of organization of protein structure

[Lehninger *et al.* Principles of Biochemistry]

Proteins are an important class of biological macromolecules which are the polymers of amino acids.

Biochemists have distinguished several levels of structural organization of proteins. They are:



The primary structure of protein: a sequence of amino acids linked together by peptide bonds (covalent bond)

The secondary structure of protein: Polypeptide folding into α helix, β sheet, or random coil (H bonds involved)

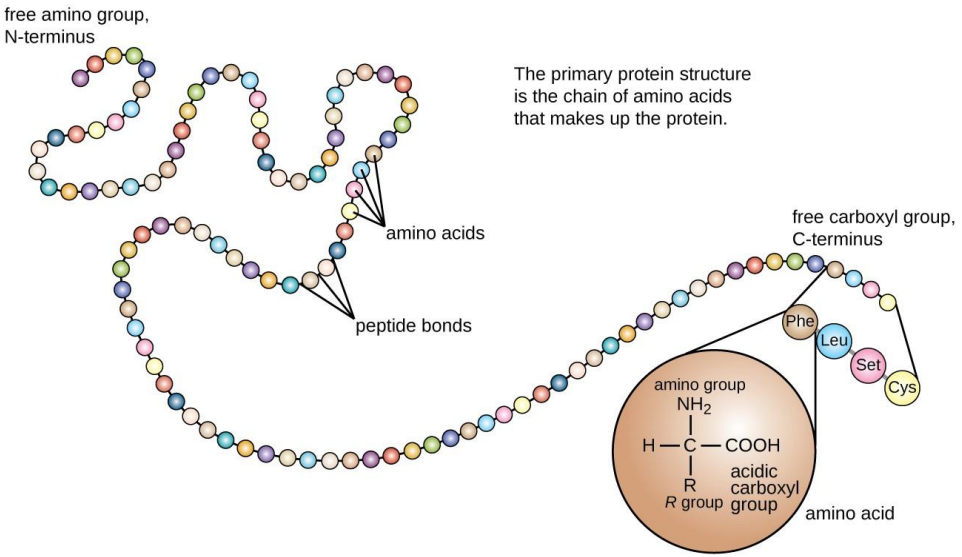
The tertiary structure of protein: 3-D folding of a single polypeptide chain (H bonds, disulfide bonds, ionic bonds, van der Waals interactions, hydrophobic interactions)

The quaternary structure of proteins: Association of two or more folded polypeptides (sub units) to form a multimeric protein (bonds and interactions similar to tertiary structure)

Protein Structure

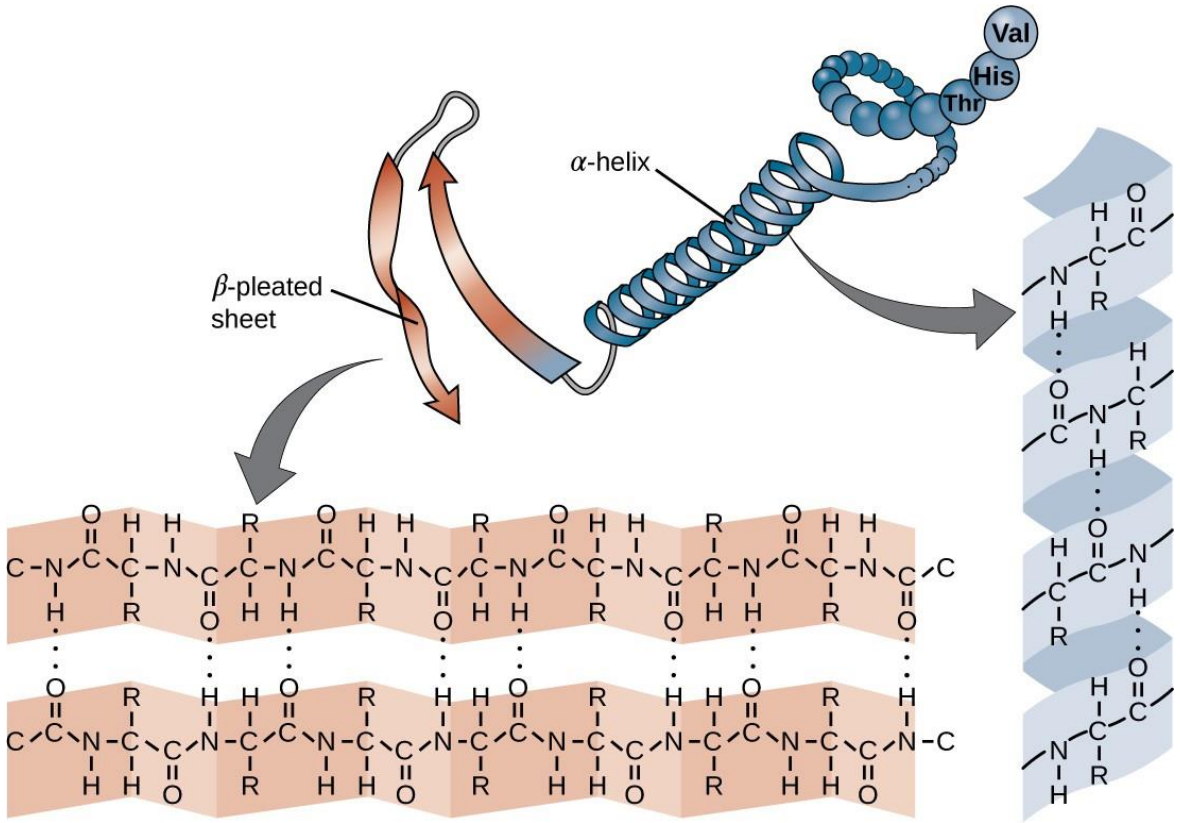
The size (length) and specific amino acid sequence of a protein are major determinants of its shape, and the shape of a protein is critical to its function. For example, in the process of biological **nitrogen fixation**, soil microorganisms collectively known as **rhizobia** symbiotically interact with roots of legume plants such as soybeans, peanuts, or beans to form a novel structure called a nodule on the plant roots. The plant then produces a carrier protein called leghemoglobin, a protein that carries nitrogen or oxygen. Leghemoglobin binds with a very high affinity to its substrate oxygen at a specific region of the protein where the shape and amino acid sequence are appropriate (the **active site**). If the shape or chemical environment of the active site is altered, even slightly, the substrate may not be able to bind as strongly, or it may not bind at all. Thus, for the protein to be fully active, it must have the appropriate shape for its function.

Protein structure is categorized in terms of four levels: primary, secondary, tertiary, and quaternary. The **primary structure** is simply the sequence of **amino acids** that make up the **polypeptide chain**.



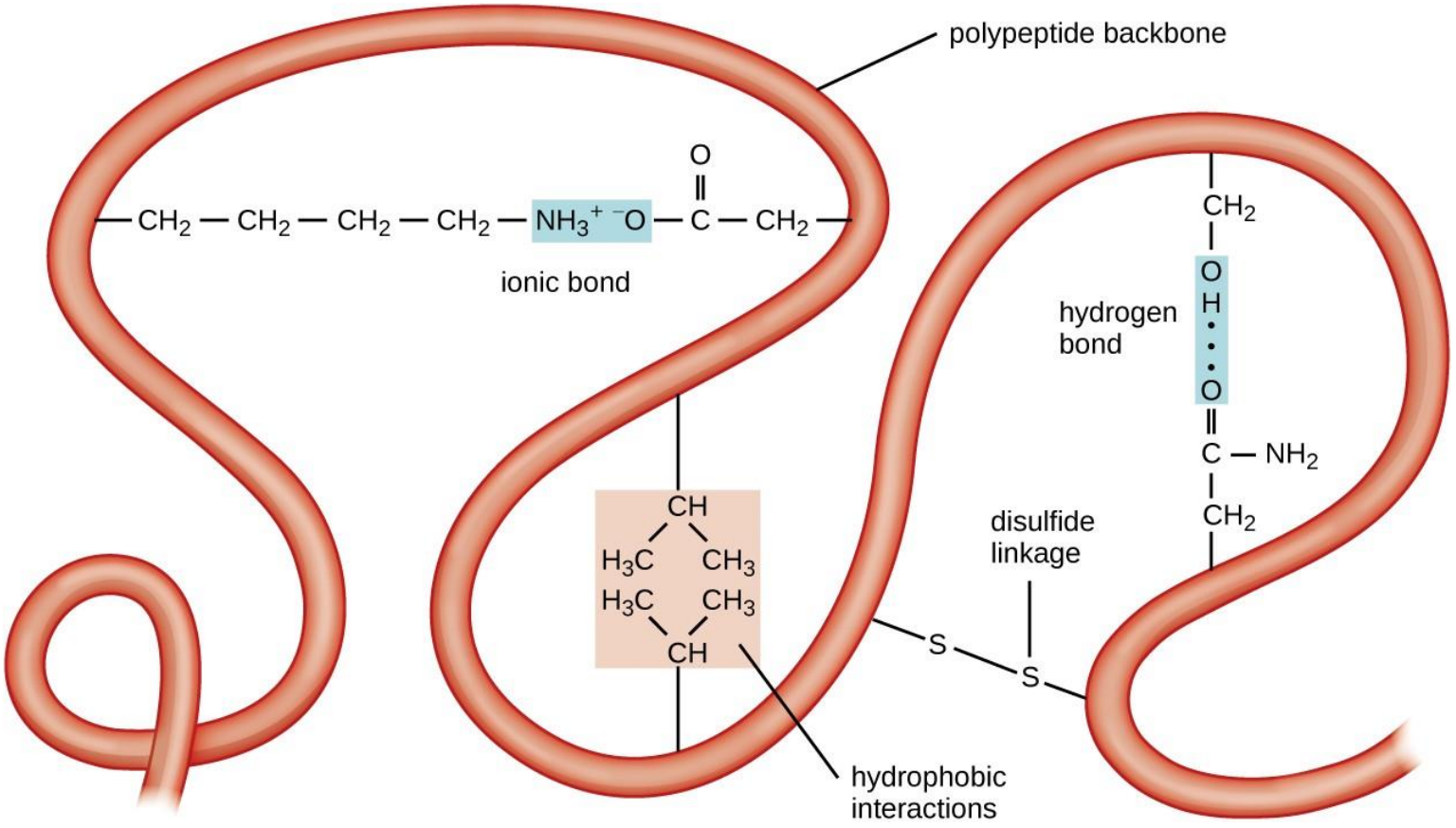
The primary structure of a protein is the sequence of amino acids.
(credit: modification of work by National Human Genome Research Institute)

The chain of amino acids that defines a protein's primary structure is not rigid but instead is flexible because of the nature of the bonds that hold the amino acids together. When the chain is sufficiently long, hydrogen bonding may occur between amine and carbonyl functional groups within the peptide backbone (excluding the *R* side group), resulting in localized folding of the polypeptide chain into helices and sheets. These shapes constitute a protein's **secondary structure**. The most common secondary structures are the α-helix and β-pleated sheet. In the **α-helix** structure, the helix is held by hydrogen bonds between the oxygen atom in a **carbonyl group** of one amino acid and the hydrogen atom of the amino group that is just four amino acid units farther along the chain. In the **β-pleated sheet**, the pleats are formed by similar **hydrogen bonds** between continuous sequences of carbonyl and **amino groups** that are further separated on the backbone of the polypeptide chain



The secondary structure of a protein may be an α-helix or a β-pleated sheet, or both.

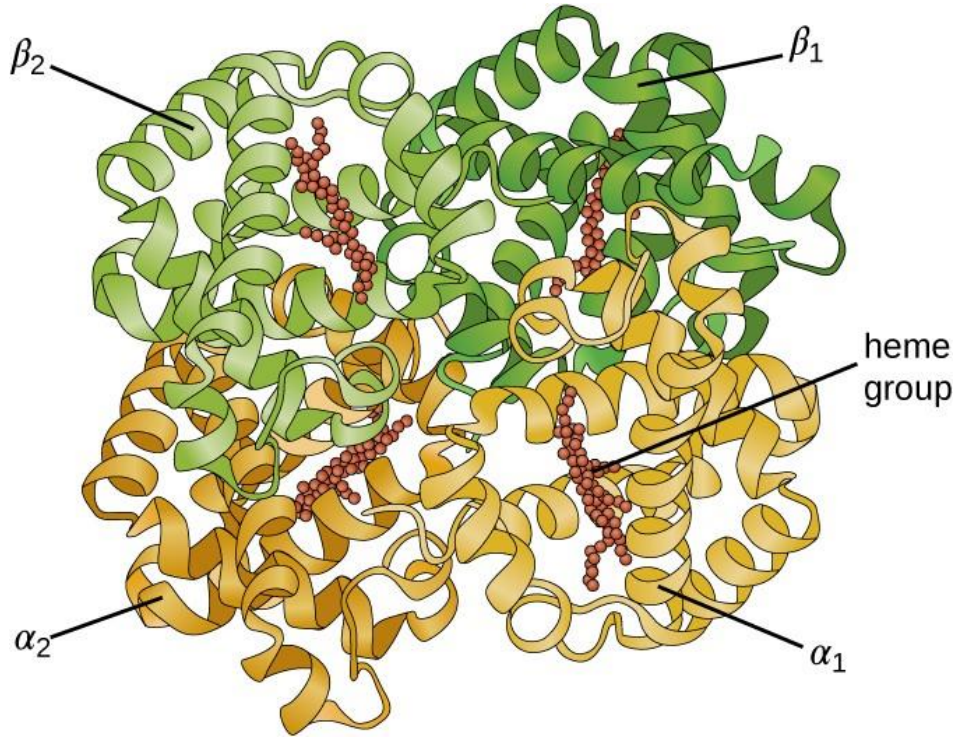
The next level of protein organization is the **tertiary structure**, which is the large-scale three-dimensional shape of a single polypeptide chain. Tertiary structure is determined by interactions between amino acid residues that are far apart in the chain. A variety of interactions give rise to protein tertiary structure, such as **disulfide bridges**, which are bonds between the sulfhydryl (–SH) functional groups on amino acid side groups; hydrogen bonds; ionic bonds; and hydrophobic interactions between nonpolar side chains. All these interactions, weak and strong, combine to determine the final three-dimensional shape of the protein and its function.



The tertiary structure of proteins is determined by a variety of attractive forces, including hydrophobic interactions, ionic bonding, hydrogen bonding, and disulfide linkages.

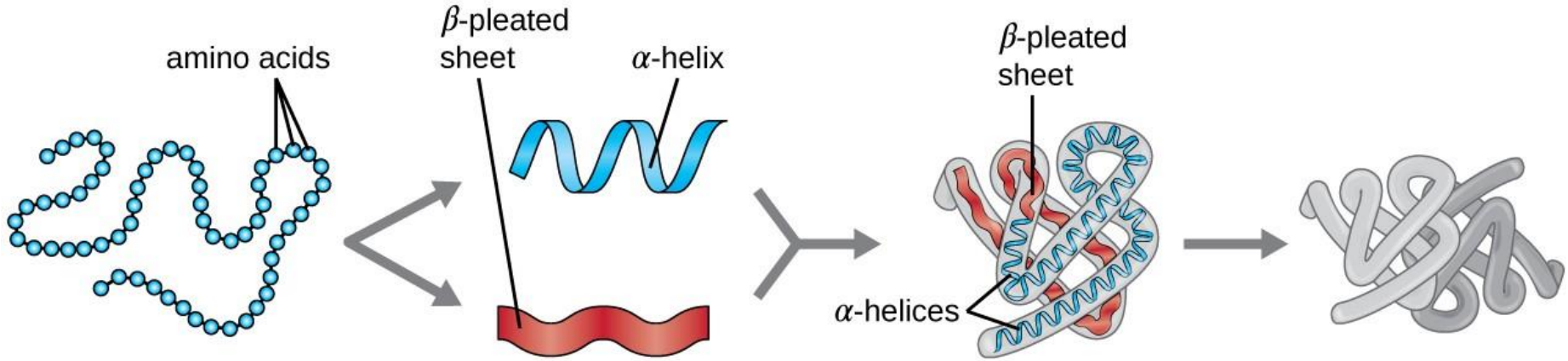
The process by which a polypeptide chain assumes a large-scale, three-dimensional shape is called **protein folding**. Folded proteins that are fully functional in their normal biological role are said to possess a **native structure**. When a protein loses its three-dimensional shape, it may no longer be functional. These **unfolded proteins** are **denatured**. Denaturation implies the loss of the **secondary structure** and **tertiary structure** (and, if present, the quaternary structure) without the loss of the primary structure.

Some proteins are assemblies of several separate **polypeptides**, also known as **protein subunits**. These proteins function adequately only when all subunits are present and appropriately configured. The interactions that hold these subunits together constitute the **quaternary structure** of the protein. The overall quaternary structure is stabilized by relatively weak interactions. Hemoglobin, for example, has a quaternary structure of four globular protein subunits: two α and two β polypeptides, each one containing an iron-based heme.



A hemoglobin molecule has two α and two β polypeptides together with four heme groups.

Another important class of proteins is the **conjugated proteins** that have a nonprotein portion. If the conjugated protein has a carbohydrate attached, it is called a **glycoprotein**. If it has a lipid attached, it is called a **lipoprotein**. These proteins are important components of membranes. Figure below summarizes the four 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

Hierarchical nature of protein structure

Primary structure (Amino acid sequence)

↓

Secondary structure （α-helix, β-sheet）

↓

Tertiary structure （Three-dimensional structure formed by assembly of secondary structures）

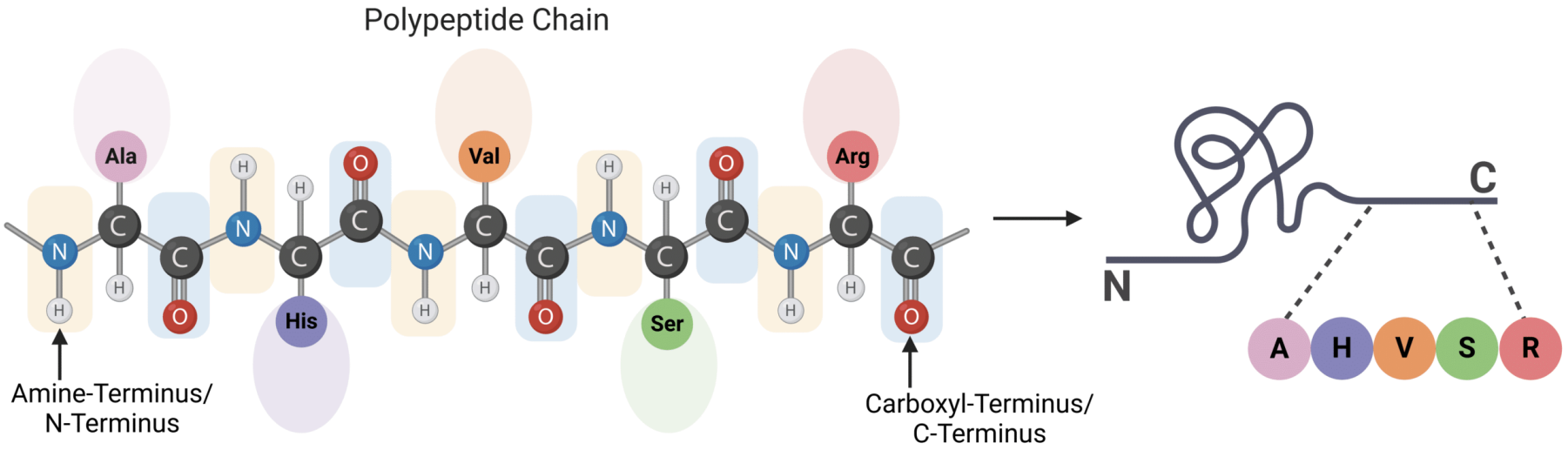
↓

Quaternary structure （Structure formed by more than one polypeptide chains）

Amino acid sequencing is the process of identifying the arrangement of amino acids in proteins and peptides. Numerous distinct amino acids have been discovered in nature but all proteins in the human body are comprised of just twenty different types. Yet these few organic molecules can attach to one another in complex three-dimensional structures of near-limitless structural varieties. This underlies the immense functional diversity of proteins—and suggests the inherent value of amino acid sequencing.

Understanding even the partial sequence of amino acids in a polypeptide chain can yield valuable insights into the identity of a protein or peptide and can help characterize its post-translational modifications.

The primary protein structure typically begins at an amine-terminus, or N-terminus; an amino acid residue with an amine group attached to the alpha-carbon. At the other end of the primary structure is the C-terminus with an unbound carboxyl group. Each of the amino acids found in nature are represented by a single or three-letter code. For instance, alanine is expressed as Ala, or A. The entire sequence of protein can subsequently be notated as a string of letters from left-to-right, corresponding to order of amino acids from terminal-to-terminal.



Amino Acid Sequencing Methods

There are two main methods of amino acid sequencing: *mass spectrometry* and *Edman degradation* with a protein sequenator.

Automated Edman amino acid sequencers are offering convenient analysis of polypeptides of up to 50 amino acids long. This process is generally characterized by seven steps:

1. Break apart disulphide bridges in the protein with a reducing agent
2. Separate the protein complex and purify the chain(s)
3. Determine the amino acid composition and terminal AAs per chain
4. Fragment each polypeptide chain
5. Recreate the AA sequence using these fragments
6. Repeat with different fragment patterns to mitigate errors

Identification via mass spectrometry is increasingly preferred as it overcomes many of the established limitations of Edman degradation. But there are various techniques within protein mass spectrometry that make amino acid sequencing via MS techniques harder to define in brief.