Introduction to DSP – Proteins and Their Structure

Your Objectives:

At the end of the lesson, you should be able to describe proteins and their structure.

Proteins are composed of chains of amino acids. A typical protein contains between 200 and 300 amino acids. Each amino acid has a specific chemical formula and structure.

Cells make distinctive proteins by arranging the different amino acids in unique sequences. For each type of protein, the cell's genetic code dictates which amino acids are added and in what order.

These carefully arranged amino acids form a chain-like structure held together by chemical bonds. Because the chemical bonds are weak, they can be easily affected by factors such as temperature and pH.

Proteins can do their job only if they keep their shape. Changes in the shape of a protein make it unable to function correctly. If the target protein (product) is exposed to the wrong temperature or pH, the protein chain will lose its shape and ability to function entirely. The loss of protein structure is called denaturation. A denatured protein cannot return to its original shape.

Protein structure is the three-dimensional arrangement of atoms in an amino acid-chain molecule. Proteins are polymers – specifically **polypeptides** – formed from sequences of amino acids, the monomers of the polymer. A single amino acid monomer may also be called a **residue**, indicating a repeating unit of a polymer. Proteins form by amino acids undergoing condensation reactions, whereby the amino acids lose one water molecule per reaction in order to attach bond to one another which has a peptide bond. By convention, a chain of under 30 amino acids is often identified as a peptide rather than as a protein. To be able to perform their biological function, proteins fold into one or more specific spatial conformations driven by a number of **non-covalent interactions,** such as hydrogen bonding, ionic interactions, Van der Waals forces and hydrophobic packing. To understand the functions of proteins at a molecular level, it is necessary to determine their three-dimensional structure. The scientific field of structural biology employs techniques such as X-ray crystallography, NMR spectroscopy, cryogenic (anstatt nur: cryo) electron microscopy (cryo-EM) and dual polarisation interferometry to determine the structure of proteins.

Protein structures range in size, from tens to several thousands of amino acids. By physical size, proteins are classified as nanoparticles ranging between 1–100 nm. Very large aggregates can be

formed from protein subunits. For example, many thousands of actin molecules assemble into a microfilament.

A protein generally undergoes reversible structural changes in performing its biological function. The alternative structures of the same protein are referred to as different conformational isomers, or simply, conformations, and transitions between them are called conformational changes.

Primary structure

The primary structure of a protein refers to the sequence of amino acids in the polypeptide chain. The primary structure is held together by peptide bonds made during the process of protein biosynthesis. The two ends of the polypeptide chain are referred to as the carboxyl terminus (C-terminus), and the amino terminus (N-terminus), based on the nature of the free group on each extremity. A counting of residues always starts at the N-terminal end (NH2-group), which is the end where the amino group is not involved in a peptide bond. The primary structure of a protein is determined by the gene corresponding to the protein. A specific sequence of nucleotides in DNA is transcribed into mRNA, which is read by the ribosome in a process called **translation.** The sequence of amino acids in insulin was discovered by Frederick Sanger, who established that proteins have defining amino acid sequences. The sequence of a protein is unique to that protein and defines the structure and function of the protein. The sequence of a protein can be determined by methods such as Edman degradation or tandem mass spectrometry. Often, however, it is read directly from the sequence of the gene, using the genetic code. It is strictly recommended to use the words "amino acid residues" when discussing proteins because, when a peptide bond is formed, a water molecule is lost, and therefore proteins are made up of amino acid residues. Post-translational modification such as phosphorylations and glycosylations are usually also considered a part of the primary structure, and cannot be read from the gene. For example, insulin is composed of 51 amino acids in two chains. One chain has 31 amino acids, and the other has 20 amino acids.

Secondary structure

Secondary structure refers to highly regular local sub-structures on the actual polypeptide backbone chain. Two main types of secondary structure, the α-helix and the β-strand or β-sheets, were suggested in 1951 by Linus Pauling et al. These secondary structures are defined by patterns of hydrogen bonds between the main-chain peptide groups. They have a regular geometry, being constrained to specific values of the dihedral angles ψ and ϕ on the Ramachandran plot. Both the α-helix and the β-sheet represent a way of saturating all the hydrogen bond donors and acceptors in the peptide backbone. Some parts of the protein are ordered but do not form any regular structures. They should not be confused with random coil, an unfolded polypeptide chain

lacking any fixed three-dimensional structure. Several sequential secondary structures may form a "supersecondary unit".

Tertiary structure

Tertiary structure refers to the three-dimensional structure of monomeric and multimeric protein molecules. The α-helixes and β-pleated-sheets are folded into a compact globular structure. The folding is driven by the non-specific hydrophobic interactions, the burial of hydrophobic residues from water, except that the structure is stable only when the parts of a protein domain lock into place by specific tertiary interactions, such as salt bridges, hydrogen bonds, and the tight packing of side chains and disulfide bonds. The disulfide bonds are extremely rare in cytosolic proteins, since the cytosol (intracellular fluid) is generally a reducing environment.

Quaternary structure

Quaternary structure is the three-dimensional structure consisting of the aggregation of two or more individual polypeptide chains (subunits) that operate as a single functional unit (multimer). The resulting multimer is stabilized by the same non-covalent interactions and disulfide bonds as with a tertiary structure. There are many possible quaternary structure organisations. Complexes of two or more polypeptides (i.e. multiple subunits) are called multimers. More specifically, it would be called a dimer when containing two subunits, a trimer with three subunits, a tetramer for four subunits, and a pentamer with five. The subunits are frequently related to one another by symmetry operations, such as a twofold axis in a dimer. Multimers made up of identical subunits are referred to with the prefix of "homo-" whereas those made up of different subunits carry the prefix of "hetero-", as with a heterotetramer: two alpha and two beta chains of hemoglobin.