Lecture 4
Discipline: Bioorganic Chemistry
Lecturer: Associate Professor, Dr. Gulnaz Seitimova

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

Objective: The aim of this lecture is to describe the structure and biological functions of proteins and polypeptides, to explain the importance of amino acid sequence in determining their properties, and to present methodologies used for determining the primary structure of proteins and peptides, including classical and modern analytical techniques.

Main Questions: Definition and general characteristics of proteins and polypeptides. Classification of proteins based on structure, solubility, and biological function. Levels of protein structure: primary, secondary, tertiary, and quaternary. Role of amino acid composition and sequence in protein function. Biological significance of proteins and polypeptides. Techniques for determining amino acid composition. Classical methods for establishing the amino acid sequence (Edman degradation). Modern mass spectrometry approaches for sequence analysis. Applications of protein and peptide sequence information in biotechnology and medicine.

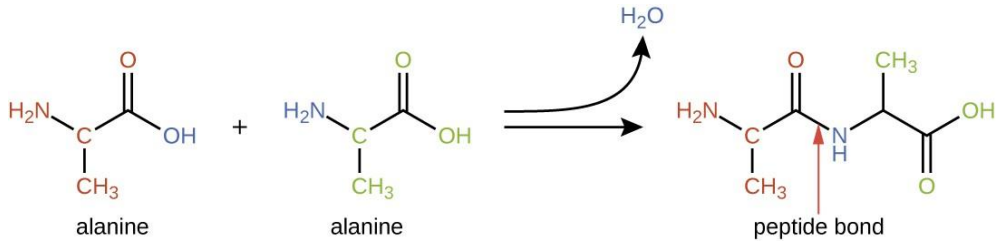
Key Notes and Theses

Proteins and Polypeptides

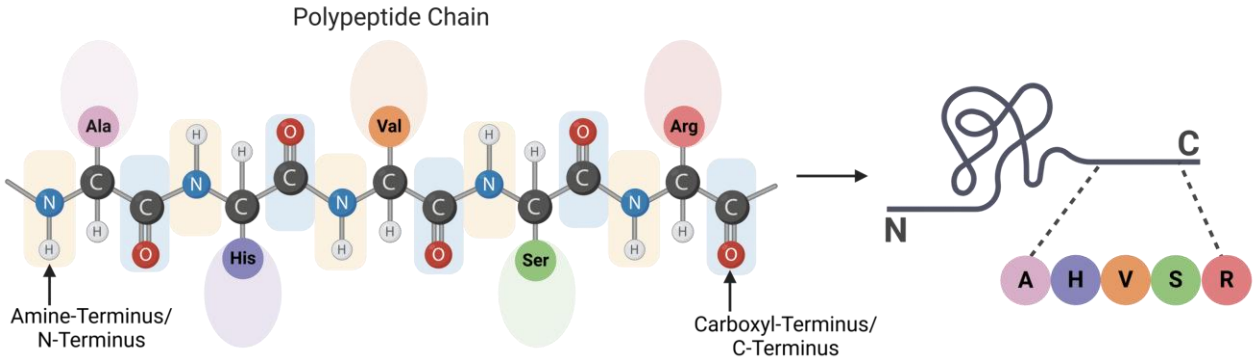
- Proteins are polypeptides with a specific sequence of amino acids that fold into a functional three-dimensional structure.
- Polypeptides are linear chains of amino acids linked via peptide bonds.
- Proteins can be simple (only amino acids) or conjugated (containing non-protein components, e.g., glycoproteins, metalloproteins).

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.



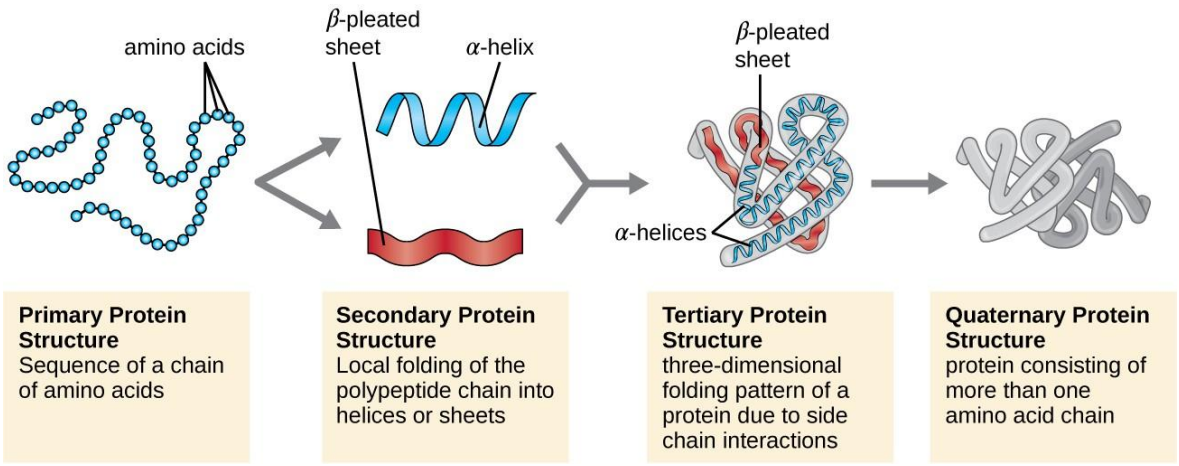
Levels of Protein Structure

1. Primary structure: linear sequence of amino acids; determines all higher-level structures.
2. Secondary structure: α-helices, β-sheets, turns, and loops stabilized by hydrogen bonds.
3. Tertiary structure: three-dimensional folding driven by hydrophobic interactions, disulfide bridges, ionic and hydrogen bonds.
4. Quaternary structure: association of multiple polypeptide chains into a functional protein complex.

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

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.

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



Biological Significance

- Structural proteins (collagen, keratin) provide mechanical support.
- Enzymes catalyze biochemical reactions.
- Transport proteins (hemoglobin, serum albumin) carry molecules.
- Signaling molecules and hormones regulate physiological processes.
- Antibodies mediate immune responses.

Amino Acid Composition Analysis

- Hydrolysis of proteins into individual amino acids using strong acids (6 M HCl, 110°C, 24 h).
- Quantitative analysis by chromatography (TLC, HPLC, ion-exchange chromatography).
- Determination of content of essential and non-essential amino acids.

Methods for Determining Amino Acid Sequence

Classical method – Edman degradation:

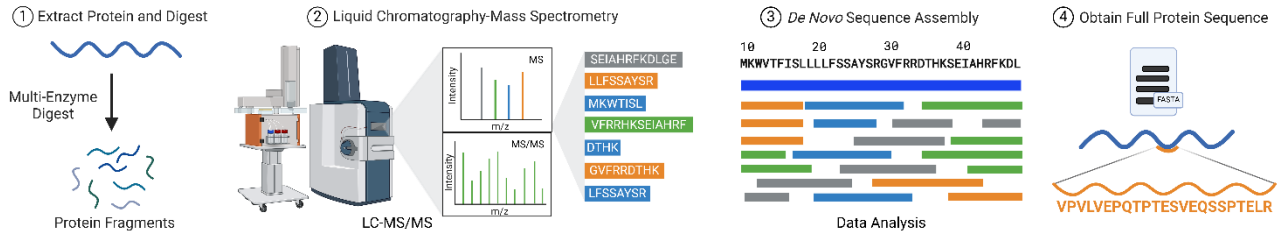
- Sequential removal of N-terminal amino acids using phenylisothiocyanate.
- Identification of each released amino acid by chromatography.
- Suitable for peptides up to ~50 amino acids.

Modern approaches – Mass spectrometry (MS):

- Peptides are ionized (MALDI, ESI) and fragmented.
- MS/MS spectra provide sequence information.
- Highly sensitive; can analyze complex mixtures and post-translational modifications.

Significance of Sequence Information

- Enables prediction of protein function and 3D structure.
- Allows engineering of proteins with enhanced properties.
- Facilitates drug design and peptide therapeutics.
- Critical for understanding genetic mutations and their biochemical consequences.



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.

Questions for Knowledge Assessment

1. What distinguishes a protein from a polypeptide?
2. Name and describe the four levels of protein structure.
3. How does amino acid sequence influence protein function?
4. What are conjugated proteins? Provide examples.
5. Explain the principle of Edman degradation for determining amino acid sequence.
6. How is mass spectrometry used to analyze protein sequences?
7. What are the biological roles of structural, enzymatic, transport, and signaling proteins?
8. Describe common methods for determining amino acid composition.
9. Why is protein sequence information important in medicine and biotechnology?

Recommended Literature

1. Nelson, D. L., Cox, M. M. (2017). *Lehninger Principles of Biochemistry* (7th ed.). New York: W.H. Freeman and Company.
2. Voet, D., Voet, J. G. (2011). *Biochemistry* (4th ed.). Hoboken, NJ: John Wiley & Sons.
3. Garrett, R. H., Grisham, C. M. (2016). *Biochemistry* (6th ed.). Boston, MA: Cengage Learning.

4. Stryer, L., Berg, J. M., Tymoczko, J. L., Gatto, G. J. (2015). *Biochemistry* (8th ed.). New York: W.H. Freeman and Company.

5. McMurry, J. (2010). *Organic Chemistry with Biological Applications* (2nd ed.). Belmont, CA: Brooks/Cole, Cengage Learning.

6. McMurry, J., Castellion, M. E. (2002). *Fundamentals of General, Organic, and Biological Chemistry* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

7. Fromm, H. J., Hargrove, M. (2012). *Essentials of Biochemistry*. Berlin, Heidelberg: Springer-Verlag.

8. Hunter, G. K. (2000). *Vital Forces: The Discovery of the Molecular Basis of Life*. San Diego, CA: Academic Press.

9. Tyukavkina, N. A., Baukov, Y. I. (2014). *Bioorganic Chemistry* (in Russian). Moscow.

10. Ovchinnikov, Y. A. (1987). *Bioorganic Chemistry* (in Russian). Moscow.

11. Rouessac, F., Rouessac, A. (2007). *Chemical Analysis: Modern Instrumentation Methods and Techniques*. Hoboken, NJ: John Wiley & Sons.

12. Jeffery, G. H., Bassett, J., Mendham, J., Denney, R. C. (1989). *Vogel's Textbook of Quantitative Chemical Analysis* (5th ed.). London: Longman; John Wiley & Sons.