

03 Amino Acids, Peptides, and Proteins

Executive Summary

Proteins are the essential molecular machines of life, mediating virtually every process within a cell. Their incredible functional diversity, from catalyzing biochemical reactions to providing structural support, originates from a remarkably simple foundation: a common alphabet of 20 standard amino acids. Each amino acid shares a core structure but is distinguished by its unique side chain, or R group. The distinct chemical properties of these R groups—their size, shape, charge, and polarity—are the fundamental determinants of a protein's character.

These amino acid building blocks are covalently linked in a specific order via peptide bonds to form long chains called polypeptides. This defined sequence of amino acids constitutes the protein's **primary structure**. The primary structure is not merely a list of components; it is the blueprint that dictates the protein's entire architecture. Based on the chemical interactions between the amino acid side chains, the linear polypeptide folds into a specific three-dimensional structure, a complex architecture comprising secondary, tertiary, and in some cases, quaternary levels of organization.

Ultimately, it is this precise three-dimensional structure that endows a protein with its specific biological function. A single alteration in the amino acid sequence can disrupt this structure and abolish function, leading to disease. Therefore, understanding the foundational principles—the properties of the 20 amino acids, the nature of the peptide bond, and the critical relationship between sequence, structure, and function—is the key to comprehending virtually all biological processes at a molecular level.

1. The Fundamental Building Blocks: Amino Acids

The vast diversity of protein function begins with a simple but powerful concept: proteins are polymers constructed from a common set of 20 amino acids. These molecules can be regarded as the "alphabet" in which the language of protein structure is written. A thorough understanding of the unique chemical properties of each amino acid is the first and most critical step in comprehending the intricate structure and function of the proteins they form.

1.1. Common Structural Features

All 20 common amino acids are **α -amino acids**, sharing a universal architecture. As shown in Figure 3-2, they possess a central carbon atom, the **α -carbon**, bonded to four distinct chemical groups:

1. A basic **amino group** ($-\text{NH}_3^+$)
2. An acidic **carboxyl group** ($-\text{COO}^-$)
3. A **hydrogen atom** ($-\text{H}$)
4. A variable **side chain**, or **R group**, which is unique to each amino acid.

Because the α -carbon is bonded to four different groups, it serves as a **chiral center**. The only exception is **glycine**, where the R group is a second hydrogen atom, making it achiral. This

tetrahedral arrangement around the chiral α -carbon allows amino acids to exist as stereoisomers.

1.2. Stereochemistry: The L-Isomer Preference

The two possible stereoisomers of an amino acid are **enantiomers**—nonsuperposable mirror images of each other, designated as L and D forms (e.g., L-Alanine and D-Alanine in Figure 3-3). A remarkable feature of life is that the amino acid residues found in proteins are almost exclusively **L-stereoisomers**. This uniformity is not accidental; it is a functional imperative. The formation of stable, repeating structures like the α -helix, which are essential for protein function, requires that the constituent amino acids all belong to the same stereochemical series. Cells specifically synthesize L-isomers because the active sites of their enzymes are themselves asymmetric, causing the reactions they catalyze to be stereospecific.

1.3. Classification by R-Group Properties

The 20 common amino acids are categorized into five main classes based on the properties of their R groups, particularly their polarity and charge at a physiological pH (around 7.0). These properties dictate how an amino acid residue behaves within a protein.

Class Name	Core Property	Significance in Proteins	Examples
Nonpolar, aliphatic	Hydrophobic and nonpolar alkyl side chains.	Tend to cluster in the protein's interior via the hydrophobic effect, stabilizing the protein's three-dimensional structure.	Glycine, Alanine, Proline, Valine, Leucine, Isoleucine, Methionine
Aromatic	Aromatic rings in the side chain; relatively nonpolar.	Also contribute to the hydrophobic effect. Tryptophan and tyrosine can absorb UV light at 280 nm, a property used to quantify proteins.	Phenylalanine, Tyrosine, Tryptophan
Polar, uncharged	R groups contain functional groups that form H-bonds.	Hydrophilic and often located on the protein surface. Cysteine can form covalent disulfide bonds, which are critical for stabilizing structure.	Serine, Threonine, Cysteine, Asparagine, Glutamine
Positively charged	Side chains carry a net positive charge at pH 7.0.	Hydrophilic and key for ionic interactions. Histidine, with a pKa near neutral pH, often acts as a proton donor/acceptor in enzyme catalysis.	Lysine, Arginine, Histidine
Negatively charged	Side chains carry a net negative charge at pH 7.0.	Hydrophilic and involved in ionic interactions (salt bridges) and metal binding. Often found in the active sites of enzymes.	Aspartate, Glutamate

1.4. Acid-Base Properties and the Zwitterion

At neutral pH, the carboxyl group of an amino acid is deprotonated ($-\text{COO}^-$) and the amino group is protonated ($-\text{NH}_3^+$). This creates a dipolar ion known as a **zwitterion**, which has both

a positive and a negative charge but a net charge of zero. This zwitterionic form can act as either an acid (proton donor) or a base (proton acceptor). The acid-base behavior of an amino acid can be visualized with a titration curve, such as that for glycine (Figure 3-10).

- **pKa:** The pKa is the pH at which an ionizable group is half-protonated and half-deprotonated. Glycine has two pKa values: $pK_1 = 2.34$ for the α -carboxyl group and $pK_2 = 9.60$ for the α -amino group.
- **Buffering Regions:** An amino acid acts as a buffer in the flat regions of its titration curve, which are centered around its pKa values. Glycine is an effective buffer near pH 2.34 and pH 9.60.
- **Isoelectric Point (pI):** The pI is the specific pH at which an amino acid has a net charge of zero and will not migrate in an electric field. For a simple amino acid with a non-ionizable R group, it is the arithmetic mean of its two pKa values: $pI = (pK_1 + pK_2)/2$.

With this understanding of individual amino acids, we can now explore how they are linked together to build functional proteins.

2. From Chains to Machines: Peptides and Proteins

Individual amino acids serve as monomers that are covalently linked to form polymers called polypeptides. The specific covalent linkage, the **peptide bond**, creates the protein's **primary structure**—the linear sequence of amino acid residues. This sequence is the fundamental foundation upon which a protein's complex three-dimensional architecture and biological function are built.

2.1. The Peptide Bond: Forming Polypeptide Chains

A peptide bond is an amide linkage formed between the α -carboxyl group of one amino acid and the α -amino group of another. This reaction, depicted in Figure 3-13, is a **condensation** (or dehydration) reaction because it involves the removal of a water molecule.

- An **oligopeptide** is a short chain formed by joining a few amino acid residues.
- A **polypeptide** is a longer chain, typically with a molecular weight below 10,000.
- The term **protein** generally refers to one or more large polypeptides with a defined biological function.

2.2. Polypeptide Directionality and Composition

By convention, polypeptide chains are read from the **amino-terminal (N-terminal)** end, which has a free α -amino group, to the **carboxyl-terminal (C-terminal)** end, which has a free α -carboxyl group (Figure 3-14). Each protein has a unique amino acid composition and, more importantly, a unique sequence. This primary structure is precisely defined and is absolutely critical for the protein's ability to fold correctly and carry out its specific function.

2.3. Diversity in Protein Structure and Composition

The universe of proteins is characterized by immense diversity in size, composition, and structure, allowing them to perform a vast array of biological functions.

- **Size Variation:** Proteins range dramatically in size. Some hormones, like oxytocin, are small peptides of only nine residues, while titin, a muscle protein, is one of the largest known, with nearly 27,000 amino acid residues.
- **Multisubunit Proteins:** Many proteins are **multisubunit**, consisting of two or more noncovalently associated polypeptide chains. If a protein has at least two identical subunits, it is said to be **oligomeric**, and the identical units are referred to as **protomers**. Hemoglobin, for example, is a classic oligomeric protein with four subunits (two identical α chains and two identical β chains) that can be viewed as a dimer of identical $\alpha\beta$ protomers.
- **Conjugated Proteins:** Some proteins, called **conjugated proteins**, contain permanently associated non-amino acid chemical components known as **prosthetic groups**. These groups are often essential for function.

Conjugated Protein Class	Prosthetic Group	Example
Lipoproteins	Lipids	β_1 -Lipoprotein of blood
Glycoproteins	Carbohydrates	Immunoglobulin G
Phosphoproteins	Phosphate groups	Glycogen phosphorylase
Hemoproteins	Heme (iron porphyrin)	Hemoglobin
Flavoproteins	Flavin nucleotides	Succinate dehydrogenase
Metalloproteins	Metal ion (e.g., Fe, Zn, Cu)	Ferritin

This incredible diversity necessitates a robust set of methods for separating and studying individual proteins from the complex mixtures found in cells.

3. Isolating and Characterizing Proteins

To understand a protein's structure and function, it must first be isolated from the thousands of other proteins and molecules inside a cell. This process, known as protein purification, is a strategic endeavor that employs a series of techniques to exploit the unique physical and chemical properties of the target protein, such as its size, charge, and binding affinity.

3.1. Core Purification Strategies: Column Chromatography

Column chromatography is a powerful and versatile method for separating proteins. In this technique, a protein mixture is passed through a column containing a solid, porous matrix (the stationary phase). As a buffered solution (the mobile phase) flows through, proteins are separated based on their differential interactions with the matrix.

- **Ion-Exchange Chromatography:** This technique separates proteins based on their net electric charge at a specific pH. The column matrix is charged, either negatively (cation exchanger) or positively (anion exchanger). Proteins with an opposite charge

bind to the matrix and are retarded, while proteins with the same charge pass through more quickly.

- **Size-Exclusion Chromatography:** Also known as gel filtration, this method separates proteins by size and shape. The matrix contains pores of a specific size. Larger proteins cannot enter the pores and thus travel a more direct path, eluting from the column first. Smaller proteins enter the pores, taking a longer, more labyrinthine path, and elute later.
- **Affinity Chromatography:** This is a highly specific method that separates proteins based on their unique binding affinity for a molecule, or **ligand**, that is covalently attached to the column matrix. Only proteins that bind the specific ligand are retained on the column, while all others wash through. The bound protein can then be eluted by adding a high concentration of free ligand or by changing the salt concentration.

3.2. Analysis of Purity and Molecular Weight: Electrophoresis

Electrophoresis is a powerful analytical tool, not a purification step, used to assess the purity of a protein sample and estimate its molecular weight. In this technique, proteins migrate through a porous gel in an electric field.

- **SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis):** This is the most common electrophoretic method. The detergent SDS is used to denature proteins and bind to them, which gives them a large, uniform negative charge-to-mass ratio. This process effectively masks the protein's intrinsic charge, causing separation to occur almost exclusively based on mass. Smaller proteins migrate more rapidly through the gel.
- **Isoelectric Focusing:** This technique separates proteins based on their isoelectric point (pI). A stable pH gradient is established in the gel, and proteins migrate until they reach the pH that matches their pI , at which point their net charge is zero and they stop moving.
- **Two-Dimensional Electrophoresis:** For high-resolution separation of complex protein mixtures, this technique combines two methods. Proteins are first separated by isoelectric focusing in one dimension, and then that gel is subjected to SDS-PAGE in a second dimension. This separates proteins by both pI and molecular weight.

These analytical tools provide a clear picture of a protein's purity and its constituent subunits, bridging the gap between a purified sample and the determination of its amino acid sequence, or primary structure.

4. The Blueprint of Function: Primary Structure

The structure of a protein can be described at four distinct levels of complexity: primary, secondary, tertiary, and quaternary (Figure 3-23). The foundational level is the **primary structure**, which is the linear sequence of amino acid residues in a polypeptide chain. This sequence is not merely a list of components; it is the fundamental blueprint that contains all the information needed to specify the protein's higher-level structures and, ultimately, its unique biological function.

4.1. The Critical Link Between Sequence and Function

The amino acid sequence of a protein is inextricably linked to its biological role. The evidence for this fundamental principle is compelling and multifaceted:

- **Functional Specificity:** Proteins that perform different functions invariably have different amino acid sequences.
- **Genetic Diseases:** Many genetic disorders, such as sickle cell anemia, result from a single amino acid change in a protein, which dramatically impairs its function.
- **Evolutionary Conservation:** When comparing proteins with similar functions across different species (e.g., human vs. bovine cytochrome c), their sequences often show significant similarities, especially in regions critical for function.

4.2. Unlocking Evolutionary History

Because protein sequences are a direct expression of genetic information, they are an invaluable resource for tracing evolutionary relationships between species. By comparing sequences, we can identify protein families and understand their ancestral origins.

- **Homologs:** These are proteins that share a common ancestor, identified by significant sequence similarity.
- **Orthologs vs. Paralogs:** Homologs can be further classified. **Orthologs** are homologous proteins found in different species (e.g., human hemoglobin and chimpanzee hemoglobin). **Paralogs** are homologous proteins found within a single species, arising from gene duplication (e.g., the different globin chains within humans).
- **Conserved Sequences:** These are segments of a protein's sequence that have changed very little over long periods of evolutionary time. Such conservation implies that these regions are critically important for the protein's structure or function.
- **Signature Sequences:** In some cases, a particular sequence motif is unique to a specific taxonomic group, serving as a "signature" that can help define evolutionary lineages.

From the simplest building blocks to the complex machinery of life, the journey of a protein begins with its linear sequence. This string of 20 common amino acids is the fundamental code that underlies the vast complexity and diversity of biological function.
