# "Structure of Proteins"

## Overview

- The 20 amino acids commonly found in proteins are joined together by peptide bonds.

- The linear sequence of the linked amino acids contains the information necessary to generate a protein molecule with a unique three-dimensional shape.

- The complexity of protein structure is best analyzed by considering the molecule in terms of four organizational levels:

    - Primary
    - Secondary
    - Tertiary
    - Quaternary

- An examination of these hierarchies of increasing complexity has revealed that certain structural elements are repeated in a wide variety of proteins.

- This suggests that there are general "rules" regarding the ways in which proteins achieve their native, functional form.

- These repeated structural elements range from:
    - Simple combinations of α-helices and β-sheets forming small motifs
    - To the complex folding of polypeptide domains of multifunctional proteins

## Primary Structure of Proteins

- The sequence of amino acids in a protein is called the primary structure of the protein.
- Understanding the primary structure of proteins is important because:
    - Many genetic diseases result in proteins with abnormal amino acid sequences
    - These abnormal sequences cause improper folding and loss or impairment of normal function
- If the primary structures of the normal and the mutated proteins are known:
    - This information may be used to diagnose or study the disease

## A. Peptide Bond

- In proteins, amino acids are joined covalently by peptide bonds
- Peptide bonds are:
    - Amide linkages between the α-carboxyl group of one amino acid and the α-amino group of another
- Example:
    - Valine and alanine can form the dipeptide valylalanine through the formation of a peptide bond
- Peptide bonds are not broken by conditions that denature proteins, such as:
    - Heating
    - High concentrations of urea
- Prolonged exposure to a strong acid or base at elevated temperatures is required to hydrolyze these bonds nonenzymically

# Naming and Characteristics of Peptides

## 1. Naming the Peptide

- By convention:
    - The free amino end (N-terminal) of the peptide chain is written to the left
    - The free carboxyl end (C-terminal) is written to the right
- Therefore, all amino acid sequences are read from the N- to the C-terminal end of the peptide.
- Linkage of many amino acids through peptide bonds results in an unbranched chain called a polypeptide
- Each component amino acid in a polypeptide is called a "residue":
    - This is because it is the portion of the amino acid remaining after the atoms of water are lost in the formation of the peptide bond.

- When a polypeptide is named:
    - All amino acid residues have their suffixes (-ine, -an, -ic, or -ate) changed to -yl
    - Exception: The C-terminal amino acid retains its original suffix
- Example: A tripeptide composed of:
    - N-terminal valine
    - Glycine
    - C-terminal leucine
    - Is named valyl glycyl leucine

2. Characteristics of the Peptide Bond

- The peptide bond has a partial double-bond character
    - It is shorter than a single bond
    - It is rigid and planar
- This prevents free rotation around the bond between the carbonyl carbon and the nitrogen of the peptide bond

- However, bonds between the:
    - α-carbons and α-amino groups
    - α-carbons and α-carboxyl groups
    - Can be freely rotated
        - Although rotation is limited by the size and character of the R-groups
- This flexibility allows the polypeptide chain to assume a variety of possible configurations
- The peptide bond is generally a trans bond (instead of cis)
    - This is due to steric interference of the R-groups when in the cis position

3. Polarity of the Peptide Bond

- Like all amide linkages:
    - The -C=O and -NH groups of the peptide bond are uncharged
    - They neither accept nor release protons over the pH range of 2-12

- Therefore, the charged groups present in polypeptides consist solely of:
    - The N-terminal (α-amino) group
    - The C-terminal (α-carboxyl) group
    - Any ionized groups present in the side chains of the constituent amino acids
- The -C=O and -NH groups of the peptide bond are polar
    - They are involved in hydrogen bonds, for example, in:
        - α-helices
        - β-sheet structures

B. Determination of the Amino Acid Composition of a Polypeptide

Step 1: Hydrolysis of the Polypeptide

- The first step in determining the primary structure of a polypeptide is to:
    - Identify and quantitate its constituent amino acids

- A purified sample of the polypeptide is:
    - Hydrolyzed by strong acid at 110°C for 24 hours
    - This treatment cleaves the peptide bonds and releases the individual amino acids

Step 2: Separation by Cation-Exchange Chromatography

- The released amino acids can be separated by cation-exchange chromatography
- In this technique:
    - A mixture of amino acids is applied to a column
    - The column contains a resin to which a negatively charged group is tightly attached
    - [Note: If the attached group is positively charged, the column becomes an anion-exchange column]
- The amino acids bind to the column with different affinities, depending on their:
    - Charges
    - Hydrophobicity
    - Other characteristics

- Each amino acid is sequentially released from the chromatography column by eluting with:
    - Solutions of increasing ionic strength and pH

Step 3: Quantitation by Ninhydrin Reaction

- The separated amino acids contained in the eluate from the column are quantitated by:
    - Heating them with ninhydrin
- Ninhydrin is a reagent that:
    - Forms a purple compound with most:
        - Amino acids
        - Ammonia
        - Amines

Quantitation of Amino Acids in a Polypeptide

- Quantitation Process:
    - After separation, amino acids are reacted with ninhydrin, which reacts with:

        - Primary amino groups ($-NH_2$),
        - Ammonia,
        - Amines.

    - Forms a purple-colored complex, except:
        - Proline and hydroxyproline give a yellow color due to secondary amine group.

- Spectrophotometric Measurement:

    - The intensity of color produced by ninhydrin-amino acid reaction is measured spectrophotometrically.
    - The absorbance at a specific wavelength (usually 570 nm) indicates the quantity of the amino acid.

- Amino Acid Analyzer:

    - An automated instrument that:
        - Performs cation-exchange chromatography,
        - Applies ninhydrin derivatization,
        - Uses spectrophotometric detection to quantify each amino acid.

## C. Peptide Sequencing from N-Terminal End (Edman Degradation)

- Sequencing Definition:
    - Determining the order of amino acids in a polypeptide from the N-terminal.
- Edman Reagent (Phenylisothiocyanate):
    - Used under mildly alkaline conditions.
    - Reacts with the free α-amino group of the N-terminal amino acid.
    - Forms a phenylthiocarbamoyl derivative, which is cleaved to yield:
        - A stable phenylthiohydantoin (PTH)-amino acid.
- Key Features:
    - The PTH-amino acid can be identified by chromatography.
    - Remaining peptide (shortened by one residue) can be cycled repeatedly for sequencing.
    - Useful for sequencing peptides up to ~50 residues.

## D. Fragmentation of Polypeptides for Sequencing

- Problem with Large Polypeptides:
    - Polypeptides >100 amino acids are too large to sequence directly.
- Solution: Fragmentation:
    - Polypeptide is cleaved at specific sites to generate smaller fragments.
    - Each fragment is individually sequenced.
    - Overlapping fragments (from different cleavage methods) help reconstruct full sequence.
- Cleavage Agents:
    - Enzymes (Proteases/Peptidases):
        - Exopeptidases: Cut terminal residues.
            - Aminopeptidases: Remove from N-terminal.
            - Carboxypeptidases: Remove from C-terminal.
        - Endopeptidases: Cleave within the chain (e.g., trypsin, chymotrypsin).
    - Chemical Cleavers: e.g., cyanogen bromide (cleaves at methionine).

## E. Primary Structure Determination by DNA Sequencing

- Genetic Basis of Protein Sequence:
    - The DNA coding sequence in a gene determines the amino acid sequence of the polypeptide.
- DNA Sequencing Advantage:
    - Direct sequencing of the DNA allows in silico translation using the genetic code.
    - Especially useful for large proteins or when peptide sequencing is difficult.
- Note:
    - Requires knowledge of reading frame and post-translational modifications for accurate interpretation.

## Limitations of Indirect Amino Acid Sequencing

- Common Practice:
    - DNA sequencing is routinely used to predict a protein's amino acid sequence.

- Limitations:
    - Disulfide bonds:
        - Cannot identify positions of disulfide bridges (-S-S-) in the folded polypeptide.
    - Post-translational modifications (PTMs):
        - Cannot detect modified amino acids (e.g., hydroxylation, phosphorylation, methylation).
        - PTMs occur after translation, affecting structure and function.

- Importance of Direct Protein Sequencing:
    - Provides the true primary sequence, including:
        - Disulfide linkage positions.
        - Identification of chemically altered residues.

Secondary Structure of Proteins

- Definition:
    - Regular, repetitive 3D arrangements of nearby amino acids in the polypeptide chain.
    - Governed by hydrogen bonding between backbone atoms (not R-groups).

- Examples of Secondary Structures:
    - α-helix
    - β-pleated sheet
    - β-bend (β-turn)
    - Collagen α-chain helix (special case—see separate notes)

## A. α-Helix Structure

- General Features:
    - Right-handed spiral with a tightly coiled backbone core.
    - R-groups project outward from the helix axis to minimize steric clashes.
    - Found in both fibrous proteins (e.g., keratin) and globular proteins (e.g., myoglobin).

- Example Proteins:
    - Keratins: Almost entirely α-helical; rigid due to many disulfide bonds.
    - Myoglobin: Highly α-helical; globular and flexible in structure..

1. Hydrogen Bonds in the α-Helix

- Intrachain Hydrogen Bonding:
    - Between carbonyl oxygen (C=O) of one amino acid and the amide hydrogen (N–H) of an amino acid 4 residues ahead.
    - Direction: Parallel to the axis of the helix.
    - Stabilizes helix by forming a continuous hydrogen-bond network.
- Result:
    - All peptide bond components (except the first and last few) are linked by intrachain H-bonds.
    - Individual H-bonds are weak, but collectively they stabilize the α-helix.

2. Amino Acids per Turn

- Turn Characteristics:
    - 3.6 amino acid residues per complete helical turn.
    - Amino acids 3–4 residues apart in primary structure are close together spatially in the helix.

## 3. Amino Acids That Disrupt α-Helix

- Proline:
    - Has a secondary amino group (imino), forming a rigid ring structure.
    - Cannot form normal hydrogen bonds → inserts a kink, breaking the helix.
- Charged Amino Acids (when in high concentration):
    - Glutamate, Aspartate (acidic),
    - Histidine, Lysine, Arginine (basic).
    - May repel each other or form ionic bonds, destabilizing the helix.
- Bulky or β-branched Amino Acids:
    - Tryptophan (bulky indole ring) → steric hindrance.
    - Valine, Isoleucine (β-branched at R-group) → disrupt tight packing.
    - Prevent smooth helical folding if present in clusters.

B. β-Sheet (β-Pleated Sheet)

- Definition:
    - A type of secondary structure formed by hydrogen bonding between peptide bond components of extended polypeptide segments.
    - Surfaces appear pleated, hence the name β-pleated sheet.
- Representation:
    - In diagrams, β-strands are often shown as broad arrows, indicating the direction from N-terminal to C-terminal.

I. Comparison of β-Sheet vs. α-Helix

- α-Helix:
    - Composed of a single polypeptide forming a right-handed coil.
    - Hydrogen bonds are parallel to the helix axis.

- β-Sheet:
    - Composed of 2 or more β-strands (either from different chains or the same chain folded back).
    - Polypeptide chains are fully extended, not coiled.
    - Hydrogen bonds are perpendicular to the backbone axis.
    - Hydrogen bonding occurs between adjacent strands.

2. Types of β-Sheets

- Antiparallel β-Sheet:
    - Adjacent strands run in opposite directions (N→C and C←N).
    - Hydrogen bonds are more linear → stronger and more stable.

- Parallel β-Sheet:
    - Strands run in the same direction (all N→C).
    - Hydrogen bonds are angled, hence slightly weaker.

- Interchain vs. Intrachain:
    - Interchain β-sheets:
        - Formed by different polypeptide chains.
        - Stabilized by interchain hydrogen bonds.

- Intrachain β-sheets:
  - Formed when a single chain folds back on itself.
  - Stabilized by intrachain hydrogen bonds.

- Structural Note in Globular Proteins:
  - β-sheets are often twisted with a right-handed curl when viewed along the backbone.
  - This twist helps form the hydrophobic core of globular proteins.

- Functional Note:
  - Both α-helices and β-sheets allow maximum hydrogen bonding within a polypeptide's interior.

C. β-Bends (Reverse Turns or β-Turns)

- Function:
  - Induce a reversal in the direction of the polypeptide chain.
  - Essential for forming compact, globular shapes.
  - Often found on the surface of proteins, helping to orient chains inward or outward.

- Association with β-Sheets:
    - Often connect adjacent antiparallel β-strands, facilitating the zig-zag structure.
- Composition:
    - Typically 4 amino acids long.
    - Common amino acids:
        - Proline: Inserts a kink due to its rigid ring structure.
        - Glycine: Frequently present due to its small size (H as R-group), providing flexibility.
- Stabilization:
    - Maintained by:
        - Hydrogen bonds (usually between the carbonyl oxygen of residue 1 and the amide hydrogen of residue 4).
        - Sometimes ionic interactions between charged side chains.

## D. Nonrepetitive Secondary Structure

- Definition:
    - The portion of a polypeptide chain not organized into α-helices or β-sheets.
    - Typically adopts loop or coil conformations.
- Proportion in Globular Proteins:
    - About 50% of a typical globular protein is nonrepetitive.
    - The rest is made of α-helices and β-sheets (repetitive structures).
- Characteristics:
    - Though nonrepetitive, these structures are not random.
    - They exhibit specific conformations essential for function.
    - Distinct from "random coils", which refer to denatured/disordered proteins (loss of 3D structure).

## E. Supersecondary Structures (Motifs)

- Definition:
    - Specific combinations of α-helices, β-sheets, and loops that form compact folding patterns.
    - Represent intermediate levels between secondary and tertiary structures.
- Location in Proteins:
    - Found mostly in the core (interior) of globular proteins.
    - Loop regions (like β-bends) connect them and are typically located on the protein surface.
- Formation Mechanism:
    - Formed by close packing of side chains from adjacent structural elements.
    - Often, adjacent α-helices and β-sheets in the sequence are also spatially adjacent in the 3D structure.

- Examples of Common Motifs:
    - Helix-loop-helix motif:
        - Common in DNA-binding proteins.
        - Frequently found in transcription factors.
        - Provides structural support for DNA interaction.

Tertiary Structure of Globular Proteins

- Definition:
    - The 3D folding of a single polypeptide chain into its final functional shape.
    - Involves both the folding of individual domains and their arrangement.

- Determinant:
    - Dictated entirely by the primary structure (amino acid sequence).

- Domains:
    - Basic structural and functional units of a protein.
    - Each domain may perform a specific function (e.g., binding site, catalytic site).

- Structural Features in Aqueous Solution:
    - Protein is compact with high atomic packing density.
    - Hydrophobic side chains:
        - Buried in the interior.
        - Shielded from aqueous environment.
    - Hydrophilic groups:
        - Located on the surface.
        - Interact with water and other polar molecules.

A. Domains

- Definition:
    - Domains are functional and structural units within a polypeptide.
    - Each domain behaves like an independent compact globular protein.

- Occurrence:
    - Polypeptides with >200 amino acids usually contain 2 or more domains.

- Structure:
    - Core of each domain is made from supersecondary structures (motifs).
    - Domains fold independently of each other within the same polypeptide chain.
- Functional Independence:
    - Each domain often has a distinct function (e.g., ligand binding, enzymatic activity, etc.).
    - Structurally self-contained, maintaining integrity even when isolated.

B. Interactions Stabilizing Tertiary Structure

- Overview:
    - Tertiary structure is dictated by the primary amino acid sequence.
    - Folding is guided by side chain interactions, forming a compact 3D structure.
    - Four main types of interactions stabilize the structure:

## 1) Disulfide Bonds

- Definition:
    - Covalent bond between the –SH groups of two cysteine residues.
    - Produces a cystine residue.
- Location:
    - Cysteines may be far apart in the primary sequence or even on different polypeptides.
    - Folding brings them into proximity to form the bond.
- Function:
    - Stabilizes the tertiary structure.
    - Protects proteins from denaturation, especially in extracellular environments.
    - Common in secreted proteins like immunoglobulins.

## 2) Hydrophobic Interactions

- Mechanism:
    - Nonpolar side chains cluster in the interior of the protein.
    - Avoid contact with the aqueous environment.
    - Interact with other hydrophobic residues.
- Contrast:
    - Polar/charged residues are exposed to the surface, interacting with water.
- In Membranes:
    - This arrangement is reversed in membrane proteins (nonpolar groups outside).
- Result:
    - Energetically favorable R-group segregation ensures stability.

3) Hydrogen Bonds

- Involvement:
    - Side chains with –OH or –NH groups (e.g., serine, threonine) participate.
    - Form bonds with electron-rich atoms like carbonyl oxygen or carboxyl groups.
- Effect:
    - Strengthens the folded structure.
    - Bonds between surface polar groups and water enhance protein solubility.

4) Ionic Interactions (Salt Bridges)

- Participants:
    - Negatively charged side chains:
        - Aspartate ($-COO^-$), Glutamate ($-COO^-$)
    - Positively charged side chains:
        - Lysine ($-NH_3^+$), Arginine ($-NH_3^+$)

- Result:
    - Formation of ionic bonds contributes to structural stability.

C. Protein Folding

- Definition:
    - Folding is driven by interactions between side chains of amino acids.
    - Occurs rapidly within seconds to minutes inside the cell.

- Folding Mechanism:
    - Side chain attractions and repulsions direct the folding path.
        - Opposite charges (e.g., lysine and glutamate) attract.
        - Like charges repel.
    - Involves:
        - Hydrogen bonding
        - Hydrophobic interactions
        - Disulfide bond formation

- Outcome:
    - Folding follows a trial-and-error pathway, favoring configurations where attractions outweigh repulsions.
    - Results in a stable, low-energy conformation.

D. Denaturation of Proteins

- Definition:
    - Denaturation is the loss of secondary and tertiary structure without breaking peptide bonds.
    - Leads to unfolding and disorganization of the protein structure.

- Denaturing Agents:
    - Heat
    - Organic solvents
    - Mechanical mixing
    - Strong acids or bases
    - Detergents
    - Heavy metal ions (e.g., lead, mercury)

- Reversibility:
    - Some proteins can refold to their original structure when the agent is removed (reversible).
    - However, most proteins become permanently denatured and disordered.
- Result:
    - Denatured proteins often become insoluble and precipitate from solution.

E. Role of Chaperones in Protein Folding

- Folding Information:
    - Instructions for folding are encoded in the primary structure.
    - But proteins often don't refold correctly after denaturation.
- Folding Timing:
    - Folding starts during translation (not after full synthesis), which helps reduce misfolding.

- Chaperones (Heat Shock Proteins):
    - Specialized proteins that assist folding.
    - Functions:
        - Prevent premature folding until full synthesis.
        - Accelerate folding (act as catalysts).
        - Shield vulnerable regions to prevent incorrect interactions.

V. Quaternary Structure of Proteins

- Definition:
    - Arrangement of two or more polypeptide subunits in a functional protein.

- Types of Proteins:
    - Monomeric proteins → single polypeptide chain.
    - Multimeric proteins → two or more subunits (can be identical or different).

- Interactions Holding Subunits:
    - Noncovalent bonds, including:
        - Hydrogen bonds
        - Ionic bonds
        - Hydrophobic interactions

Quaternary Subunit Function and Isoforms

- Subunit Functionality:
    - Subunits in multimeric proteins may:
        - Function independently, or
        - Cooperate with each other.
            - Example: In hemoglobin, oxygen binding to one subunit increases the affinity of the other subunits for oxygen (positive cooperativity).

- Isoforms:
    - Proteins that perform the same function but have different primary structures.
    - May arise from:
        - Different genes, or
        - Tissue-specific processing of the product of a single gene.

- Isozymes:
    - If isoforms are enzymes, they are called isozymes.

## VI. Protein Misfolding

- Nature of Folding:
    - Folding is a complex trial-and-error process.
    - Misfolded proteins can occur due to errors.
- Cellular Response:
    - Most misfolded proteins are:
        - Tagged for degradation.
        - Broken down by the cell's quality control mechanisms.
    - However, with age, these systems become less efficient.
- Result:
    - Intracellular/extracellular aggregates of misfolded proteins may accumulate.
    - These deposits are linked to degenerative diseases.

## A. Amyloid Disease

- Causes of Misfolding:
    - Spontaneous misfolding
    - Gene mutations that alter protein structure.
    - Abnormal proteolytic cleavage of normal proteins, leading to misfolding.
- Amyloid Characteristics:
    - Misfolded proteins form long fibrillar aggregates with β-pleated sheet structure.
    - These insoluble aggregates are called amyloids.
- Diseases Involved:
    - Amyloids are implicated in several degenerative disorders, especially:
        - Alzheimer disease (AD)
- Amyloid-β (Aβ) in Alzheimer Disease:
    - Aβ peptide contains 40–42 amino acids.
    - Derived from amyloid precursor protein (APP)—a single transmembrane protein.
    - APP is cleaved proteolytically to form Aβ.

- Pathogenic Properties of Aβ:
    - Aggregates into non-branching fibrils with β-pleated sheets.
    - Neurotoxic to brain cells.
    - Central to cognitive decline in AD.
- Aβ Deposition Sites:
    - Found in brain parenchyma and around blood vessels.
- Genetic vs Sporadic Forms:
    - Most AD cases are sporadic.
    - About 5-10% of cases are familial (genetically inherited).

Neurofibrillary Tangles in Alzheimer Disease

- Secondary Pathological Feature:
    - Accumulation of neurofibrillary tangles inside neurons.

- Key Protein:
    - Composed of abnormal tau (τ) protein.
    - Normal tau: Helps with microtubule assembly.
    - Defective tau: Disrupts the function of normal tau, contributing to neuronal dysfunction.

B. Prion Disease

Overview

- Prion protein (PrP) is implicated in transmissible spongiform encephalopathies (TSEs).
- Examples of TSEs:
    - Creutzfeldt-Jakob disease (humans)
    - Scrapie (sheep)
    - Bovine spongiform encephalopathy (BSE) or "mad cow disease" (cattle)

Discovery of the Infectious Agent

- Scientists found that the infectious agent in scrapie was a single protein, without any detectable nucleic acid.

- This infectious protein is designated PrP^Sc:
    - Sc = scrapie
    - Highly resistant to proteolytic degradation
    - Forms insoluble fibrillar aggregates, similar to amyloids

The Normal Form: PrP^C

- PrP^C = Cellular prion protein
- Found normally in mammalian brains
    - Located on the surface of neurons and glial cells
- Encoded by the same gene as PrP^Sc
- No difference in:
    - Primary structure
    - Post-translational modifications

## Pathogenic Mechanism

- The key difference lies in the 3D conformation:
    - Normal PrP^C: Rich in α-helices
    - Infectious PrP^Sc: Increased β-sheet content
- Conformational change makes PrP^Sc:
    - Protease-resistant
    - Capable of forming aggregates
- PrP^Sc acts as a template:
    - Induces the conversion of normal PrP^C → pathogenic PrP^Sc
    - Leads to progressive accumulation of infectious protein

## Clinical Impact

- TSEs are invariably fatal
- Currently, no treatment exists to reverse or halt the disease