Structural Bioinformatics

Hafsath.C.A

Department Of Computer Science, CUSAT

July 13, 2022

Overview

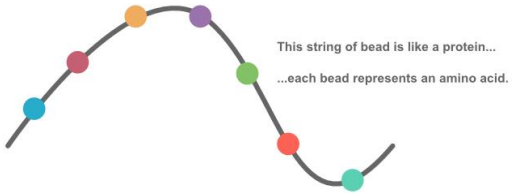
Protein Structure Basics

- Proteins perform most essential biological and chemical functions in a cell.
- Play important roles in structural, enzymatic, transport, and regulatory functions.
- The protein functions are strictly determined by their structures.

Protein Structure Basics

- The building blocks of proteins are twenty naturally occurring amino acids(residue)
- Proteins are made up of hundreds of smaller units called amino acids that are attached to one another by peptide bonds, forming a long chain.
- Protein as a string of beads where each bead is an amino acid.

Figure 1:



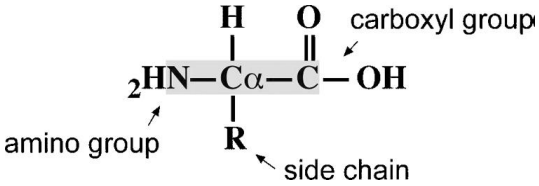
Hierarchy

- Protein structures can be organized into four levels of hierarchies with increasing complexity
  - ⋆ **Primary structure**: Amino acid sequence of its polypeptide chain
  - ⋆ **Secondary structure**: Spatial arrngement of the polypeptide backbone
  - ⋆ **Tertiary structure**: Three dimensional structure of entire polypeptide
  - ⋆ **Quaternary structure**: Three dimensional structure of proteins that are composed of two or more polypeptide chains

α – *Amino acids*

- They are small molecules that contain a free amino group (NH2) and a free carboxyl group (COOH). Both of these groups are linked to a central carbon (C), which is attached to a hydrogen and a side chain group (R).
- The chemical reactivities of the R groups determine the specific properties of the amino acids.

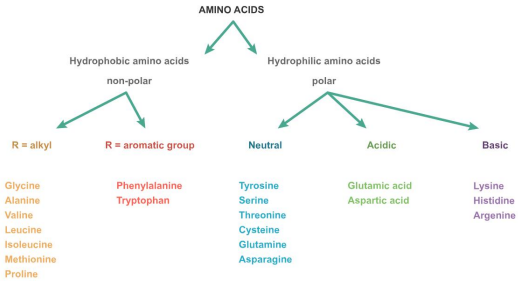
Figure 2: General Structure of Amino acid



Amino acid Classification

Amino acids can be grouped into several categories based on the chemical and physical properties of the side chains, such as size and affinity for water

Figure 3: Amino acid Classaification



List Of Amino acids

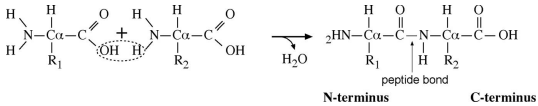
Figure 4: Twenty Standard Amino Acids Grouped by Their Common Side-Chain Features

Amino Acid Group	Amino Acid Name	Three- and One-Letter Code	Main Functional Features
Small and nonpolar	Glycine Alanine Proline	Gly, G Ala, A Pro, P	Nonreactive in chemical reactions; Pro and Gly disrupt regular secondary structures
Small and polar	Cysteine Serine Threonine	Cys, C Ser, S Thr, T	Serving as posttranslational modification sites and participating in active sites of enzymes or binding metal
Large and polar	Glutamine Asparagine	Gln, Q Asn, N	Participating in hydrogen bonding or in enzyme active sites
Large and polar (basic)	Arginine Lysine Histidine	Arg, R Lys, K His, H	Found in the surface of globular proteins providing salt bridges; His participates in enzyme catalysis or metal binding
Large and polar (acidic)	Glutamate Aspartate	Glu, E Asp, D	Found in the surface of globular proteins providing salt bridges
Large and nonpolar (aliphatic)	Isoleucine Leucine Methionine Valine	Ile, I Leu, L Met, M Val, V	Nonreactive in chemical reactions; participating in hydrophobic interactions
Large and nonpolar (aromatic)	Phenylalanine Tyrosine Tryptophan	Phe, F Tyr, Y Trp, W	Providing sites for aromatic packing interactions; Tyr and Trp are weakly polar and can serve as sites for phosphorylation and hydrogen bonding

Peptide Formation

- The peptide formation involes two amino acids covalently joined together between the carboxyl group of one amino acid and the amino group of another

Figure 5: Peptide link formation



- Condensation reaction(removal of elements of water from the two molecules) between the carboxyl group of one amino acid and the amino group of another.
- The hydroxyl group of the carboxyl group and a hydrogen of the amino group are lost to give rise to a water molecule and a dipeptide.

Peptide Formation

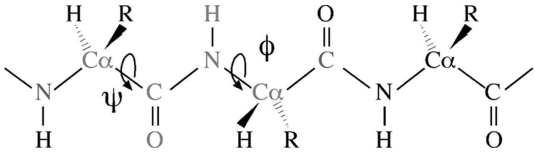
- The resulting product is called a dipeptide and the newly formed covalent bond connecting the two amino acids is called a **peptide bond**.
- Multiple amino acids can be joined together to form a longer chain of amino acid polymer.
- A linear polymer of more than fifty amino acid residues is referred to as a polypeptide and a polypeptide also called **protein**.
- The residues in a peptide or polypeptide are numbered beginning with the residue containing the amino group, referred to as the **N-terminus** , and ending with the residue containing the carboxyl group, known as the **C-terminus**
- The actual sequence of amino acid residues in a polypeptide determines its ultimate structure and function

Dihedral Angles

- A peptide bond is actually a partial double bond owing to shared electrons between O=C–N atoms
- The rigid double bond structure forces atoms to lie in the same plane, called the **peptide plane**
- The angle of rotation by the two bonded pairs of atoms around the peptide bond is referred to as the dihedral angle (also called the **tortional angle**).
- One is the dihedral angle along the N–Cα *bond*(ϕ)
- the other is the angle along the Cα–*C bond*(ψ)
- Various combinations of those angles allow the proteins to fold in many different ways.

Dihedral Angles

Figure 6: Dihedral angles

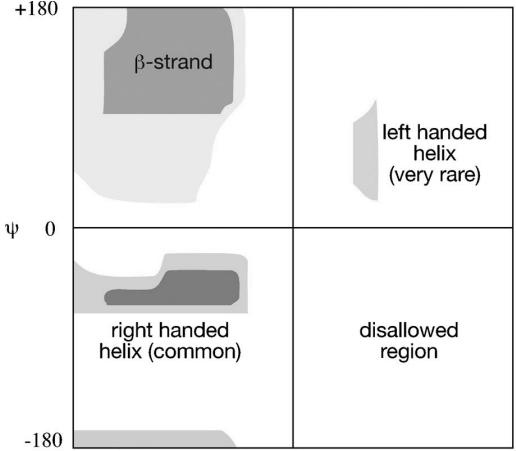


Ramachandran Plot

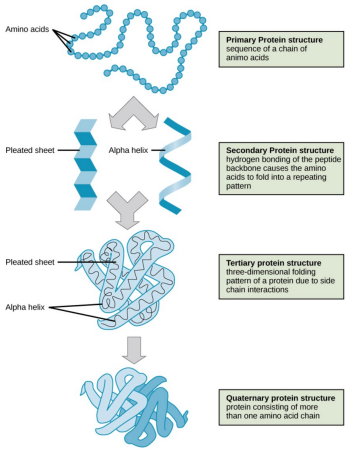
- The rotation of phi and psi is not completely free because of the planar nature of the peptide bond
- When phi and psi angles of amino acids of a particular protein are plotted against each other, the resulting diagram is called a Ramachandran plot.
- This plot maps the entire conformational space of a peptide and shows sterically allowed and disallowed regions
- It can be very useful in evaluating the quality of protein models.

Ramachandran Plot

Figure 7: A Ramachandran plot



Hierarchy Of Protein Structure



Hierarchy

- Protein structures can be organized into four levels of hierarchies with increasing complexity
  - ⋆ **Primary structure**: Amino acid sequence of its polypeptide chain
  - ⋆ **Secondary structure**: Spatial arrngement of the polypeptide backbone
  - ⋆ **Tertiary structure**: Three dimensional structure of entire polypeptide
  - ⋆ **Quaternary structure**: Three dimensional structure of proteins that are composed of two or more polypeptide chains

Stabilizing Forces

- Protein structures from secondary to quaternary are maintained by noncovalent forces.
  - ⋆ **electrostatic interactions**
  - ⋆ **van der Waals forces**
  - ⋆ **hydrogen bonding.**
- Electrostatic interactions are a significant stabilizing force in a protein structure. They occur when excess negative charges in one region are neutralized by positive charges in another region
- The result is the formation of salt bridges between oppositely charged residues.
-

Stabilizing Forces

- Hydrogen bonds are a particular type of electrostatic interactions involving hydrogen from one residue and oxygen from another
- Hydrogen from the hydrogen bond donor group such as the N–H group is slightly positively charged, whereas oxygen from the hydrogen bond acceptor group such as the C=Ogroup is slightly negatively charged. When they come within a close distance (¡3 Å), a partial bond is formed between them, resulting in a hydrogen bond

Stabilizing Forces

- Van der Waals forces are instantaneous interactions between atoms when they become transient dipoles.
- A transient dipole can induce another transient dipole nearby and the dipoles of the two atoms can be reversed a moment later. The oscillating dipoles result in an attractive force.
- The van der Waals interactions are weaker than electrostatic and hydrogen bonds and thus only have a secondary effect on the protein structure.

SECONDARY STRUCTURES

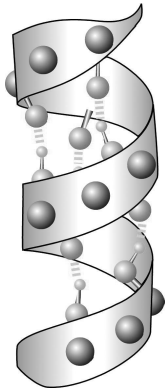
- Local structures of a protein with regular conformations are known as secondary structures
- Stabilized by hydrogen bonds formed between carbonyl oxygen and amino hydrogen of different amino acids

- Chief elements of secondary structures are $\alpha - helices\ and\ \beta - sheets$.

$\alpha -$ *Helices*

- An α $-$
  helix has a main chain backbone conformation and exhibiting a rightward spiral form.
- There are 3.6 amino acids per helical turn
- The structure is stabilized by hydrogen bonds formed between the main chain atoms of residues i and i $+$ 4.
- The hydrogen bonds are nearly parallel with the helical axis
- Hydrophobic residues of the helix tend to face inside and hydrophilic residues of the helix face outside.
- every third residue along the helix tends to be a hydrophobic residue.
- Ala, Gln, Leu, and Met are commonly found in an $\alpha -$ *helix*, *butnotPro*, *Gly*, *andTyr*

α – *Helices*



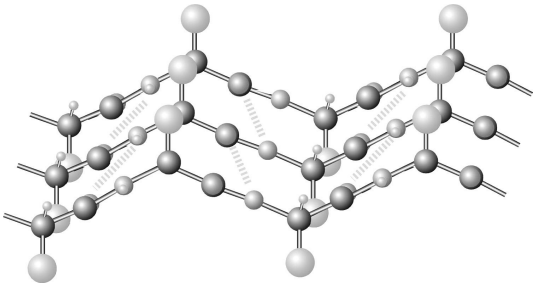


Figure 8: A ribbon diagram of an

β – Sheets

- It is a fully extended configuration built up from several spatially adjacent regions(-strand) of a polypeptide chain
- The -strand conformation is pleated with main chain backbone zigzagging and side chains positioned alternately on opposite sides of the sheet.
- -Strands are stabilized by hydrogen bonds between residues of adjacent strands
- -strands near the surface of the protein tend to show an alternating pattern of hydrophobic and hydrophilic regions, whereas strands buried at the core of a protein are nearly all hydrophobic.
- The -strands form a parallel sheet, antiparallel sheet or mixture of both
- The and angles are also widely distributed in the upper left region and it is more difficult to predict -sheets than - helices

Coils and Loops

- There are also local structures that do not belong to regular secondary structures and they are coils or loops
- The loops are often characterized by sharp turns or hairpin-like structures. If the connecting regions are completely irregular, they belong to random coils.
- Residues in the loop or coil regions tend to be charged and polar and located on the surface of the protein structure.
- They are often the evolutionary variable regions where mutations, deletions and insertions frequently occur and these are often the active sites of proteins

TERTIARY STRUCTURES

- The overall packing and arrangement of secondary structures form the tertiary structure
- The tertiary structure can come in as either globular or membrane proteins
- The globular proteins exists in solvents through hydrophilic interactions with solvent molecules
- the membrane proteins exists in membrane lipids and is stabilized through hydrophobic interactions with the lipid molecules. of a protein

Globular Proteins

- Globular proteins are usually soluble and surrounded by water molecules
- spherical shape with polar or hydrophilic residues on the surface and hydrophobic residues in the core
- Common examples of globular proteins are enzymes,myoglobins, cytokines, and protein hormones

Integral Membrane Proteins

- Membrane proteins exist in lipid bilayers of cell membranes
- Because they are surrounded by lipids, the exterior of the proteins spanning the membrane must be very hydrophobic to be stable
- Common examples of membrane proteins are rhodopsins, cytochrome c oxidase, and ion channel proteins

DETERMINATION OF PROTEIN THREE-DIMENSIONAL STRUCTURE

Protein three-dimensional structures are obtained using two popular experimental techniques

- **x-ray crystallography**
- **nuclear magnetic resonance (NMR) spectroscopy**

X-ray Crystallography

- In x-ray protein crystallography, proteins need to be grown into large crystals in which their positions are fixed in a repeated, ordered fashion
- The protein crystals are illuminated with an intense x-ray beam
- Thex-rays are deflected by the electron clouds producing a regular pattern of diffraction
- The diffraction pattern can be converted into an electron density map using Fourier transform.
- To interpret a three-dimensional structure from two-dimensional electron density maps requires solving the phases in the diffraction data. The phases refer to the relative timing of different diffraction waves hitting the detector.

X-ray Crystallography

- Phase solving can be carried out by two methods
  - molecular replacement
    - # Molecular replacement uses a homologous protein structure as template to derive aninitial estimate of the phases
  - multiple isomorphous replacement
    - # Multiple isomorphous replacement derives phases by comparing electron intensity changes in protein crystals containing heavy metalatomsandthe ones without heavy metalatoms

X-ray Crystallography

- Once the phases are available, protein structures can be solved by modeling with amino acid residues that best fit the electron density map.
- The quality of the final model is measured by an R factor
- The R factor is expressed as a percentage of difference between theoretically reproduced diffraction data and experimentally determined diffraction data.
- R values can range from 0.0, which is complete agreement, to 0.59, which is complete disagreement

Nuclear Magnetic Resonance Spectroscopy

- NMR spectroscopy detects spinning patterns of atomic nuclei in a magnetic field
- A radiofrequency radiation is used to induce transitions between nuclear spin states in a magnetic field
- Interactions between spinning isotope pairs produce radio signal peaks that correlate with the distances between them
- By interpreting the signals observed using NMR,proximity between atoms can be determined
- Knowledge of distances between all labeled atoms in a protein allows a protein model to be built that satisfies all the constraints.