# Proteins

- **Biomolecule, macromolecule**
  - more than 50% of the dry weight of cells → proteins
- Polymer of amino acids connected into linear chains
- Strings of symbols
- Machinery of life
  - play central role in the structure and function of cells
  - regulate and carry out many biological functions
- Enzymes or catalytic proteins (trypsin, DNA polymerase)
- Contractile proteins (actin, myosin)
- Structural or cytoskeletal proteins (tropocollagen, keratin)
- Effector proteins (insulin, epidermal growth factor)
- Defense proteins (immunoglobulins, thrombin)
- Receptors (CD4, acetylcholine receptor)
- Repressor proteins (Jun, Fos)
- Chaperones (GroEL, DnaK)
- Storage proteins (ferritin, gludin)
Protein

(a) amino acid

(b) peptide bond
Zwitterion

Charged ion, but overall is neutral!

Amino group: weak base; Carboxyl group: weak acid

\[
\text{acidic environment} \quad \xrightarrow{\text{pH 2.0}} \quad \text{basic (alkaline) environment} \quad \xrightarrow{\text{pH 9.0}}
\]

pH = 7 at 25°C (neutral environment)

\[\text{pH} = -\log_{10}(\text{[H}^+\text{]}) \Rightarrow [\text{H}^+] \text{ in pure water is } 10^{-7} \text{ mol/l} \]
Isoelectric point (pI)

Let's consider amino acid alanine (R group is CH₃).

When

\[ \text{pH} = \text{pI} \approx 6 \]

the net charge of alanine is 0

\( pK_1 \) (\( pK_2 \)) – half of the alanines in the solution are zwitterions, while the rest are cations (anions)

http://www.indstate.edu/thcme/mwking/amino-acids.html
A note on stereoisomers

- there are L- and D-amino acids
- only L amino acids are used in proteins
- D-amino acids are found in the cell walls of bacteria (protects them from the enzymes of the host organism)

D-alanine

L-alanine

http://chemed.chem.purdue.edu/genchem/topicreview/bp/1biochem/amino2.htm
Staggered vs. Non-staggered (Aligned)

Ethane:
$\text{CH}_3\text{CH}_3$

Valine:
$\text{CH}_3\text{CH}_3$

**Rotamers**: side chain conformations that occur most frequently.
Peptide Bonds

- they are planar, nearly
- they are very strong

a) Pauling’s theoretical model (1951)
b) experimentally determined bond lengths
Amino Acids

(a) Hydrophobic amino acids

- Ala, Alanine
- Val, Valine
- Phe, Phenylalanine
- Pro, Proline
- Ile, Isoleucine
- Leu, Leucine
- Met, Methionine

©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group
Amino Acids

D Asp, Aspartic acid

E Glu, Glutamic acid

K Lys, Lysine

R Arg, Arginine
Amino Acids

- Ser, Serine
- Thr, Threonine
- Tyr, Tyrosine
- His, Histidine
- Cys, Cysteine
- Asn, Asparagin
- Gln, Glutamine
- Trp, Tryptophan
Amino Acids

Aliphatic side chains: A, V, L, I
Hydroxyl-containing residues: S, T, Y
Acidic residues (- charged): D, E
Amide-containing residues: N, Q
Sulfur-containing residues: C, M
Basic residues (+ charged): K, R, (H)
Aromatic residues: F, Y, W

Other: P, H, G
Ramachandran Plot

1. **α region** – corresponds to the residues found typically in the alpha helices
2. **β region** – corresponds to the residues found typically in the beta sheets
3. **L region** – corresponds to the residues typically found in the left-handed helices

- most angles are not allowed because of steric collisions
  - between atoms of the same residue
  - between atoms of the neighboring residues
- the allowed combinations can be calculated
a) observed Ramachandran angles for all residues except glycine
b) observed Ramachandran angles for glycine
Ramachandran Angles

- Almost completely describe overall fold of a protein

- Why?
  - Bond *lengths* are approximately fixed
  - Bond *angles* are approximately fixed
  - Only Ramachandran (torsion, dihedral) angles are variable
  - Only 2 variables per amino acid (for the backbone)
Peptide Bond Revisited

- Usually trans
  - with $\omega \approx 180^\circ \pm 6^\circ$ rms

- Occasionally cis
  - with $\omega \approx 0^\circ$
  - $\sim 1/4$ of prolines
  - very infrequently glycines
  - almost never other amino acids

Michael S. Chapman
Levels of Protein Structure

Native state (conformation) – conformation at which protein shows its activity
Helical Structure

a) idealized diagram
b) the same as a) but with approximate positions for main-chain atoms
c) schematic diagram of an alpha helix
d) a ball and stick model
Helical Structure
Other Helical Conformations

- **$3_{10}$ helix ($\Phi = -49$, $\Psi = -26$)**
  - found in proteins when a regular helix is distorted by the presence of unfavorable residues, near turn regions or in short helices
  - hydrogen bonds between $i$ and $i+3$ (instead of $i+4$)
  - 3 residues per turn and 10 backbone atoms between donor and acceptor atom
  - tighter and narrower
- **$\pi$ helix ($\Phi = -57$, $\Psi = -70$)**
  - more loosely coiled
  - hydrogen bonds between $i$ and $i+5$ (instead of $i+4$)
  - 4.4 residues per turn
  - can be very long
- **Poly(Pro) helices**
  - all cis with 3.3 residues per turn, right-handed (type I) (-83, +158)
  - all trans with 3 residues per turn, left-handed (type II) (-78, +149)
- **Poly(Gly) chains**
  - type I: beta conformation
  - type II: similar to poly(Pro) helix with 3 residues per turn
Polyproline II Helices (PPII)

- shorter than regular helices (4-5 residues)
- longer (physically) than regular helices (rise per turn is twice that of regular helices – 9.3Å)
- seem to be stabilized by main chain – water hydrogen bonds
- found mostly on protein surface
- preference for hydrophilic residues and proline
- Gln, Ser, Arg and Ala are found in PPII regions, Gly is rare
- involved in protein-protein interactions
- important roles in signal transduction, transcription, cell motility, etc.
- contain sequence motifs, e.g. PXXP
- dominant element of secondary structure in unfolded proteins (Horng and Raines, Prot. Sci. 2006)
Anti-parallel $\beta$-sheet
Anti-parallel $\beta$-sheet

Typical length: 5-10 residues
Parallel $\beta$-sheet
Only about 20% of sheets are of mixed type.

Almost all sheets have twisted strands with fixed handedness (right-handed twist)
a) histogram showing the frequency of hairpin loops of different lengths in 62 proteins

b) the two most frequently occurring two-residue hairpin loops
Turn types

Type I turn

Type II turn

Jan Feng, Protein Structure
Super Secondary Structure

Helix-loop-helix motif

- helix-loop-helix motif with $\text{Ca}^{2+}$ atom attached

![Diagram](Image)

**Table 2.2** Amino acid sequences of calcium-binding EF motifs in three different proteins

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
<th>Calcium-binding Residues</th>
<th>Hydrophobic Core</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farvalbumin</td>
<td>V K K A F A I D D O D K S G F I E E D E L K L F L O N E</td>
<td>Orange</td>
<td>Light Green</td>
</tr>
<tr>
<td>Calmodulin</td>
<td>E K E A F S L D D K D D G T I T T K E L G T V M R S L</td>
<td>Orange</td>
<td>Light Green</td>
</tr>
<tr>
<td>Troponin-C</td>
<td>D A D C F I E D K N A D G F I D I E E L G E I L R A T</td>
<td>Orange</td>
<td>Light Green</td>
</tr>
</tbody>
</table>

Calcium-binding residues are orange, and residues that form the hydrophobic core of the motif are light green. The helix-loop-helix region shown underneath is colored as in Figure 2.13.
Domains

Four Helix Bundle

- hydrophobic residues tend to be on the inside
- polar residues tend to be on the outside of proteins

a) four helix bundle – red cylinders are helices while green parts are loops

b) projection from above

Introduction to Protein Structure by Branden and Tooze
Disulphide Bonds (Bridges)

- covalent bond between two cysteins (i.e. their sulfur atoms)
- require oxidative environment
- present in extracellular proteins
- stabilize proteins
- “create” so-called long-range interactions

Introduction to Protein Structure by Branden and Tooze
Levinthal’s Calculation

- Cyrus Levinthal 1968

- **Q:** do proteins explore all possible conformations before they adopt a specific 3-D structure?

- **A:** let’s consider a simplified problem
  - each residue can adopt one of the three discrete groups from the Ramachandran plot (alpha, beta, loop)
  - a switch between conformations can be done in $10^{-12}$ seconds
  - then, a protein with 150 residues would need to explore $3^{150}$ possible states, which is $10^{71}$
  - at the rate of $10^{-12}$ a protein would need $\sim10^{50}$ years

- we know that protein folds between 0.1s and 1000s
Protein Folding Problem

- How do proteins fold into a specific 3-D structure?
- How does the primary structure of a protein determine its secondary and tertiary structure?

- there are two conditions a protein needs to meet
  - there must be a single, stable, folded conformation (thermodynamic condition)
  - a protein must fold on an appropriate time scale (kinetic condition)

- thus, only a small amount of conformational space is explored
- also, there must exist a specific folding pathway
- the paradox how proteins quickly fold into specific 3-D conformations is called a protein folding problem