Principles of Protein Structure
Amino Acids - Basic Building Blocks of Proteins

**FIGURE 2.2** The 20 different side chains of the amino acids.
Amino Acids Have

**FIGURE 2.1** Illustration of the CORN rule for the L-conformation of amino acids. The \( \alpha \)-carbon is viewed down the H-C bond.
Molecular Chirality

Crystals and molecules can also be either right-handed or left-handed (such as quartz crystals), although many crystals and molecules have no handedness. The discovery of chirality itself is related to crystals. In his famous experiments in 1848, Louis Pasteur recrystallized a salt of tartaric acid and obtained two kinds of small crystals; their shapes were mirror images of each other.

Molecules of many substances can also be right-handed or left-handed. Living organisms contain a large number of such molecules. All naturally occurring amino acids are chiral (except one of them: glycine).

Here is an illustration showing a pair of hands with a pair of right-handed and left-handed amino acid molecules. This illustration appeared in a book by R. N. Bracewell discussing the possibility of intelligent life in outer space.
Proteins are Polymers

FIGURE 2.3 The peptide bond. The top left figure shows the definition of a torsion angle. The middle figure, where we look along the bond between atoms B and C, shows how we can determine the angles between the bonds AB and CD. The top right drawing shows the names of the torsion angles. In a trans peptide $\omega \approx 180^\circ$, since the atoms of the peptide bond tends to form a plane (the amide plane). Bottom: The distances and angles between the atoms of a peptide bond.
Cis Proline in the Active Site of HCV NS2

Cis trans Proline Isomerase

- Cyclophilins are a family of proteins that catalyze the isomerization of peptide backbones at a proline
- Critically important for protein folding
- Cyclosporine is a potent immunosuppressant and an inhibitor of cyclophilins
- Commonly used after organ transplant
Four Levels of Protein Structure

- **Primary, 1°**
  - Amino acid sequence; Covalent bonds

- **Secondary, 2°**
  - Local conformation of main-chain atoms ($\Phi$ and $\Psi$ angles); non-covalent interactions (h-bonds)

- **Tertiary, 3°**
  - 3-D arrangement of all the atoms in space (main-chain and side-chain); non-covalent interactions

- **Quaternary, 4°**
  - 3-D arrangement of subunit chains, non-covalent interactions
Four Levels of Protein Structure

- Primary, $1^\circ$
  
  TPEEKSAVTALWGKV

- Secondary, $2^\circ$

- Tertiary, $3^\circ$

- Quaternary, $4^\circ$
**Side Chain Conformation**

**FIGURE 2.6** Illustration of the torsion angles in a side chain. The angle $\chi_1$ is the angle between the N-C$\alpha$ bond and the bond from C$\beta$ to C$\gamma$ (or one of the $\gamma$ atoms). **Left:** A view down the C$\alpha$-C$\beta$ bond showing the staggered conformation corresponding to a $\chi_1$ angle of $-60^\circ$. The other staggered conformations correspond to a rotation of $120^\circ$ in the positive or negative direction. **Right:** The common rotamer for a Glu side chain with $\chi_1$ close to $-60^\circ$ and $\chi_2$ close to $180^\circ$ is shown.
Protein Backbone Conformation

**Figure 2.4** Ramachandran plot. Top: allowed regions based on steric clashes according to Ramachandran’s original analysis for alanine (left) and glycine (right). Bottom: Observed angles based on about 1000 experimentally determined protein structures, non-glycine residues (left) and glycines (right). The β region is split into two sections. Note that the observed distributions differ considerably from the expected distributions from the steric clashes, especially for glycine residues. (Reprinted with permission from Hövöller et al. (2002) Conformations of amino acids in proteins. *Acta Cryst D58*: 768–776. Copyright (2003) Elsevier.)
FIGURE 2.7  ■ The $\alpha$-helix. Left: The main chain and C$\beta$ atoms (gray) of an $\alpha$-helix. The pitch (rise per turn) is 5.4 Å. Right: The same $\alpha$-helix showing the side chains. The backbone is drawn schematically, with the C$\beta$ atoms pointing towards the N-terminus of the helix (down).
α Helix

- If N-terminus is at bottom, then all peptide N-H bonds point “down” and all peptide C=O bonds point “up”.
- N-H of residue n is H-bonded to C=O of residue n+4.
- α-Helix has:
  - 3.6 residues per turn
  - Rise/residue = 1.5 Å
  - Rise/turn = 5.4Å
Secondary Structure: Alpha Helices

Right handed

Left handed
$3_{10}$ and $\pi$ Helices

**Figure 2.8** A $3_{10}$ helix from *Aplysia limacina* myoglobin (PDB: 1MBA) and a $\pi$-helix from methane monoxygenase hydroxylase from *Methylococcus capsulatus* (PDB: 1MTY).

$3_{10}$ helix H-bonds $n$ and $n+3$

$\pi$ helices H-bonds $n$ and $n+5$
α Helix

- **R-groups in α-helices:**
  - extend radially from the core,
  - shown in helical wheel diagram.
  - Can have varied distributions

**Polar**

**Hydrophobic**

**Amphipathic**
Secondary Structure: Beta Sheet

Antiparallel beta sheet

Parallel beta sheet
β Sheet

- Stabilized by H-bonds between N-H & C=O from adjacent stretches of strands
- Peptide chains are fully extended pleated shape because adjacent peptides groups can’t be coplanar.
Parallel
Not optimum H-bonds; less stable

Anti-parallel
Optimum H-bonds; more stable
Beta Turn – 2 Conformations

Only Difference
Tertiary Structure

- Charge based interactions
  - 62R:163E
  - 55E:170R

- Hydrophobic interactions
  - 189V
  - 201L
  - 213I
  - 215L
  - 266L

- Disulfide bond
  - 203C:259C

1HSA Peptide bound to Class I MHC
Quaternary Structure
# Figure B.2

The beginning of the header and of the first part of the list of atomic coordinates of a PDB entry (1ZA7). The coordinates of the first two residues in the entry (Gln A40 and Gly A41) are given. The first three columns of numbers are the coordinates of the atoms in Å in relation to a suitable coordinate system. The fourth column of numbers is the occupancy (1.0 for full occupation is normal). In high-resolution structures, some side chains may be modeled in alternative conformations, and multiple coordinates for atoms are given their respective occupancy. In addition, bound molecules may have partial occupancy indicated by this number. The next column gives an estimate of the thermal disorder (the B-factors; see Fig. B.3).
Interatomic Distances

Atomic coordinates

\[ \mathbf{a}_1 = (a_{1x}, a_{1y}, a_{1z}) \quad \quad \quad \quad \quad \mathbf{a}_2 = (a_{2x}, a_{2y}, a_{2z}) \]

Bond length

\[ b_{12} = \sqrt{(a_{2x} - a_{1x})^2 + (a_{2y} - a_{1y})^2 + (a_{2z} - a_{1z})^2} \]
Bond lengths remain constant:

- N-CA: $1.459 \pm 0.003 \text{ Å}$
- CA-C: $1.527 \pm 0.003 \text{ Å}$
- C-N: $1.329 \pm 0.001 \text{ Å}$

**Bonded distances don't change much:**

- Chain A backbone
  - PDB file 3e7l
  - Bond lengths: $1.45 \pm 0.003 \text{ Å}$
Bond Angles

Atomic coordinates

\[ a_1 = (a_{1x}, a_{1y}, a_{1z}) \quad a_2 = (a_{2x}, a_{2y}, a_{2z}) \quad a_3 = (a_{3x}, a_{3y}, a_{3z}) \]

Bond vectors

\[ b_{12} = (a_{2x} - a_{1x}, a_{2y} - a_{1y}, a_{2z} - a_{1z}) \quad b_{23} = (a_{3x} - a_{2x}, a_{3y} - a_{2y}, a_{3z} - a_{2z}) \]

Scalar product

\[ b_{12} \cdot b_{23} = (a_{2x} - a_{1x})(a_{3x} - a_{2x}) + (a_{2y} - a_{1y})(a_{3y} - a_{2y}) + (a_{2z} - a_{1z})(a_{3z} - a_{2z}) \]

\[ b_{12} \cdot b_{23} = b_{12} b_{23} \cos(\pi - \theta_{123}) \]

\[ \cos \theta_{123} = -b_{12} \cdot b_{23} / (b_{12} b_{23}) \]

\[ \cos(\pi - \theta_{123}) = \cos \pi \cos \theta_{123} + \sin \pi \sin \theta_{123} = -\cos \theta_{123} \]
Distribution of Backbone Angles

- N-CA-C: 112.0±1.5°
- CA-C-N: 116.4±0.5°
- C-N-CA: 121.7±0.7°
Distribution of B-factors

**Figure B.3** A plot of the B-factors of the Cα atoms of Ras (one of the three identical chains in the entry 1ZA7). The B-factors are normally higher in surface loops and at the termini of the chain.
Motifs, Topologies and Folds: $\beta$ Sheet

The arrangement of secondary structure elements that give rise to a folded entity.

**Figure 2.14** A $\beta$-hairpin and a four-stranded up-and-down sheet.
Motifs, Topologies and Folds: \( \beta \) Sheet

**FIGURE 2.15** Two up-and-down sheets, the open sheet in the coat protein subunit of phage MS2 (PDB: 2MS2), and the closed cylinder in a bacterial porin, a protein from the outer membrane of *E. coli* (PDB: 2OMF).

**FIGURE 2.16** Up-and-down sheets in a propeller structure: the seven-bladed propeller of the \( \beta \)-subunit of transducin (PDB: 1GOT).

**FIGURE 2.17** A schematic drawing of a Greek key motif and the same motif in a protein (Micrococal nuclease, PDB: 2SNS). The 5-stranded sheet and the helix in the loop connecting strands 3 and 4 is an example of the OB (oligonucleotide/oligosaccharide binding) fold found in many proteins.

**FIGURE 2.18** Left: A jellyroll topology (top) can be seen as a \( \beta \)-hairpin rolled up. Right: The coat protein of the plant virus STNV, a simple jellyroll fold (PDB: 2BUK).
Motifs, Topologies and Folds: $\beta\alpha\beta$

**Figure 2.19** - A $\beta$-helix protein: pectin lyase from Aspergillus niger (PDB: 1IDK).

**Figure 2.20** - Schematic drawing of a $\beta\alpha\beta$ unit. Following the direction of the polypeptide chain a right-hand screw is formed.

**Figure 2.21** - Left: Schematic drawing of the topology called the Rossmann fold. Right: The NAD-binding domain of Sulfolobus solfataricus alcohol dehydrogenase, a typical Rossmann fold (order of strands 654123). The first $\beta\alpha\beta$ motif of the domain is in blue, the second in red and the connection in yellow. The nucleotide is bound at the C-terminal ends of the $\beta$-strands at the center of the sheet. The phosphates of the NAD molecule are located at the N-termini of two helices where their binding is favored by the helical dipoles (PDB: 1R37).

**Figure 2.22** - The TIM barrel of triose phosphate isomerase from Plasmodium falciparum. The order of the strands is 12345678. The active site is occupied by a transition-state analog, phosphoglycolate. In all TIM barrel structures, the active site is found at the C-terminal end of the $\beta$-strands (PDB: 1LYX).
Motifs, Topologies and Folds: $\alpha$-helical

**FIGURE 2.23** Schematic drawing of an up-and-down four-helix bundle and the coat protein of tobacco mosaic virus (PDB: 2TMV).

**FIGURE 2.24** The myoglobin structure where most of the helices pack with an approximately 50° angle between helix axes. The protein binds a heme group (PDB: 1MBA).
Motifs, Topologies and Folds: $\alpha$-helical

Middle Domain of eIF4G - scaffold protein for translation initiation factors.
Protein Domains

- A folded unit of protein
- normally formed around a hydrophobic core
- Tend to be globular
- 150-250 amino acids
Protein Domains

Many Eukaryotic proteins consist of multiple domains

- Limited proteolysis can be used to experimentally determine domain limits.
Limited Trypsin Digestion in the Absence and Presence of dsRNA

<table>
<thead>
<tr>
<th>dsRNA</th>
<th>Helicase-RD</th>
<th>Full-length RIG-I</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>0 15 30 60 120</td>
<td>0 15 30 60 120</td>
</tr>
<tr>
<td></td>
<td>+ + + + +</td>
<td>- - - - -</td>
</tr>
<tr>
<td>116</td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 2.26  ■ Domain swapping of a protein. Normally the protein is a monomer (top) where the two domains interact within the monomer. Bottom: The two domains of the protein interact in the same way with a different monomer.
Location of His 143, Glu 163, Cys 184
HIV Protease

Retroviral aspartic protease - dimerization forms a single active site
Structural Comparison

• Structural comparison requires a 3-Dimension search

• Minimize the search by using just secondary structural elements - Potential problem?

• DALI server
  http://ekhidna.biocenter.helsinki.fi/dali_server/

• PDBe/fold
  http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmsserver

• VAST
SCOP: Structural Classification Of Proteins

SCOP database classifies domains not entire proteins

**Figure B.5** SCOP classification. The example shows the hierarchy for immunoglobulin domains. The number of different subdivisions of a group is shown in parentheses.
How to classify proteins with same general arrangement of secondary structure but are connected differently?