COMPUTATIONAL STRUCTURAL BIOLOGY

STRUCTURE, SIMULATION, FUNCTION & PREDICTION

Lecture 2

Michael Levitt
Structural Biology, Stanford

http://csb.stanford.edu/class
LECTURE 2

Polypeptide Chain
Amino Acids
Degrees of Freedom
Reverse Turns
The Alpha Helix
The Beta Sheet
Polypeptide Chain
Concept 2.1
THE POLYPEPTIDE CHAIN

Chemical Structure

The Peptide Group
CHEMICAL STRUCTURE

- The amino acid side chain is attached to the CA atom.
- It has L chirality always pointing up in the orientation shown.
- An amino acid runs from the NH group to the CO group.
THE PEPTIDE GROUP

- The peptide group between amino acids is rigid and planar (resonance).
- It has co-linear bonds entering and leaving.
- The peptide group has a strong dipole moment due to partial charges on NH and CO groups.
Amino Acids
Concept 2.2
AMINO ACIDS

Amino Acid Side Chains

Amino Acid Names

Amino Acid Classification 1

Amino Acids: Gly & Pro

Amino Acids: Cys

Amino Acids: Leu & Phe

Amino Acids: Glu & Arg

Amino Acids: Val & Ile

Amino Acid Classification 2
AMINO ACID SIDE CHAINS

- There are 20 naturally occurring amino acids.
- It is worth learning the amino acid names.
- There are many textbooks that say much more about amino acids.

Greek letters are not used in computers.
# Amino Acid Names

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Abbreviation</th>
<th>One Letter</th>
<th>Full Name</th>
<th>Abbreviation</th>
<th>One Letter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
</tr>
<tr>
<td>Aspartic Acid</td>
<td>Asp</td>
<td>D</td>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
</tr>
<tr>
<td>Glutamic Acid</td>
<td>Glu</td>
<td>E</td>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
<td>ARginine</td>
<td>Arg</td>
<td>R</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
<td>Valine</td>
<td>Val</td>
<td>V</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
<td>TYrosine</td>
<td>Tyr</td>
<td>Y</td>
</tr>
</tbody>
</table>

It is strongly advised that you learn these names.
AMINO ACID CLASSIFICATION 1

Number of Sidechain Heavy Atoms

0 1 2 3 4 5 6 7 8 10

Amino Acid Classification:

- Polar (hydrophilic)
- Negative
- Positive
- Bulky
- Aromatic
- Nonpolar (hydrophobic)

There are many properties.
They cluster logically.
AMINO ACIDS GLY AND PRO

- Glycine is simplest, without a sidechain. It is very flexible.

- Proline is the only cyclic amino acid used in proteins. It is very rigid.
AMINO ACIDS CYS

- Cystine is special in that it forms covalent bonds.

- These are called SS bonds or disulfide bridges.

- They make closed loops in the polypeptide chain.
AMINO ACIDS LEU AND PHE

- Leucine is a branched aliphatic hydrophobic amino acid.

- Phenylalanine is an aromatic hydrophobic amino acid. It is one of the largest amino acids.
AMINO ACIDS GLU AND ARG

- Glutamic Acid is a large polar amino acid.
  It accepts hydrogen bonds and is normally ionized (-1e).

- Arginine is a large polar amino acid.
  It donates hydrogen bonds and is normally ionized (+1e).
AMINO ACIDS VAL AND ILE

- Valine is a branched aliphatic hydrophobic amino acid. It branches at the CB atom.

- Isoleucine is also a branched aliphatic hydrophobic amino acid.

These amino acids are "Bulky"
AMINO ACID CLASSIFICATION 2

Aromatic

Bulky

Nonpolar (hydrophobic)

Polar (hydrophilic)

Negative

Positive

Special

Key

Learn to recognize the amino acids.

ACG

TPV

MLI

NDE

KFR

Michael Levitt 04
Degrees of Freedom
Concept 2.3
DEGREES OF FREEDOM

Backbone Degrees of Freedom
Steric Clashes Limit ($\phi, \psi$)
Ramachandran Diagram
Different Backbone Conformations
Contour Plots
($\phi, \psi$) Distributions: A, G, P, N
($\phi, \psi$) Distributions: A, M, F, V
Side Chain Conformations
**BACKBONE DEGREES OF FREEDOM**

- The torsion angle rotating about the N-CA bond is called $\phi$
- The torsion angle rotating about the CA-C bond is called $\psi$
- Together they are the $(\phi, \psi)$ angles
STERIC CLASHES LIMIT PHI, PSI

- The peptide O atom is much bigger than the peptide H atom. It can clash with the side chain.

- As the CO group is closer for rotation about $\phi$ than for $\psi$, the clash is worse.
RAMACHANDRAN DIAGRAM

- Map the clashes. (Done in 1963).
- Green is allowed.

Non-GLY amino acids

GLY special No CB
DIFERENT BACKBONE CONFORMATIONS

- Straight chain (used for beta-sheet).
- CO groups point in opposite directions.
- Bent chain (used for alpha-helix).
- CO groups point in same direction.
CONTOUR PLOTS

- Count number of $N(\phi, \psi)$ found in known proteins and then convert to a pseudo energy using $-\log(N(\phi, \psi))$
PHI, PSI DISTRIBUTIONS A, G, P, N

- GLY and PRO are special.
- ASN is a bit like GLY.
- Most are like ALA due to CB atom.
PHI, PSI DISTRIBUTIONS A, M, F, V

- Four amino acids with very different side chains have similar ($\phi, \psi$) preferences.
- They all have a CB atom.
SIDE CHAIN CONFORMATIONS

- Chi1 distributions are very non-uniform.
- The preferred chi1 values depend on the amino acid side chain and the backbone conformation.

\[ \chi \text{ is defined by N-CA-CB-CG.} \]
Reverse Turns
Concept 2.4
REVERSE TURNS

Beta Turns
Virtual Bonds
Beta Turns Types
**BETA TURNS**

- The chain directions changes 180 degrees.

- This can be done in many ways giving types: I, I', II, II', IV & VIII.

- There are sequence restrictions. For example, in Type II', position 2 must be GLY.

*From Mathews and van Holde: Biochemistry 2/e. © The Benjamin/Cummings Publishing Co., Inc.*
VIRTUAL BONDS

- $\alpha_i$ is defined by $\text{CA}_{i-1} - \text{CA}_i - \text{CA}_{i+1} - \text{CA}_{i+2}$
- $\alpha_i$ is approximately $\psi_i + \phi_{i+1} + 180$
**Beta Turn Types**

There are four allowed $(\phi, \psi)$ values:

- $\alpha_R (-60, -40)$
- $\beta_R (-120, +120)$
- $\alpha_L (+60, +40)$
- $\beta_L (+120, -120)$

For turn need small alpha angle:

$$\alpha_i = \psi_i + \phi_{i+1} + 180 = 0$$

Try allowed conformations at $i$ and $i+1$:

- $\alpha_R \alpha_R$ (I): $\alpha = -40 - 60 + 180 = 80$
- $\alpha_L \alpha_L$ (I’): $\alpha = +40 + 60 + 180 = 80$
- $\beta_R \alpha_L$ (II): $\alpha = +120 + 60 + 180 = 0$
- $\beta_L \alpha_R$ (II’): $\alpha = -120 - 60 + 180 = 0$
- $\alpha_R \beta_L$ (VIII): $\alpha = -40 - 120 + 180 = 20$
- $\alpha_L \beta_R$ (VIII’): $\alpha = +40 + 120 + 180 = -20$
The Alpha Helix Concept 2.5
THE ALPHA-HELIX

Proteins have Alpha-Helices
The Alpha-Helix
The Helix Surface
Helix Surface Ridges
Helix Ridge Lines
Different Possible Helices
3_{10} Explained
Alpha-helix Dipole
**PROTEINS HAVE HELICES**

- Myoglobin is built almost entirely from alpha-helices.
- Most other proteins have some helix.
THE ALPHA HELIX

- Pauling et al. 1951 original.
- The alpha helix is formed by NH...O=C hydrogen bonds.
THE HELIX SURFACE

- Make a helical net of the helix surface.
- Wrap a piece of paper around the helix, mark the positions of the sidechains and unwrap.

3.6 residues/turn, 5.4 Å/turn, 4/3 helix
HELIX SURFACE RIDGES

- Sidechains that are in positions $i$ and $i+3$ form a ridge that is a left-handed helix.
- Sidechains in positions $i$ and $i+4$ form a ridge that is a right-handed helix.
HELIX RIDGE LINES

i to i+3 ridge

i to i+4 ridge
DIFERENT POSSIBLE HELICES

- The $3_{10}$ helix is thin and has 3.0 residues/turn.
- The $\pi$ or $5_{16}$ helix is fat and has 4.2 residues/turn.
- The $\alpha$ helix is just right. It is a $4_{13}$ helix with 3.6 residues/turn.
310 EXPLAINED

- Find a hydrogen bonded loop.
- Count residues by number of CA atoms in the loop. Here 3.
- Count atoms in loop including O and H. Here 10.
ALPHA-HELIX DIPOLE 1

- The peptide group has a strong dipole moment due to partial charges on NH and CO groups.
**Alpha-Helix Dipole 2**

- In the α-helix, peptide dipoles line up to give the helix a large dipole.

- This dipole is equivalent to having a charge of +0.5e at the N-terminus and -0.5e at the C-terminus.

- The N-terminal +0.5e charge is important in many enzymes. It acts like a short positively charged side chain.
The Beta Sheet Concept 2.6
BETA SHEETS

Proteins have Beta-Strands

The Beta-Strand

Beta Sheets

Beta Sheet Sidedness

Beta Sheets are Solid

Beta Sheets are Twisted

The Beta-Hairpin
PROTEINS HAVE STRANDS

- Immunoglobulin domains are built entirely from β-strands.
- Most other proteins have some beta structure.
THE BETA STRAND

Extended chain is flat.

β strands twist

N-H...O=C Hydrogen bonds

Off diagonal so strand twists.  N-H...O=C Hydrogen bonds
BETA SHEETS

All parallel

Mixed parallel & anti-parallel
BETA SHEETS ARE SOLID

The hydrogen bonds bring the atoms of the strands in close contact forming a solid surface, which is flexible.
BETASHEET SIDEDNESS

- Every second residue in a \( \beta \)-strand is on the same side of the \( \beta \)-sheet.

- \( \beta \)-sheet can be totally amphipathic. Sequence is SLSLSLSLS.

Side chains on bottom of sheet  Side chains on top of sheet

Hydrophobic side  Hydrophilic side

©Michael Levitt 04
REAL SHEETS ARE TWISTED 1

Top view of parallel sheet in flavodoxin
REAL SHEETS ARE TWISTED 2
THE BETA HAIRPIN

- The beta-hairpin can be very twisted.
- This is because the beta-strand is twisted.
- It is possible to have long twisted hairpins as they do not involve increasing distortion.