Molecular Cell Biology

Proteins: Composition and Structure

Cooper
Amino Acids form a Polypeptide

methionine (Met)

aspartic acid (Asp)

leucine (Leu)

tyrosine (Tyr)

polypeptide backbone

side chains

amino terminus or N-terminus

peptide bonds

peptide bond

carboxyl terminus or C-terminus

H2O

H2O

H2O

Wednesday, August 29, 2012
## Basic chemistry of AA side chains

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

<table>
<thead>
<tr>
<th>POLAR AMINO ACIDS</th>
<th>NONPOLAR AMINO ACIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>negative</td>
<td>nonpolar</td>
</tr>
<tr>
<td>positive</td>
<td>nonpolar</td>
</tr>
<tr>
<td>uncharged polar</td>
<td>nonpolar</td>
</tr>
</tbody>
</table>

Wednesday, August 29, 2012
Acid/Base Properties of Amino Acids

The “salt” or conjugate base, $A^-$, is the ionized form of a weak acid. By definition, the dissociation constant of the acid, $K_a$, is:

$$K_a = \frac{[H^+][A^-]}{[HA]}$$

$$pH = pK_a + \log\frac{[A^-]}{[HA]}$$

Henderson-Hasselbalch Equation
Acid/Base Properties of Amino Acids

The "salt" or conjugate base, $A^-$, is the ionized form of a weak acid. By definition, the dissociation constant of the acid, $K_a$, is:

$$K_a = \frac{[H^+][A^-]}{[HA]}$$

Henderson-Hasselbalch Equation

$$pH = pK_a + \log \frac{[A^-]}{[HA]}$$
Some Definitions

- $pK_a =$ pH at which 50% of protons have been removed
- $pI =$ isoelectric point = pH at which net charge = 0
- $pK_1 =$ most acidic group
- $pK_2 =$ next most acidic group
- $pK_3 =$ etc.
Sample Titration Curves

Glycine

\[ \text{NH}_3 + \text{CH}_2 \text{COOH} \overset{pK_1}{\rightleftharpoons} \text{NH}_3 \text{CH}_2 \text{COO}^- \]

\[ \text{NH}_3 \text{CH}_2 \text{COO}^- \overset{pK_2}{\rightleftharpoons} \text{NH}_2 \text{CH}_2 \text{COO}^- \]

- \( pK_1 = 2.34 \)
- \( pK_2 = 9.60 \)
- \( pI = 5.97 \)

\( \text{pH} \)

\( \text{OH}^- \) (equivalents)
Sample Titration Curves

Histidine

$\text{p}K_1 = 1.82$

$\text{p}K_R = 6.0$

$\text{p}K_2 = 9.17$
## pK Values of Ionizable Side Chains

<table>
<thead>
<tr>
<th>Group</th>
<th>Acid $\rightleftharpoons$ base $+$ $H^+$</th>
<th>Typical pK*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminal carboxyl</td>
<td>$-\text{COOH} \rightleftharpoons -\text{COO}^- + H^+$</td>
<td>3.1</td>
</tr>
<tr>
<td>Aspartic and glutamic acid</td>
<td>$-\text{COOH} \rightleftharpoons -\text{COO}^- + H^+$</td>
<td>4.4</td>
</tr>
<tr>
<td>Histidine</td>
<td>$-\text{CH}_2\text{HN}_2\text{NH} \rightleftharpoons -\text{CH}_2\text{N}^- + H^+$</td>
<td>6.5</td>
</tr>
<tr>
<td>Terminal amino</td>
<td>$-\text{NH}_3^+ \rightleftharpoons -\text{NH}_2 + H^+$</td>
<td>8.0</td>
</tr>
<tr>
<td>Cysteine</td>
<td>$-\text{SH} \rightleftharpoons -\text{S}^- + H^+$</td>
<td>8.5</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>$\equiv\text{OH} \rightleftharpoons \equiv\text{O}^- + H^+$</td>
<td>10.0</td>
</tr>
<tr>
<td>Lysine</td>
<td>$-\text{NH}_3^+ \rightleftharpoons -\text{NH}_2 + H^+$</td>
<td>10.0</td>
</tr>
<tr>
<td>Arginine</td>
<td>$-\text{NH}_2\text{NH}_2\text{NH}^- \rightleftharpoons -\text{NH}_2\text{NH}_2\text{H}^- + H^+$</td>
<td>12.0</td>
</tr>
</tbody>
</table>

*P*K values depend on temperature, ionic strength, and the microenvironment of the ionizable group.*
Hydrophobicity Values of Side Chains

(Free Energies of Transfer)

<table>
<thead>
<tr>
<th>Residue</th>
<th>Normalized consensus$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
<td>0.48</td>
</tr>
<tr>
<td>Ala</td>
<td>0.62</td>
</tr>
<tr>
<td>Val</td>
<td>1.08</td>
</tr>
<tr>
<td>Leu</td>
<td>1.06</td>
</tr>
<tr>
<td>Ile</td>
<td>1.38</td>
</tr>
<tr>
<td>Pro</td>
<td>0.12</td>
</tr>
<tr>
<td>Cys</td>
<td>0.29</td>
</tr>
<tr>
<td>Met</td>
<td>0.64</td>
</tr>
<tr>
<td>Thr</td>
<td>-0.05</td>
</tr>
<tr>
<td>Ser</td>
<td>-0.18</td>
</tr>
<tr>
<td>Phe</td>
<td>1.19</td>
</tr>
<tr>
<td>Trp</td>
<td>0.81</td>
</tr>
<tr>
<td>Tyr</td>
<td>0.26</td>
</tr>
<tr>
<td>Gln</td>
<td>-0.85</td>
</tr>
<tr>
<td>Asn</td>
<td>-0.78</td>
</tr>
<tr>
<td>Glu</td>
<td>-0.74</td>
</tr>
<tr>
<td>Asp</td>
<td>-0.90</td>
</tr>
<tr>
<td>His</td>
<td>-0.40</td>
</tr>
<tr>
<td>Lys</td>
<td>-1.50</td>
</tr>
<tr>
<td>Arg</td>
<td>-2.53</td>
</tr>
</tbody>
</table>
Amide Side Chains: Asn and Gln

- Asn and Gln have distinct roles because of the extra length of the Gln side chain, from the single methylene group.
- Form H-bonds through amide groups. Amide $\delta$ oxygen is more important than the $\delta$ nitrogen in Asn H-bonding. Asn, but not Gln, can form a “pseudopeptide” bond by hydrogen bonding back with the main chain.
- Gln is a relatively indifferent, plain vanilla residue that goes reasonably well with almost anything and has no extreme properties or violent preferences or aversion. Asn, in contrast, is an interesting, quirky, residue with many unique properties.
- Gln side chain can cyclize with its $\alpha$-amino group to form pyroglutamic acid.
- Gln is not the most conservative substitution for Asn. As a helix starter, for example, Ser is the most conservative replacement for Asn.
Aliphatic Hydroxyl Side Chains: Ser & Thr

- Short-chain OH residues. OH can be either an H-bond donor or acceptor
- Chemically reactive (especially Ser): found in active sites (e.g., serine proteases), can undergo phosphorylation, carbohydrate attachment.
- Ser is common in tight turns, where it may H-bond with neighboring backbone NH or CO groups. Most often, interacting with solvent.
- Ser is common in disordered regions. Has no hydrophobic character and can interact with solvent.
- Thr prefers β structure, especially antiparallel β, which typically has one side exposed to solvent.
- Extra methyl group makes Thr more hydrophobic than Ser, so Thr is more often in the interior and less often in turns and disordered structures.
Acidic Side Chains: Asp and Glu

- Asp: Common helix starter and turn-forming residue. Glu is also a turn-forming residue, as are all hydrophilic residues, but is less position-specific than Asp.

- Asp and Glu are more frequent in the first turn of alpha-helices, whereas Lys, Arg, and His are more common toward the C-terminus. Those positions promote interaction with the helix dipole.

- Asp and Glu bind Ca\(^{2+}\). Usually 6 oxygens arranged in a pocket.

- Charged residues generally on the outside and form salt links (H-bonds with oppositely charged groups). Rare on the inside - if present, almost always paired.

Aspartate (Asp, D)  pK=4.4

Glutamate (Glu, E)  pK=4.4

Asp: Common helix starter and turn-forming residue. Glu is also a turn-forming residue, as are all hydrophilic residues, but is less position-specific than Asp.

Asp and Glu are more frequent in the first turn of alpha-helices, whereas Lys, Arg, and His are more common toward the C-terminus. Those positions promote interaction with the helix dipole.

Asp and Glu bind Ca\(^{2+}\). Usually 6 oxygens arranged in a pocket.

Charged residues generally on the outside and form salt links (H-bonds with oppositely charged groups). Rare on the inside - if present, almost always paired.
Basic Amino Acids: Lys and Arg

- Positively charged and on the protein surface with rare exceptions.
- Form salt links with negatively charged macromolecules, such as nucleic acids. Histones are rich in Lys and Arg.
- Often present at active sites. Lys more often participates in binding, and Arg more often in catalysis.
- Lys side chain is highly mobile, thus often invisible on crystal-structure electron-density maps. Interacts well with solvent water. Provides solubility for many globular proteins.
- Arg side chains buried more often than Lys, on average, but rarely totally. Arg side chains usually make extensive van der Waals interactions, and they can curl around to produce a flat hydrophobic surface capable of conservatively replacing an Ile.
- Arg guanidinium group at the end has five hydrogen-bond donors held in a large, rigid, planar array. Often interact with water.
Imidazole Side Chain: His

- Unique - pK near neutrality, so it can gain or lose its positive charge by changes in surroundings.
- Major roles in active sites or as a controllable element in conformational changes.
- Can be buried, allowing it to bind the buried charge of a reactive group.

Histidine (His, H)

\[ \text{Histidine (His, H)} \]
\[ \text{pK}=6.5 \]
Valine (Val, V)

- Important in protein folding.
- Excluded from the surface. Variety of size and shape allows good fit for hydrophobic interior.
- Met - sulfur may have some H-bonding capability.
- Can be on surface in transmembrane segments to interact with membrane lipids.

Leucine (Leu, L)

Isoleucine (Ile, I)

Methionine (Met, M)
Aromatic Side Chains

• Contribute to Absorbance at 280 nm
• Phe is completely hydrophobic.
• Tyr and Trp have a single H-bond, which has only a small net effect on the hydrophobicity of Trp but makes Tyr almost indifferent to inside vs. outside.
• Major constituents of the hydrophobic core.
• Tyr can undergo sulfation and phosphorylation.
Glycine and Alanine

**Glycine (Gly, G)**
- Smallest side chain.
- Symmetrical and broader region of the $\phi$ and $\psi$ plot is accessible.
- Found in bends and turns where side chain cannot fit.
- Facilitates movement at hinge regions.

**Alanine (Ala, A)**
- Default amino acid: short side chain, no chemical reactivity, no unusual conformational properties, and fairly happy on interior or surface.
- Prefers $\alpha$-helices: most stable compact H-bonded structure. Strongest preference for a middle-helix location.
- Turns prefer more hydrophilic residues, and nonrepetitive structure in general favors amino acids with side-chain H-binding capability. A $\beta$-sheet prefers to be covered by larger side chains.
Proline

- Constraint on backbone. Ring keeps $\phi$ near -60$^\circ$.
- Prohibited in middle of helices, but often starts them, at position 1.
- Cannot fit into the regular, internal parts of beta-sheet.
- Good for making turns - has highest specific positional preference of any residue.
- Occurrence in turns and virtual absence from the interior of regular secondary structures, puts Pro as almost always on the surface.
- Can occur with a cis rather than trans peptide bond. Sequence preferences for cis versus trans prolines. The positions immediately before the cis-Pro is heavily enriched in F, Y, and L and very unfavorable for branched side chains. Low frequency of charge residues before and after cis-Pro.
- Affects protein stability based on entropy effects. One degree of freedom is missing, so Pro has a smaller loss of entropy on folding than any other residue.
Cysteine

- Can bind metal ions or form disulfides
- Free SH groups are relatively uncommon. Cys is usually buried in proteins, probably to protect the reactive side chain group.
- Disulfides: Cysteines must be between 4 - 7.5 Å apart. Difficult to predict which Cys residues will pair.
- Cannot form disulfides between 2 Cys residues in the same α-helix w/o distortion.
- Disulfides are important to stabilize protein structure, but have little role in folding, when are their reduced state.
Three Kinds of Bonds that Affect Folding

electrostatic attractions

hydrogen bond

van der Waals attractions
Hydrogen Bonds

- Sharing of a proton between donor and acceptor groups
- Strength ~ 2-5 kcal/mol
- Distance ~ 2.8-3 Å
Salt Bridges

- Electrostatic bonds: Oppositely charged groups
- 4-7 kcal/mol
van der Waals Interactions

- Electrostatic interactions cannot account for all the non-covalent interactions observed between molecules (especially uncharged ones)

- Atoms with dipoles (and higher order multipoles) induce and interact with dipoles in other atoms via dispersion forces
Importance of Hydrophobic AAs

- **Unfolded polypeptide**
- **Folded conformation in aqueous environment**

**Diagram:**
- Polar side chains
- Nonpolar side chains
- Hydrophobic core region contains nonpolar side chains
- Polar side chains on the outside of the molecule can form hydrogen bonds to water
Hydrophobic Interactions

Water not able to form H-bonds with certain side chains
Folding of the Polypeptide Chain into Secondary and Tertiary Structures

(Primary Structure is the Sequence)
Bond angles and resonance in the peptide bond keep 6 atoms in a plane.
Frequencies of Angles for phi and psi in a Given Protein

Values in this region make an α-helix.
Examples of secondary structure: the α-helix

- amino acid side chain
- oxygen
- H-bond
- carbon
- hydrogen
- nitrogen

α helix

0.54 nm

(C)
Examples of secondary structure: the β-sheet
Two kinds of $\beta$-sheet: Antiparallel and Parallel Strand
Tertiary Structure: Structured Domains can interact to define structure at a higher level

Atomic structure: Red is hydrophobic
Ribbon models of three common folds

- **Cluster of α-helices**: Cytochrome b<sub>562</sub>
- **Mixed α and β**: Part of lactate dehydrogenase
- **All β**: Variable domain of immunoglobulin light chain
One polypeptide can fold into multiple domains: Src protein kinase
Multiple polypeptides can fold together into multiple domains

- F-box protein (substrate-binding arm)
- E2 ubiquitin-conjugating enzyme
- adaptor protein 2 (Skp1)
- adaptor protein 1
- scaffold protein (cullin)
Homologous proteins in different organisms often have similar sequences and nearly identical folds.

Sequences and folds for similar DNA-binding Proteins from yeast and fruit fly.
Regions of similar sequence can be used to seek out related proteins across a wide gap in phylogeny.
Sometimes domains of conserved structure are shuffled and expressed in otherwise unrelated proteins.
Proteins with domains of similar structure are thought to have related functions.

Three examples of DNA-binding proteins:

- **yeast**: Ep1, PHD, PHD, Ep2
- **worm**: Ep1, PHD, PHD, Ep2, Br
- **human**: Znf, Ep1, PHD, PHD, Ep2, Br, BMB
Proteins can form by assembly of more than one polypeptide chain:

Cro repressor protein
Hemoglobin: well-studied multi-subunit protein
Protein fold defines binding sites on the protein surface
Cytoskeleton protein actin is one example.
Extracellular protein collagen is another
Capsid proteins of viruses are another
Schematic of virus coat assembly
Tobacco Mosaic Virus: Protein and RNA assemble to make a virion
How much does one bond affect the binding constant?

The relationship between free-energy differences and equilibrium constants (37°C)

\[
\frac{[AB]}{[A][B]} = K
\]

<table>
<thead>
<tr>
<th>equilibrium constant</th>
<th>free-energy difference of AB minus free energy of A + B (kcal/mole)</th>
<th>free-energy difference of AB minus free energy of A + B (kJ/mole)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>-1.4</td>
<td>-5.9</td>
</tr>
<tr>
<td>(10^2)</td>
<td>-2.8</td>
<td>-11.9</td>
</tr>
<tr>
<td>(10^3)</td>
<td>-4.3</td>
<td>-17.8</td>
</tr>
<tr>
<td>(10^4)</td>
<td>-5.7</td>
<td>-23.7</td>
</tr>
<tr>
<td>(10^5)</td>
<td>-7.1</td>
<td>-29.7</td>
</tr>
<tr>
<td>(10^6)</td>
<td>-8.5</td>
<td>-35.6</td>
</tr>
<tr>
<td>(10^7)</td>
<td>-9.9</td>
<td>-41.5</td>
</tr>
<tr>
<td>(10^8)</td>
<td>-11.3</td>
<td>-47.4</td>
</tr>
<tr>
<td>(10^9)</td>
<td>-12.8</td>
<td>-53.4</td>
</tr>
<tr>
<td>(10^{10})</td>
<td>-14.2</td>
<td>-59.4</td>
</tr>
<tr>
<td>(10^{11})</td>
<td>-15.6</td>
<td>-65.3</td>
</tr>
</tbody>
</table>
End