Amino Acids and Proteins

Amino acid (1)

Amino acid (2)

Peptide bond

Dipeptide

Water

Dr. Anita S. Goswami-Giri
Department of Chemistry
VPM’s B. N. Bandodkar College of Science, Thane-1
anitagoswami@yahoo.com
Amino Acids: Building Blocks of Protein

• Proteins are heteropolymers of $\alpha$-amino acids
• Amino acids have following biological functions:
  – Capacity to polymerize
  – Useful acid-base properties
  – Varied physical properties
  – Varied chemical functionality

General Structure of $\alpha$-amino acids
Most $\alpha$-Amino Acids are Chiral

- The $\alpha$-carbon has always four substituents and is tetrahedral.
- All (except proline) have an acidic carboxyl group, a basic amino group, and an alpha hydrogen connected to the $\alpha$-carbon.
- Each amino acid has an unique fourth substituent $R$.
- In glycine, the fourth substituent is also hydrogen.
# Amino Acids: Classification

Common amino acids can be placed in five basic groups depending on their R substituent's:

- **Nonpolar, aliphatic** - (7) - (Ala, Val, Ile, Leu, Met)
- **Aromatic** - (3) - (Phe, Tyr, Typ)
- **Polar, uncharged** - (5) - (Ser, Thr, Asn, Gln, Cys)
- **Positively charged** - (3) - (Arg, His, Lys)
- **Negatively charged** - (2) - (Asp, Glu)
Aliphatic Amino Acids

D. Amino Acids with Hydrophobic Side Chain

Alanine (Ala)  Valine (Val)  Isoleucine (Ile)  Leucine (Leu)  Methionine (Met)
Aromatic Amino Acids

Phenylalanine (Phe)  Tyrosine (Tyr)  Tryptophan (Trp)

\[ \text{HO} \quad \text{HO} \quad \text{HO} \]
\[ \text{O} \quad \text{O} \quad \text{O} \]
\[ \text{NH}_2 \quad \text{NH}_2 \quad \text{NH}_2 \]
\[ \text{F} \quad \text{Y} \quad \text{W} \]

\[ \text{pKa 10.10} \]
Charged Amino Acids

A. Amino Acids with Electrically Charged Side Chains

Positive

Arginine (Arg)  
R  
pKa 2.03

Histidine (His)  
H  
pKa 1.70

Lysine (Lys)  
K  
pKa 2.15

Negative

Aspartic Acid (Asp)  
D  
pKa 1.95

Glutamic Acid (Glu)  
E  
pKa 2.16

pKa 9.00

pKa 9.09

pKa 9.16

pKa 9.66

pKa 9.58

pKa 6.04

pKa 10.67

pKa 3.71

pKa 4.15

pKa 2.10
Polar Amino Acids

B. Amino Acids with Polar Uncharged Side Chains

Serine (Ser) \[ S \]  
\[
\text{HO} \quad \text{O} \quad \text{NH}_2 \\
\text{OH} \quad \text{pKa} \ 9.05
\]
\[
\text{pKa} \ 2.13
\]

Threonine (Thr) \[ T \]  
\[
\text{HO} \quad \text{O} \quad \text{NH}_2 \\
\text{HO} \quad \text{pKa} \ 8.96
\]
\[
\text{pKa} \ 2.20
\]

Asparagine (Asn) \[ N \]  
\[
\text{HO} \quad \text{O} \quad \text{NH}_2 \\
\text{NH}_2 \quad \text{O} \quad \text{pKa} \ 8.76
\]
\[
\text{pKa} \ 2.16
\]

Glutamine (Gln) \[ Q \]  
\[
\text{HO} \quad \text{O} \quad \text{NH}_2 \\
\text{O} \quad \text{NH}_2 \\
\text{pKa} \ 9.00
\]
\[
\text{pKa} \ 2.18
\]
Special Amino Acids

Glycine (Gly)  \[\text{HO-CONH}_2\]  pKa 2.34

Proline (Pro)  \[\text{HO-CONH} \_\text{C}_2\text{H}_5\]  pKa 1.95

Cysteine (Cys)  \[\text{HO-CONH-SH}\]  pKa 1.91

Selenocysteine (Sec)  \[\text{HO-CONH-SeH}\]  pKa 1.9
# 20 amino acids and its side chain R-group

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>1ltr</th>
<th>3ltr</th>
<th>Side chain</th>
<th>NonPolar</th>
<th>pKa</th>
<th>Polar</th>
<th>charge class</th>
<th>Small</th>
<th>Tiny</th>
<th>Aromatic</th>
<th>vol. (vdW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>G</td>
<td>Gly</td>
<td>-H</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>48</td>
</tr>
<tr>
<td>Alanine</td>
<td>A</td>
<td>Ala</td>
<td>-CH₃</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>67</td>
</tr>
<tr>
<td>Serine</td>
<td>S</td>
<td>Ser</td>
<td>-CH₂OH</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>73</td>
</tr>
<tr>
<td>Cysteine</td>
<td>C</td>
<td>Cys</td>
<td>-CH₂SH</td>
<td>X</td>
<td>8.18</td>
<td>-</td>
<td>acidic</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>86</td>
</tr>
<tr>
<td>Proline</td>
<td>P</td>
<td>Pro</td>
<td>CH₂CH₂CH₂-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>90</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>D</td>
<td>Asp</td>
<td>-CH₂COOH</td>
<td>-</td>
<td>3.9</td>
<td>X</td>
<td>acidic</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>91</td>
</tr>
<tr>
<td>Threonine</td>
<td>T</td>
<td>Thr</td>
<td>-CH(OH)CH₃</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>weak acidic</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>93</td>
</tr>
<tr>
<td>Asparagine</td>
<td>N</td>
<td>Asn</td>
<td>-CH₂CONH₂</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>96</td>
</tr>
<tr>
<td>Valine</td>
<td>V</td>
<td>Val</td>
<td>-CH(CH₃)₂</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>Aliphatic</td>
<td>105</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>E</td>
<td>Glu</td>
<td>CH₂CH₂COO⁻</td>
<td>-</td>
<td>4.07</td>
<td>X</td>
<td>acidic</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>109</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Q</td>
<td>Gln</td>
<td>CH₂CH₂CON⁻</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>114</td>
</tr>
<tr>
<td>Histidine</td>
<td>H</td>
<td>His</td>
<td>-CH₂-C₃H₃N₂</td>
<td>6.04</td>
<td>X</td>
<td>-</td>
<td>weak basic</td>
<td>-</td>
<td>-</td>
<td>Aromatic</td>
<td>118</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>I</td>
<td>Ile</td>
<td>-CH(CH₃)CH₂</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Aliphatic</td>
<td>124</td>
</tr>
<tr>
<td>Leucine</td>
<td>L</td>
<td>Leu</td>
<td>-CH₂CH(CH₃)₂</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Aliphatic</td>
<td>124</td>
</tr>
<tr>
<td>Methionine</td>
<td>M</td>
<td>Met</td>
<td>CH₂CH₂SCH₃</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>124</td>
</tr>
<tr>
<td>Lysine</td>
<td>K</td>
<td>Lys</td>
<td>-(CH₂)₄NH₂</td>
<td>10.54</td>
<td>X</td>
<td>-</td>
<td>basic</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>135</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>F</td>
<td>Phe</td>
<td>-CH₂C₆H₅</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Aromatic</td>
<td>135</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Y</td>
<td>Tyr</td>
<td>-CH₂C₆H₄OH</td>
<td>10.46</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Aromatic</td>
<td>141</td>
</tr>
<tr>
<td>Arginine</td>
<td>R</td>
<td>Arg</td>
<td>-(CH₂)₃NH-C(NH)₂</td>
<td>12.48</td>
<td>X</td>
<td>-</td>
<td>strongly basic</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>148</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>W</td>
<td>Trp</td>
<td>-CH₂C₈H₆N</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Aromatic</td>
<td>163</td>
</tr>
</tbody>
</table>
Strecker synthesis of amino acids

- The condensation reaction of an aldehyde with ammonium chloride in the presence of cyanide yields an α-aminonitrile. After hydrolyzed, it gives the desired amino acid. The method is used commercially for the production of racemic mixture of methionine from methional.
**Reaction mechanism**

**Step 1**
The carbonyl oxygen of an aldehyde is protonated, followed by a nucleophilic attack of ammonia to the carbonyl carbon. After subsequent proton exchange, water is cleaved from the iminium ion intermediate.

**Step 2**
A cyanide ion then attacks the iminium carbon yielding an aminonitrile.
**Step 3**
The nitrile nitrogen of the aminonitrile is protonated, and the nitrile carbon is attacked by a water molecule. A 1,2-diamino-diol is then formed after proton exchange and a nucleophilic attack of water to the former nitrile carbon.

**Step 4**
Ammonia is subsequently eliminated after the protonation of the amino group, and finally the deprotonation of a hydroxyl group produces an amino acid.
Gabriel phthalamide synthesis.

Step 1

Potassium phthalamide is a \( \cdot \text{NH}_2 \)-synthon which allows the preparation of primary amines by reaction with alkyl halides.
Mechanism of the Gabriel Synthesis

Step 1
Potassium phthalimide is a $\text{NH}_2$-synthon which allows the preparation of primary amines by reaction with alkyl halides.

Step 2
Product is cleaved by reaction with base or hydrazine, which leads to a stable cyclic product. After alkylation, the phthalimide is not nucleophile and does not react anymore.

Step 3
Product is cleaved by reaction with base or hydrazine, which leads to a stable cyclic product.
1.2.2 Peptides and Peptide bonds

Peptide bond in a di-peptide

“Peptides” are small condensation products of amino acids

They are “small” compared to proteins (di, tri, tetra… oligo-)

The bond length (C=O-N-H) of peptide is 10% shorter than that of C-N amine bonds because in peptide bond shows double bond characteristic (40%) due to resonance as consequence the shape of molecule is planner and reduce the rigidity of polypeptide
Peptide Ends are Not the Same

Numbering starts from the amino terminus

$AA_1$  $AA_2$  $AA_3$  $AA_4$  $AA_5$

Amino-terminal end

Carboxyl-terminal end

LENINGER PRINCIPLE OF BIOCHEMISTRY 5TH EDITION 2008
Neutral

Nonpolar, Aliphatic R Groups

Glycine

Alanine

Proline

Valine

Leucine

Isoleucine

Methionine
Aromatic amino acids (R Groups)

Also Hydrophobic

These amino acid side chains absorb UV light at 270-280nm
Polar, Uncharged R Groups

These amino acids side chains can form hydrogen bonding

Cysteine can form disulfide bonds
Basic R Groups

Positively charged R groups

Lysine
Arginine
Histidine
Acidic R Groups

Negatively charged R groups

Aspartate

Glutamate
Merrifield solid phase polypeptide synthesis

R = NH-NH_Fmoc or SO_2-NH_2

SPPS
Merrifield solid phase polypeptide synthesis
Merrifield solid phase polypeptide synthesis

\[ \text{NHFmoc} \quad \text{Ph(OMe)}_2 \quad \xrightarrow{\text{SPPS}} \quad \text{Gly-Asp-Asn-Asn-Ser-Asn-Asp-Gly-Trp-NH} \]

\[ + \quad \text{N}_2\text{-PEGs} \]

\[ \xrightarrow{\text{PEGylation}} \quad \text{PEGylated lipopeptides} \]
For the synthesis of peptides, Bruce Merrifield was awarded the (1984) Nobel Prize in chemistry. In *Merrifield Synthesis*, solid-phase polymer support is used. It is multistep reaction having significant length of residues and the product must be carefully purified after each step to prevent unwanted cross-reactions. To facilitate the tedious and time consuming purifications, and reduce the material losses that occur in handling, a clever modification of this strategy has been developed. This procedure, known as the *Merrifield Synthesis*.

After its inventor R. Bruce Merrifield, involves attaching the C-terminus of the peptide chain to a polymeric solid, usually having the form of very small beads. Separation and purification is simply accomplished by filtering and washing the beads with appropriate solvents. The reagents for the next peptide bond addition are then added, and the purification steps repeated.
The final step, in which the completed peptide is released from the polymer support, is a simple benzyl ester cleavage. A series of equations of the Merrifield synthesis is -

Merrifield Synthesis

Step 1: 

\[
\text{BOC-NH} \quad \text{Esterification} \quad \text{BOC-NH} \quad \text{Deprotect} \quad \text{NH}_2
\]

Step 2: 

\[
\text{BOC-NH} \quad \text{DCC} \quad \text{Deprotect} \quad \text{NH}_2
\]

Step 3: 

\[
\text{BOC-NH} \quad \text{1) DCC} \quad \text{2) Deprotect} \quad \text{NH}_2
\]

Step 4: 

\[
\text{BOC-NH} \quad \text{1) DCC} \quad \text{2) Deprotect} \quad \text{NH}_2
\]
Proteins: general idea of primary, secondary, tertiary & quaternary structure

Primary Structure of Protein - The bond length (C=O-N-H) of peptide is 10% shorter than that of C-N amine bonds because in peptide bond shows double bond characteristic (40%) due to resonance as consequence the shape of molecule is planner and reduce the rigidity of polypeptide. 3.6 amino acid residues per turn (36 amino acid per long 10 turn)
General idea of primary, secondary of protein

<table>
<thead>
<tr>
<th>Primary Structure</th>
<th>Linear sequence of amino acids</th>
<th>Peptide bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>R group peptide bond</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Secondary structure</th>
<th>Localized organization of the parts of the polypeptide chain: α - helix, β - pleated sheath</th>
<th>Backbone Hydrogen bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>e.g α-keratin</td>
<td>Per turn of the alpha helix containing 3.6 amino acids. e.g myoglobin and haemoglobin, consists mainly of alpha helices</td>
</tr>
</tbody>
</table>

β-Sheet

α-helix
General idea of tertiary & quaternary structure of protein

<table>
<thead>
<tr>
<th>Tertiary structure</th>
<th>Overall three dimensional arrangement of the polypeptide chain</th>
<th>Hydrophobic interactions, Hydrogen bonds (non covalent bonds in general) and sulphur bridges.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Secondary structure and loops come together to smallest tertiary structure unit. (100-200 amino acids). The bond interaction includes disulphide bridges and non covalent interactions like ionic bonds, hydrogen bonds, van der waals and hydrophobic interactions.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quaternary structure</th>
<th>The association of two or more polypeptides into a multi – subunit complex</th>
<th>Protein to have quaternary structure – it should form more than one polypeptide chain and these chain cannot be attached by covalent bond among them</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemoglobin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>is made up of 4 protein molecules 2α &amp; 2β subunits</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Five factors are critical for protein folding and stability:

1. **Hydrogen bonds**: Hydrogen bonding occurs between the backbone of the amine group and the oxygen of the carbonyl group in case of amino acids.

2. **Ionic bonds**: Electrostatic interactions occur between two oppositely charged molecules. Ionic interactions are weaker in water than in vacuum, this is due to a different dielectric constant faced in water between opposing charges within the protein's structure.

3. **Hydrophobic effect**: The hydrophobic interaction originates from the tendency of non-polar molecules to minimize their interactions with water. When non-polar molecules interact with water, these molecules tend to cluster together in the center to form a micelle.

4. **Van der Waals forces**: Van der Waals forces exist between non-polar molecules at close range. Of the three van der Waals interactions, interactions between permanent dipoles is the strongest, dipole-induced dipole interactions are weaker than permanent dipole and the London dispersion forces are the weakest. While van der Waals forces between individual atoms are weak, the sum of van der Waals forces resulting from interactions between many atoms in large macromolecules can be substantial. The strength of van der Waals interactions varies with the distance between the atoms and is maximal at the van der Waals contact distance.

5. **Disulfide bridges**: A disulfide bond can be form between two cysteines through oxidation. These are also the strongest covalent bonds within a protein's tertiary structure.
Thank you