Amino Acids, Peptides & Proteins

- Amino Acids
- Zwitterion or ampholyte
- Peptides
- Proteins
- Structure of Proteins
- Separation & purification
- Protein Sequencing
- Protein Synthesis
Molecular network system in a cell

Our Life Is Maintained by Molecular Network Systems

Proteins Play Key Roles in a Living System

- Three examples of protein functions
  - Catalysis: Almost all chemical reactions in a living cell are catalyzed by protein enzymes.
  - Transport: Some proteins transport various substances, such as oxygen, ions, and so on.
  - Information transfer: For example, hormones.

Alcohol dehydrogenase oxidizes alcohols to aldehydes or ketones
Haemoglobin carries oxygen
Insulin controls the amount of sugar in the blood
Estimated Functional Roles
(by % of Proteins) of the Proteome in a Complex Organism

Protein Functions

- Transport
- Regulatory
- Motor
- Fibrous Protein
- Enzyme
- Immunoglobulin
20 Amino acids

- Glycine (G)
- Alanine (A)
- Valine (V)
- Isoleucine (I)
- Leucine (L)
- Proline (P)
- Methionine (M)
- Phenylalanine (F)
- Tryptophan (W)
- Asparagine (N)
- Glutamine (Q)
- Serine (S)
- Threonine (T)
- Tyrosine (Y)
- Cysteine (C)
- Aspartic acid (D)
- Glutamic acid (E)
- Lysine (K)
- Arginine (R)
- Histidine (H)

White: Hydrophobic, Green: Hydrophilic, Red: Acidic, Blue: Basic

α-amino acids

- All proteins are composed of amino acids
- Twenty common amino acids
- All amino acids are primary amino acids except for proline
- A primary amine group is attached to the α-carbon of carboxyl group
- Except for Glycine, all other amino acids have at least one chiral centers
- All chiral amino acids are belong to L-amino acids
Amino acids vary in

Size
Structure
Electric charge
Solubility in water
Classification of amino acids

Classified by polarity of side chains
hydrophobic: water fearing, non-polar side chains
hydrophilic: water loving, polar neutral
positively charged
negatively charged
aromatic

Nonpolar, aliphatic R groups

Glycine  Alanine  Proline  Valine

Leucine  Isoleucine  Methionine
Aromatic R groups

Phenylalanine  Tyrosine  Tryptophan

Polar, uncharged R groups

Serine  Threonine  Cysteine

Asparagine  Glutamine
Positively charged R groups

Lysine

Arginine

Histidine

Negatively charged R groups

Aspartate

Glutamate
Figure 2.6
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company

Box 3.1 figure 1
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company
Titration curve of amino acids

Neutral side chain

Acidic side chain

Basic side chain

\[
\begin{align*}
\text{Nonionic form:} & & \text{Zwitterionic form:} \\
\text{Zwitterion as acid:} & & \text{Zwitterion as base:}
\end{align*}
\]
### Figure 3.11

<table>
<thead>
<tr>
<th>Methyl-substituted carboxyl and amino groups</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Diagram" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Carboxyl and amino groups in glycine</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image2.png" alt="Diagram" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pK\textsubscript{a}</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetic acid</td>
<td>CH\textsubscript{3} - COOH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methylamine</td>
<td>CH\textsubscript{3} - NH\textsubscript{2}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The normal pK\textsubscript{a} for a carboxyl group is about 4.8.
The normal pK\textsubscript{a} for an amino group is about 10.6.

Repulsion between the amino group and the departing proton lowers the pK\textsubscript{a} for the carboxyl group, and oppositely charged groups lower the pK\textsubscript{a} by stabilizing the zwitterion.

Electronegative oxygen atoms in the carboxyl group pull electrons away from the amino group, lowering its pK\textsubscript{a}.

### Figure 3.12a

![Graph](image3.png)

Net charge: +1 0 -1 -2

Glutamate:
- pK\textsubscript{1} = 2.19
- pK\textsubscript{2} = 9.67
- pK\textsubscript{A} = 4.25
**TABLE 3–1** Properties and Conventions Associated with the Common Amino Acids Found in Proteins

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Abbreviation/symbol</th>
<th>$M_r$</th>
<th>$pK_a^{1}$ (---COOH)</th>
<th>$pK_a^{2}$ (---NH$_2^+$)</th>
<th>$pK_a$ (R group)</th>
<th>$pI$</th>
<th>Hydrophathy index*</th>
<th>Occurrence in proteins (%)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonpolar, aliphatic R groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly G</td>
<td>75</td>
<td>2.34</td>
<td>9.60</td>
<td>5.97</td>
<td>−0.4</td>
<td>7.2</td>
<td></td>
</tr>
<tr>
<td>Alanine</td>
<td>Ala A</td>
<td>89</td>
<td>2.34</td>
<td>9.69</td>
<td>6.01</td>
<td>1.8</td>
<td>7.8</td>
<td></td>
</tr>
<tr>
<td>Proline</td>
<td>Pro P</td>
<td>115</td>
<td>1.99</td>
<td>10.96</td>
<td>6.48</td>
<td>1.6</td>
<td>5.2</td>
<td></td>
</tr>
<tr>
<td>Valine</td>
<td>Val V</td>
<td>117</td>
<td>2.32</td>
<td>9.62</td>
<td>5.97</td>
<td>4.2</td>
<td>6.6</td>
<td></td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu L</td>
<td>131</td>
<td>2.36</td>
<td>9.60</td>
<td>5.98</td>
<td>3.8</td>
<td>9.1</td>
<td></td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile I</td>
<td>131</td>
<td>2.36</td>
<td>9.68</td>
<td>6.02</td>
<td>4.5</td>
<td>5.3</td>
<td></td>
</tr>
<tr>
<td>Methionine</td>
<td>Met M</td>
<td>149</td>
<td>2.28</td>
<td>9.21</td>
<td>5.74</td>
<td>1.9</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>Aromatic R groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe F</td>
<td>165</td>
<td>1.83</td>
<td>9.13</td>
<td>5.48</td>
<td>2.8</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr Y</td>
<td>181</td>
<td>2.20</td>
<td>9.11</td>
<td>10.07</td>
<td>5.66</td>
<td>−1.3</td>
<td>3.2</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp W</td>
<td>204</td>
<td>2.38</td>
<td>9.30</td>
<td>5.89</td>
<td>−0.9</td>
<td>1.4</td>
<td></td>
</tr>
</tbody>
</table>


### TABLE 3-1 Properties and Conventions Associated with the Common Amino Acids Found in Proteins

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Abbreviation/ symbol</th>
<th>$M_r$</th>
<th>$pK_a$ (−COOH)</th>
<th>$pK_a$ (−NH$_2^-$)</th>
<th>$pK_a$ (R group)</th>
<th>$pI$</th>
<th>Hydropathy index*</th>
<th>Occurrence in proteins (%)$^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polar, uncharged R groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serine</td>
<td>Ser S</td>
<td>105</td>
<td>2.11</td>
<td>9.15</td>
<td>5.68</td>
<td>−0.8</td>
<td>−0.48</td>
<td>6.8</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr T</td>
<td>119</td>
<td>2.11</td>
<td>9.62</td>
<td>5.87</td>
<td>−0.7</td>
<td>−0.89</td>
<td>5.9</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys C</td>
<td>121</td>
<td>1.96</td>
<td>10.28</td>
<td>8.18</td>
<td>5.07</td>
<td>2.5</td>
<td>1.9</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn N</td>
<td>132</td>
<td>2.02</td>
<td>8.80</td>
<td>5.41</td>
<td>−3.5</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln Q</td>
<td>146</td>
<td>2.17</td>
<td>9.13</td>
<td>5.65</td>
<td>−3.5</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>Positively charged R groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys K</td>
<td>146</td>
<td>2.18</td>
<td>8.95</td>
<td>10.53</td>
<td>9.74</td>
<td>−3.9</td>
<td>5.9</td>
</tr>
<tr>
<td>Histidine</td>
<td>His H</td>
<td>155</td>
<td>1.82</td>
<td>9.17</td>
<td>6.00</td>
<td>7.59</td>
<td>−3.2</td>
<td>2.3</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg R</td>
<td>174</td>
<td>2.17</td>
<td>9.04</td>
<td>12.48</td>
<td>10.76</td>
<td>−4.5</td>
<td>5.1</td>
</tr>
<tr>
<td>Negatively charged R groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aspartate</td>
<td>Asp D</td>
<td>133</td>
<td>1.88</td>
<td>9.60</td>
<td>3.65</td>
<td>2.77</td>
<td>−3.5</td>
<td>5.3</td>
</tr>
<tr>
<td>Glutamate</td>
<td>Glu E</td>
<td>147</td>
<td>2.19</td>
<td>9.67</td>
<td>4.25</td>
<td>3.22</td>
<td>−3.5</td>
<td>6.3</td>
</tr>
</tbody>
</table>

$^*$A scale combining hydrophobicity and hydrophilicity of R groups; it can be used to measure the tendency of an amino acid to seek an aqueous environment (− value) or a hydrophobic environment (+ value). See Chapter 11, from Kyte J, Doolittle, R.F. (1982) A simple method for displaying the hydrophobic character of a protein. J. Mol. Biol. 157, 105–132.


---

### Peptide Bond

![Peptide Bond](https://via.placeholder.com/150)
Peptides

- Dipeptide
- Tripeptide
- Tetrapeptide
- Oligopeptide
- Polypeptide
- Protein

Peptide Bond

\[
\begin{align*}
R_1 & \quad \text{N-terminal Amino Terminal} \\
\text{NH}_2 & \\
\text{C-terminal Carboxy Terminal} & \quad R_2 \\
\end{align*}
\]
Replace -ine by -yl
but keep the last -ine!

Ala-Gly-Arg  Alanylglycylarginine

Structure of Proteins

Primary structure
Secondary structure
Tertiary structure
Quaternary structure
TABLE 3–2 Molecular Data on Some Proteins

<table>
<thead>
<tr>
<th>Protein</th>
<th>Molecular weight</th>
<th>Number of residues</th>
<th>Number of polypeptide chains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytochrome c (human)</td>
<td>13,000</td>
<td>104</td>
<td>1</td>
</tr>
<tr>
<td>Ribonuclease A (bovine pancreas)</td>
<td>13,700</td>
<td>124</td>
<td>1</td>
</tr>
<tr>
<td>Lysozyme (chicken egg white)</td>
<td>13,930</td>
<td>129</td>
<td>1</td>
</tr>
<tr>
<td>Myoglobin (equine heart)</td>
<td>16,890</td>
<td>153</td>
<td>1</td>
</tr>
<tr>
<td>Chymotrypsin (bovine pancreas)</td>
<td>21,600</td>
<td>241</td>
<td>3</td>
</tr>
<tr>
<td>Chymotrypsinogen (bovine)</td>
<td>22,000</td>
<td>245</td>
<td>1</td>
</tr>
<tr>
<td>Hemoglobin (human)</td>
<td>64,500</td>
<td>574</td>
<td>4</td>
</tr>
<tr>
<td>Serum albumin (human)</td>
<td>68,500</td>
<td>609</td>
<td>1</td>
</tr>
<tr>
<td>Hexokinase (yeast)</td>
<td>102,000</td>
<td>972</td>
<td>2</td>
</tr>
<tr>
<td>RNA polymerase (E. coli)</td>
<td>450,000</td>
<td>4,158</td>
<td>5</td>
</tr>
<tr>
<td>Apolipoprotein B (human)</td>
<td>513,000</td>
<td>4,536</td>
<td>1</td>
</tr>
<tr>
<td>Glutamine synthetase (E. coli)</td>
<td>619,000</td>
<td>5,628</td>
<td>12</td>
</tr>
<tr>
<td>Titin (human)</td>
<td>2,993,000</td>
<td>26,926</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3–2
Lehninger Principles of Biochemistry, Fifth Edition
© 2006 W.H. Freeman and Company
### TABLE 3–3  Amino Acid Composition of Two Proteins

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Cytosyme c</th>
<th>Cytosyme chymotrypsinogen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>Arg</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Asn</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Asp</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Cys</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Glu</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Gly</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>His</td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>Ile</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Leu</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>Lys</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>Met</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Phe</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Pro</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Ser</td>
<td>1</td>
<td>28</td>
</tr>
<tr>
<td>Thr</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>Trp</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Tyr</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Val</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>104</strong></td>
<td><strong>245</strong></td>
</tr>
</tbody>
</table>

*In some common analyses, such as acid hydrolysis, Arg and Lys are not readily distinguished from each other and an isotopic-labeled Arg or Lys (e.g., 15N-labeled Lys and 13C-labeled Arg) is used to quantify the contribution of the two amino acids.*

Table 3–3  Lehninger Principles of Biochemistry, Fifth Edition
© 2006 W.H. Freeman and Company

### TABLE 3–4  Conjugated Proteins

<table>
<thead>
<tr>
<th>Class</th>
<th>Prosthetic group</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipoproteins</td>
<td>Lipids</td>
<td>( \beta )-Lipoprotein of blood</td>
</tr>
<tr>
<td>Glycoproteins</td>
<td>Carbohydrates</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>Phosphoproteins</td>
<td>Phosphate groups</td>
<td>Casein of milk</td>
</tr>
<tr>
<td>Hemoproteins</td>
<td>Heme (iron porphyrin)</td>
<td>Hemoglobin</td>
</tr>
<tr>
<td>Flavoproteins</td>
<td>Flavin nucleotides</td>
<td>Succinate dehydrogenase</td>
</tr>
<tr>
<td>Metalloproteins</td>
<td>Iron</td>
<td>Ferritin</td>
</tr>
<tr>
<td></td>
<td>Zinc</td>
<td>Alcohol dehydrogenase</td>
</tr>
<tr>
<td></td>
<td>Calcium</td>
<td>Calmodulin</td>
</tr>
<tr>
<td></td>
<td>Molybdenum</td>
<td>Dinitrogenase</td>
</tr>
<tr>
<td></td>
<td>Copper</td>
<td>Plastocyanin</td>
</tr>
</tbody>
</table>

Table 3–4  Lehninger Principles of Biochemistry, Fifth Edition
© 2006 W.H. Freeman and Company
Peptide & Protein Charge

Arg—Ser—Gly—Asn—Gly—Phe—Pro—Lys—Met—Glu

\[ \text{pI} = ? \]

<table>
<thead>
<tr>
<th>Group</th>
<th>pKa</th>
<th>pH=1</th>
<th>pH=3</th>
<th>pH=7</th>
<th>pH=10</th>
<th>pH=11</th>
<th>pH=13</th>
</tr>
</thead>
<tbody>
<tr>
<td>-COOH</td>
<td>2.2</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-NH2</td>
<td>8.8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Glu</td>
<td>4.3</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lys</td>
<td>10.8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Arg</td>
<td>12.5</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Net Charge</td>
<td>+3</td>
<td>+2</td>
<td>+1</td>
<td>0</td>
<td>-1</td>
<td>-2</td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{pI} = (8.8 + 10.8)/2 = 9.8 \]
Protein Separation/Purification

- In general, proteins contain > 40 residues
  - Minimum needed to fold into tertiary structure
- Usually 100-1000 residues; percent of each AA varies
- Proteins separated based on differences in size and composition
- Proteins must be pure to analyze, determine structure/function
- Factors to control (to avoid denaturation or chemical degradation)
  - pH
  - Presence of enzymes
  - Temperature
  - Reactive thiol groups
  - Exposure to air, water

Methods of Separation/Purification

- Solubility (salts, solvents, pH, temperature)
- Chromatography
  - Ion exchange
  - Gel filtration
  - Affinity
- Electrophoresis
Figure 3-16
Lehrsturz Principles of Biochemistry, Fifth Edition
© 2006 W.H. Freeman and Company

Reservoir
Protein sample (mobile phase)
Solid porous matrix (stationary phase)
Porous support
Effluent

Proteins
A
B
C

Figure 3-17a
Lehrsturz Principles of Biochemistry, Fifth Edition
© 2006 W.H. Freeman and Company

Polymer beads with negatively charged functional groups
Protein mixture is added to column containing cation exchangers.
Proteins move through the column at rates determined by their net charge at the pH being used. With cation exchangers, proteins with a more negative net charge move faster and elute earlier.

Ion-exchange chromatography
**Size-exclusion chromatography**

Figure 3-17b
Lehninger Principles of Biochemistry, Fifth Edition
© 2000 W.H. Freeman and Company

**Affinity chromatography**

Figure 3-17c
Lehninger Principles of Biochemistry, Fifth Edition
© 2000 W.H. Freeman and Company
TABLE 3–5  A Purification Table for a Hypothetical Enzyme

<table>
<thead>
<tr>
<th>Procedure or step</th>
<th>Fraction volume (mL)</th>
<th>Total protein (mg)</th>
<th>Activity (units)</th>
<th>Specific activity (units/mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Crude cellular extract</td>
<td>1,400</td>
<td>10,000</td>
<td>100,000</td>
<td>10</td>
</tr>
<tr>
<td>2. Precipitation with ammonium sulfate</td>
<td>280</td>
<td>3,000</td>
<td>96,000</td>
<td>32</td>
</tr>
<tr>
<td>3. Ion-exchange chromatography</td>
<td>90</td>
<td>400</td>
<td>80,000</td>
<td>200</td>
</tr>
<tr>
<td>4. Size-exclusion chromatography</td>
<td>80</td>
<td>100</td>
<td>60,000</td>
<td>600</td>
</tr>
<tr>
<td>5. Affinity chromatography</td>
<td>6</td>
<td>3</td>
<td>45,000</td>
<td>15,000</td>
</tr>
</tbody>
</table>

Note: All data represent the status of the sample after the designated procedure has been carried out. Activity and specific activity are defined on page 91.

Table 3-5
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. W. Norton and Company

\[
\text{Na}^+ - \text{O} - \text{SO} - \text{O} - (\text{CH}_2)_{11}\text{CH}_3
\]

Sodium dodecyl sulfate (SDS)

Unnumbered 3 p8
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. W. Norton and Company
An ampholyte solution is incorporated into a gel.

A stable pH gradient is established in the gel after application of an electric field.

Protein solution is added and electric field is reapplied.

After staining, proteins are shown to be distributed along pH gradient according to their pI values.

**TABLE 3-6 The Isoelectric Points of Some Proteins**

<table>
<thead>
<tr>
<th>Protein</th>
<th>pI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pepsin</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Egg albumin</td>
<td>4.6</td>
</tr>
<tr>
<td>Serum albumin</td>
<td>4.9</td>
</tr>
<tr>
<td>Urease</td>
<td>5.0</td>
</tr>
<tr>
<td>β-Lactoglobulin</td>
<td>5.2</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>6.8</td>
</tr>
<tr>
<td>Myoglobin</td>
<td>7.0</td>
</tr>
<tr>
<td>Chymotrypsinogen</td>
<td>9.5</td>
</tr>
<tr>
<td>Cytochrome c</td>
<td>10.7</td>
</tr>
<tr>
<td>Lysozyme</td>
<td>11.0</td>
</tr>
</tbody>
</table>
Protein Sequencing

Traditional Chemical Method (Sanger Method)

Genetic Method

Proteomics
A Common Strategy for Protein Sequencing

Protein chemists follow a basic strategy when they attempt to determine the sequence of most proteins. This strategy is outlined below. Keep in mind that this strategy is only a guide, and should not inhibit your own ingenuity in solving the sequence of a protein.

**Determine the Amino Acid Composition**

In order to know which amino acids and how many of each amino acid there are in a polypeptide, we must break the peptide bonds. This can be accomplished with strong acids (i.e. 6N HCl) or strong bases or by exhaustive enzymatic digestion. By performing an acid hydrolysis or base hydrolysis experiment you obtain a minimum length for the polypeptide. The hydrolyzed amino acids are then grouped into different groups according to their pI values, and each amino acid is isolated by HPLC with pH gradient buffer, and determined by reacting with ninhydrin. The concentration of individual amino acid are normalized to integral number, and by comparing with the molecular weight, the actual number of individual amino acids within that particular protein can be determined.

**Hydrolysis**

- A polypeptide can be hydrolyzed by refluxing with 6M hydrochloric acid for 24h
- The individual amino acids can be separated from each other using a cation-exchange resin
  - An acidic solution of the amino acids is passed through the cation-exchange column; the strength of adsorption varies with the basicity of each amino acid (the most basic are held most strongly)
  - Washing the column with a sequence of buffered solutions causes the amino acids to move through it at different rates
• In the original method, the column eluant is treated with ninhydrin, a dye used for detecting and quantifying each amino acid as it comes off the column

\[
\begin{align*}
\text{Indane-1,2,3-trione} & \quad \text{Ninhydrin} \\
\end{align*}
\]

\[
\begin{align*}
\text{Zinc chloride} + \text{NH}_3 \quad \text{Purple anion}
\end{align*}
\]

• In modern practice, analysis of amino acid mixtures is routinely accomplished using high performance liquid chromatography (HPLC)

– Sanger N-Terminal Analysis
  • The N-terminal end of the polypeptide is labeled with 2,4-dinitrofluorobenzene and the polypeptide is hydrolyzed
    – The labeled N-terminal amino acid is separated from the mixture and identified

\[
\begin{align*}
\text{Labeled N-terminal amino acid} & \quad \text{Mixture of amino acids}
\end{align*}
\]

• The Sanger method is not as widely used as the Edman method
– C-Terminal Analysis

• Enzymes called carboxypeptidases hydrolyze C-terminal amino acids selectively
  – The enzyme continues to release each newly exposed C-terminal amino acid as the peptide is hydrolyzed; it is necessary to monitor the release of C-terminal amino acids as a function of time to identify them
• Primary Structure of Polypeptides and Proteins
  - The sequence of amino acids in a polypeptide is called its primary structure
    - Several methods exist to elucidate the primary structure of peptides
  - Edman Degradation
    - Edman degradation involves sequential cleavage and identification of N-terminal amino acids
    - Edman degradation works well for polypeptide sequence analyses up to approximately 60 amino acid residues
      - The N-terminal residue of the polypeptide reacts with phenyl isothiocyanate
      - The resulting phenylthiocarbamyl derivative is cleaved from the peptide chain
      - The unstable product rearranges to a stable phenylthiohydantoin (PTH) which is purified by HPLC and identified by comparison with PTH standards
    - Automated amino acid sequencing machines use the Edman degradation and high performance liquid chromatography (HPLC)
      - One Edman degradation cycle beginning with a picomolar amount of polypeptide can be completed in approximately 30 minutes
      - Each cycle results in identification of the next amino acid residue in the peptide
– Complete Sequence Analysis

- The Sanger and Edman methods of analysis apply to short polypeptide sequences (up to about 60 amino acid residues by Edman degradation)

- For large proteins and polypeptides, the sample is subjected to partial hydrolysis with dilute acid to give a random assortment of shorter polypeptides which are then analyzed
  - The smaller polypeptides are sequenced, and regions of overlap among them allow the entire polypeptide to be sequenced
Larger polypeptides can also be cleaved into smaller sequences using *site-specific* reagents and enzymes

- The use of these agents gives more predictable fragments which can again be overlapped to obtain the sequence of the entire polypeptide
- Cyanogen bromide (CNBr) cleaves peptide bonds only on the C-terminal side of methionine residues
  Mass spectrometry can be used to determine polypeptide and protein sequences

<table>
<thead>
<tr>
<th>Reagent (biological source)*</th>
<th>Cleavage points†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trypsin (bovine pancreas)</td>
<td>Lys, Arg (C)</td>
</tr>
<tr>
<td><em>Submaxillaris</em> protease (mouse submaxillary gland)</td>
<td>Arg (C)</td>
</tr>
<tr>
<td>Chymotrypsin (bovine pancreas)</td>
<td>Phe, Trp, Tyr (C)</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em> V8 protease (bacterium <em>S. aureus</em>)</td>
<td>Asp, Glu (C)</td>
</tr>
<tr>
<td>Asp-N-protease (bacterium <em>Pseudomonas fragi</em>)</td>
<td>Asp, Glu (N)</td>
</tr>
<tr>
<td>Pepsin (porcine stomach)</td>
<td>Leu, Phe, Trp, Tyr (N)</td>
</tr>
<tr>
<td>Endoproteinase Lys C (bacterium <em>Lysobacter enzymogenes</em>)</td>
<td>Lys (C)</td>
</tr>
<tr>
<td>Cyanogen bromide</td>
<td>Met (C)</td>
</tr>
</tbody>
</table>

*All reagents except cyanogen bromide are proteases. All are available from commercial sources.
†Residues furnishing the primary recognition point for the protease or reagent; peptide bond cleavage occurs on either the carbonyl (C) or the amino (N) side of the indicated amino acid residues.
Repeat Steps 3 and 4 to Determine Sub-sequences and Create Overlapping

The initial cleavage is generally made as specific as possible in order to generate large peptide fragments. It is easy to arrange fewer fragments. These fragments can be positioned relative to one another after treatment of the original polypeptide by a second cleavage procedure that generates fragments whose sequences extend across the initial cleavage points (referred to as overlapping peptides).

Locate the Disulfide Bonds

No primary structure analysis of a cysteine-containing protein can be regarded as complete before the presence and location of disulfide bonds has been established.

Reconstruct the Original Protein.

From the overlapping peptides and information gained from the original protein, a unique sequence for the protein or polypeptide of interest can be worked out.

**Figure 3-25**
Lehninger Principles of Biochemistry, Fifth Edition
© 2000 W.H. Freeman and Company
“Ladder sequencing” involves analyzing a polypeptide digest by mass spectrometry, wherein each polypeptide in the digest differs by one amino acid in length; the difference in mass between each adjacent peak indicates the amino acid that occupies that position in the sequence. Mass spectra of polypeptide fragments from a protein can be compared with databases of known polypeptide sequences, thus leading to an identification of the protein or a part of its sequence by matching.
The Blind Men and the Elephant

by John Godfrey Saxe

American poet John Godfrey Saxe (1816–1887) based the following poem on a fable which was told in India many years ago.

The Blind Men and the Elephant

It was six men of Indostan
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind
The First approached the Elephant,
And happening to fall
Against his broad and sturdy side,
At once began to bawl:
“God bless me! but the Elephant
Is very like a wall!”

The Second, feeling of the tusk,
Cried, “Ho! what have we here
So very round and smooth and sharp?
To me ‘tis very clear
This wonder of an Elephant
Is very like a spear!”

The Third approached the animal,
And happening to take
The squirming trunk within his hands,
Thus boldly up and spake:
“I see,” quoth he, “the Elephant
Is very like a snake!”

The Fourth reached out an eager hand,
And felt about the knee.
“What most this wondrous beast is like
Is mighty plain,” quoth he;
“Tis clear enough the Elephant
Is very like a tree!”

The Fifth, who chanced to touch the ear,
Said: “E’en the blindest man
Can tell what this resembles most;
Deny the fact who can
This marvel of an Elephant
Is very like a fan!”

The Sixth no sooner had begun
About the beast to grope,
Than, seizing on the swinging tail
That fell within his scope,
“I see,” quoth he, “the Elephant
Is very like a rope!”

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!

A Fan
A Rope
A Tree
A Wall
A Spear
A Snake
Determine the sequence of protein from the following experimental details.

a. A peptide had the following amino acid composition obtained by acid hydrolysis and amino acid analysis: Arg, Phe, Glu, Asp, Lys, Met, Pro, Ser, His, Trp. N-terminal and C-terminal analyses of the peptide yielded no amino acids. The peptide strongly absorbed UV light at 280 nm.

b. Treatment of the original peptide with trypsin yield 2 peptides. T-1 had an amino acid composition after acid hydrolysis of Ser, Asp, Lys and yielded in Edman degradation Asp 1st and Ser 2nd. T-2 had amino acid composition after acid hydrolysis of His, Phe, Arg, Pro, Glu, Met, Trp and yielded in Edman degradation Met 1st, Glu 2nd, Trp 3rd and His 4th.

c. Treatment of the original peptide with chymotrypsin gave one large peptide with amino acid composition obtained by acid hydrolysis and amino acid analysis the same as the original peptide. Edman degradation yielded His 1st, Phe 2nd, Pro 3rd and Arg 4th.

d. Treatment of the original peptide with cyanogen bromide (CNBr) gave one large peptide with amino acid composition obtained by acid hydrolysis and amino acid analysis the same as the original peptide, except there was no Met and it contained an unknown amino acid. Edman degradation yielded Glu 1st, Trp 2nd, His 3rd and Phe 4th.
<table>
<thead>
<tr>
<th>Number of residues in the final polypeptide</th>
<th>Overall yield of final peptide (%) when the yield of each step is:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>96.0%</td>
</tr>
<tr>
<td>11</td>
<td>64</td>
</tr>
<tr>
<td>21</td>
<td>42</td>
</tr>
<tr>
<td>31</td>
<td>28</td>
</tr>
<tr>
<td>51</td>
<td>12</td>
</tr>
<tr>
<td>100</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Table 3-8
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company