Chapter 24
Proteins
Introduction

➔ The three major groups of biological polymers are polysaccharides, proteins and nucleic acids

➔ Proteins have many diverse functions; they are major components of the following biomolecules

 użytk Enzymes and hormones which catalyze and regulate biological reactions
 użytkow Muscles and tendons which provide the body with means for movement
 użytk Hemoglobin which carries oxygen to all parts of the body
 użytkow Antibodies they are integral parts of the immune system

➔ All proteins are polyamides

 użytk Their monomeric units are one of about 20 \( \alpha \)-amino acids

\[
\text{HO-} \alpha \text{NH}_2 \quad \text{R}
\]

An \( \alpha \)-amino acid

\( R \) is a side chain at the \( \alpha \) carbon that determines the identity of the amino acid (Table 24.1).

A portion of a protein molecule

Amide (peptide) linkages are shaded.

\( R_1 \sim R_5 \) may be any of the possible side chains.
Proteins have several levels of structure

- **Primary structure** refers to the exact sequence of amino acids along a protein chain.
- **Secondary** and **tertiary structures** refer to the further bending and folding of the primary structure.
- **Quaternary structure** refers to the aggregation of more than one polyamide chain.

All amino acids except glycine are chiral and have the L configuration (as related to glyceraldehyde) at the α carbon.

![Fischer projections for an L-α-amino acid and L-glyceraldehyde](image)
Amino acids

- Structure and Names
  - 22 amino acids but only 20 amino acids comprise the building blocks for synthesis of proteins
  - The remaining 2 amino acids are derived by modification after biosynthesis of the protein
    - Hydroxyproline and cystine are synthesized from proline and cysteine, respectively, after the protein chain has been synthesized
  - Cysteine is oxidized under mild conditions to the disulfide cystine
    - The reaction is reversible
    - This linkage is important in maintaining the overall shape of a protein

- Essential Amino Acids
  - Essential amino acids are not made by higher animals and must be part of the diet
    - There are 8 essential amino acids for adult humans (see Table 24.1)
<table>
<thead>
<tr>
<th>Structure of R</th>
<th>Name</th>
<th>Abbreviations&lt;sup&gt;a&lt;/sup&gt;</th>
<th>pK&lt;sub&gt;a1&lt;/sub&gt; α-CO&lt;sub&gt;2&lt;/sub&gt;H</th>
<th>pK&lt;sub&gt;a2&lt;/sub&gt; α-NH&lt;sub&gt;3&lt;/sub&gt;&lt;sup&gt;+&lt;/sup&gt;</th>
<th>pK&lt;sub&gt;a3&lt;/sub&gt; R group</th>
<th>pI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutal Amino Acids</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>—H</td>
<td>Glycine</td>
<td>G or Gly</td>
<td>2.3</td>
<td>9.6</td>
<td></td>
<td>6.0</td>
</tr>
<tr>
<td>—CH₃</td>
<td>Alanine</td>
<td>A or Ala</td>
<td>2.3</td>
<td>9.7</td>
<td></td>
<td>6.0</td>
</tr>
<tr>
<td>—CH(CH₃)₂</td>
<td>Valine&lt;sup&gt;b&lt;/sup&gt;</td>
<td>V or Val</td>
<td>2.3</td>
<td>9.6</td>
<td></td>
<td>6.0</td>
</tr>
<tr>
<td>—CH₂CH(CH₃)₂</td>
<td>Leucine&lt;sup&gt;b&lt;/sup&gt;</td>
<td>L or Leu</td>
<td>2.4</td>
<td>9.6</td>
<td></td>
<td>6.0</td>
</tr>
<tr>
<td>—CHCH₂CH₃</td>
<td>Isoleucine&lt;sup&gt;b&lt;/sup&gt;</td>
<td>I or Ile</td>
<td>2.4</td>
<td>9.7</td>
<td></td>
<td>6.1</td>
</tr>
<tr>
<td>—CH₂</td>
<td>Phenylalanine&lt;sup&gt;b&lt;/sup&gt;</td>
<td>F or Phe</td>
<td>1.8</td>
<td>9.1</td>
<td></td>
<td>5.5</td>
</tr>
<tr>
<td>—CH₂CONH₂</td>
<td>Asparagine</td>
<td>N or Asn</td>
<td>2.0</td>
<td>8.8</td>
<td></td>
<td>5.4</td>
</tr>
<tr>
<td>—CH₂CH₂CONH₂</td>
<td>Glutamine</td>
<td>Q or Gln</td>
<td>2.2</td>
<td>9.1</td>
<td></td>
<td>5.7</td>
</tr>
<tr>
<td>—CH&lt;sub&gt;2&lt;/sub&gt;CONH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>Tryptophan&lt;sup&gt;b&lt;/sup&gt;</td>
<td>W or Trp</td>
<td>2.4</td>
<td>9.4</td>
<td></td>
<td>5.9</td>
</tr>
<tr>
<td>—O</td>
<td>Proline</td>
<td>P or Pro</td>
<td>2.0</td>
<td>10.6</td>
<td></td>
<td>6.3</td>
</tr>
</tbody>
</table>

<sup>a</sup>Abbreviations used in this table are: G or Gly, A or Ala, V or Val, L or Leu, I or Ile, F or Phe, N or Asn, Q or Gln, W or Trp, P or Pro.
<table>
<thead>
<tr>
<th>Structure of</th>
<th>Name</th>
<th>Abbreviations $^a$</th>
<th>pK$_{\text{a1}}$ $\alpha$-CO$_2$H</th>
<th>pK$_{\text{a2}}$ $\alpha$-NH$_3^+$</th>
<th>pK$_{\text{a3}}$ R group</th>
<th>pI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral Amino Acids</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{CHOH}$</td>
<td>Serine</td>
<td>S or Ser</td>
<td>2.2</td>
<td>9.2</td>
<td></td>
<td>5.7</td>
</tr>
<tr>
<td>$\text{CHOH}$</td>
<td>Threonine$^b$</td>
<td>T or Thr</td>
<td>2.6</td>
<td>10.4</td>
<td></td>
<td>6.5</td>
</tr>
<tr>
<td>$\text{CH}_2\text{OH}$</td>
<td>Tyrosine</td>
<td>Y or Tyr</td>
<td>2.2</td>
<td>9.1</td>
<td>10.1</td>
<td>5.7</td>
</tr>
<tr>
<td>$\text{HOOC}$</td>
<td>Hydroxyproline</td>
<td>Hyp</td>
<td>1.9</td>
<td>9.7</td>
<td></td>
<td>6.3</td>
</tr>
<tr>
<td>(complete structure)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{CH}_2\text{SH}$</td>
<td>Cysteine</td>
<td>C or Cys</td>
<td>1.7</td>
<td>10.8</td>
<td>8.3</td>
<td>5.0</td>
</tr>
<tr>
<td>$\text{CH}_2\text{S}$</td>
<td>Cystine</td>
<td>Cys-Cys</td>
<td>1.6</td>
<td>7.9</td>
<td></td>
<td>5.1</td>
</tr>
<tr>
<td>$\text{CH}_2\text{S}$</td>
<td>Methionine$^b$</td>
<td>M or Met</td>
<td>2.3</td>
<td>9.2</td>
<td></td>
<td>5.8</td>
</tr>
<tr>
<td>Structure of R</td>
<td>Name</td>
<td>Abbreviations(^a)</td>
<td>pK(_{a_1}) (\alpha\text{-CO}_2\text{H})</td>
<td>pK(_{a_2}) (\alpha\text{-NH}_3^+)</td>
<td>pK(_{a_3}) R group</td>
<td>pI</td>
</tr>
<tr>
<td>---------------</td>
<td>------</td>
<td>-------------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>---</td>
</tr>
<tr>
<td>(\text{R Contains an Acidic (Carboxyl) Group})</td>
<td>(\text{—CH}_2\text{CO}_2\text{H})</td>
<td>Aspartic acid</td>
<td>D or Asp</td>
<td>2.1</td>
<td>9.8</td>
<td>3.9</td>
</tr>
<tr>
<td>(\text{—CH}_2\text{CH}_2\text{CO}_2\text{H})</td>
<td>Glutamic acid</td>
<td>E or Glu</td>
<td>2.2</td>
<td>9.7</td>
<td>4.3</td>
<td>3.2</td>
</tr>
<tr>
<td>(\text{R Contains a Basic Group})</td>
<td>(\text{—CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{NH}_2)</td>
<td>Lysine(^b)</td>
<td>K or Lys</td>
<td>2.2</td>
<td>9.0</td>
<td>10.5(^c)</td>
</tr>
<tr>
<td>(\text{—CH}_2\text{CH}_2\text{CH}_2\text{NH}—\text{C—NH}_2)</td>
<td>Arginine</td>
<td>R or Arg</td>
<td>2.2</td>
<td>9.0</td>
<td>12.5(^c)</td>
<td>10.8</td>
</tr>
<tr>
<td>(\text{—CH}_2\text{N}—\text{H})</td>
<td>Histidine</td>
<td>H or His</td>
<td>1.8</td>
<td>9.2</td>
<td>6.0(^c)</td>
<td>7.6</td>
</tr>
</tbody>
</table>

\(^a\)Single-letter abbreviations are now the most commonly used form in current biochemical literature.

\(^b\)An essential amino acid.

\(^c\)pK\(_{a_3}\) is of protonated amine of R group.
Amino Acids as Dipolar Ions

- In the dry solid state amino acids exist as dipolar ions (zwitterions).
- In aqueous solution an equilibrium exists between the dipolar ion, the cationic and the anionic forms of the amino acid.
  - The predominant form depends on the pH of the solution.

\[
\begin{align*}
\text{H}_3\text{NCHCO}_2\text{H}^+ & \rightleftharpoons \text{H}_2\text{O}^+ + \text{H}_3\text{NCHCO}_2^- \\
\text{H}_2\text{NCHCO}_2^- & \rightleftharpoons \text{H}_2\text{O}^+ + \text{H}_2\text{NCHCO}_2^- \\
\end{align*}
\]

- Cationic form (predominant in strongly acidic solutions, e.g., at pH 0)
- Dipolar ion
- Anionic form (predominant in strongly basic solutions, e.g., at pH 14)

- At low pH the amino acid exists primarily in the cationic form.
- At high pH the amino acid exists primarily in the anionic form.
- At some intermediate pH called the pI (isoelectric point), the concentration of the dipolar ion is at a maximum and the concentrations of anionic and cationic forms are equal.
- Each individual amino acid has a characteristic pI (see Table 24.1).
  - Entire proteins also have a characteristic pI.
The amino acid alanine has a neutral side chain and can be used to illustrate the fundamental behavior of an amino acid at various pHs.

- At low pH alanine exist as the cation.
- $pK_{a1}$ of alanine (for ionization of the carboxylic acid proton) is 2.3, considerably lower than the $pK_a$ of a normal carboxylic acid.

$$\text{CH}_3\text{CHCO}_2\text{H} \rightleftharpoons \text{CH}_3\text{CHCO}_2^- + \text{H}_3\text{O}^+$$

Cationic form of alanine

$\text{CH}_3\text{CHCO}_2\text{H}$

$\text{CH}_3\text{CH}_2\text{CO}_2\text{H}$

$\text{CH}_3\text{CH}_2\text{CO}_2^- \rightleftharpoons \text{CH}_3\text{CH}_2\text{CO}_2^- + \text{H}_3\text{O}^+$

Propanoic acid

$\text{pK}_{a1} = 2.3$

$\text{pK}_{a2} = 9.7$

- $pK_{a2}$ of alanine (for ionization of a proton from the protonated amino group) is 9.7.

The isoelectric point, $pI$, for alanine is the average of the two $pK_a$ values $i.e. (pK_{a1} + pK_{a2})/2$.
When base is slowly added to fully protonated alanine, a pH is reached where half of the carboxylic acid groups are deprotonated.

This pH of 2.3 is the value of $pK_{a1}$.

The Henderson-Hasselbach equation predicts this result:

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

As more base is added, the pI is reached and the molecule is electrically neutral; this point is reached when exactly one equivalent of base is added.

As more base is added and pH 9.7 is reached, half of the aminium groups will be deprotonated.

Addition of more base will eventually produce only the anionic amino acid.
Lysine, which contains a basic side-chain, has a more complex equilibrium

- The pI for lysine will be high because of the presence of two basic groups.
- The pI for lysine is the average of the monocation ($pK_{a2}$) and the dipolar ion ($pK_{a3}$).

![Chemical reaction diagram]

- Dicationic form of lysine ($pK_{a1} = 2.2$)
- Monocationic form ($pK_{a2} = 9.0$)
- Dipolar ion ($pK_{a3} = 10.5$)
- Anionic form
Synthesis of $\alpha$-Amino Acids

The first three methods result in racemic mixtures of amino acids

- **Direct Ammonolysis of an $\alpha$-Halo Acid**
  - Yields tend to be poor in this reaction
  
  \[
  \text{RCHCO}_2\text{H} + \text{NH}_3 \text{(excess)} \rightarrow \text{RCHCO}_2^- + \text{NH}_3
  \]

- **From Potassium Phthalimide**
  - This is a variation of the Gabriel synthesis and yields are usually high
The Strecker Synthesis

Treatment of an aldehyde with ammonia and hydrogen cyanide yields an α-aminonitrile which is hydrolyzed to the α-amino acid.

The reaction proceeds via an intermediate imine.

\[
\begin{align*}
\text{RCHO} + \text{NH}_3 + \text{HCN} & \rightarrow \text{RCHCN} \xrightarrow{\text{H}_2\text{O}, \text{heat}} \text{RCHCO}_2^- \\
& \quad \downarrow \text{NH}_2 \\
& \quad \alpha\text{-Amino nitrile} \\
\end{align*}
\]

\[
\begin{align*}
\text{RCH} + :\text{NH}_3 & \leftrightarrow \text{RCHNH}_3 \leftrightarrow \text{RCHNH}_2 \xrightarrow{-\text{H}_2\text{O}} \text{RCH} = \text{NH} \leftrightarrow \text{RCH} = \text{NH} \xrightarrow{\text{H}_2\text{O}^+, \text{heat}} \text{RCH} - \text{NH}_2 \\
& \quad \alpha\text{-Amino acid} \\
\end{align*}
\]

Chapter 24
Resolution of DL-Amino Acids

A racemic amino acid mixture can be resolved by (1) conversion to a racemic mixture of \( N \)-acylamino acids, followed by (2) hydrolysis with a deacylase enzyme that selectively deacylates the L-acylamino acid.
Asymmetric Synthesis of Amino Acids

- Enantioselective syntheses that produce only the desired naturally occurring amino acid enantiomers are ideal.
- One important method involves asymmetric hydrogenation of an enamide using a chiral transition metal catalyst.

This method was used to synthesize L-dopa, a chiral amino acid used in the treatment of Parkinson’s disease.

Asymmetric Synthesis of L-DOPA

\[
\begin{align*}
\text{H}_3\text{CO} & \quad \text{AcO} \\
\text{NHAc} & \quad \text{COOH} \\
\end{align*}
\]

\[
\text{H}_3\text{CO} \quad \text{AcO} \\
\text{NHAc} \quad \text{COOH}
\]

100% Yield, 95% ee

\[
\begin{align*}
\text{H}_3\text{CO} & \quad \text{AcO} \\
\text{NHAc} & \quad \text{COOH} \\
\end{align*}
\]

L-DOPA

\[
\begin{align*}
\text{H}_3\text{CO} & \quad \text{AcO} \\
\text{NHAc} & \quad \text{COOH} \\
\end{align*}
\]

\[
\text{(R,R)-DiPAMP} \\
\text{(chiral ligand for rhodium)}
\]

COD = 1,5-Cyclooctadiene

Chapter 24
A similar method is used to synthesize (S)-phenylalanine, needed for preparation of Aspartame.

**Asymmetric Synthesis of Aspartame**

\[
\begin{align*}
\text{(2,3-dimethylphosphino)phenyl} & \quad \text{Rh(I) (cat.), H}, \quad \text{83\% ee} \\
\text{(catalytic asymmetric hydrogenation)} \\
\text{MeOH, HA} \\
\end{align*}
\]

\[
\begin{align*}
\text{(S)-Phenylalanine methyl ester} \\
(97\% \text{ ee after recrystallization})
\end{align*}
\]

\[
\begin{align*}
\text{(R,R)-PNPP} \\
(\text{chiral ligand for rhodium})
\end{align*}
\]

Aspartame
Polypeptides and Proteins

- Enzymes polymerize amino acids by forming amide linkages.
- The polymer is called a *peptide* and the amide linkages are called *peptide bonds* or *peptide linkages*.
- Each amino acid in the peptide is called an amino acid residue.
- Proteins can contain one or more polypeptide chains and other associated molecules or metal ions.

![Chemical structure of a dipeptide]

A dipeptide
Polypeptides are customarily written with the $N$-terminal residue to the left.

Three letter or one letter abbreviations are usually used as a short hand to indicate the sequence of a polypeptide.
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.

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

\[ \text{Labeled polypeptide} \]
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.

![Graph showing amino acid content](image)

**PTH Standards**
10 picomoles each

Retention time (minutes)
- **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.

  ![Chemical Reaction Diagram]

  - 2,4-Dinitrofluorobenzene (DNFB) + Polypeptide $\xrightarrow{\text{HCO}_3^- (-HF)}$ Labeled polypeptide

  - Labeled N-terminal amino acid + Mixture of amino acids $\xrightarrow{\text{H}_2\text{O}^+}$ Separate and identify

- **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.
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.

- Example: A pentapeptide is known to contain the following amino acids:

  Val₂, Leu, His, Phe

- Using DNFB and carboxypeptidase, the N-terminal and C-terminal amino acids are identified.

  Val (Val, His, Phe) Leu

- The pentapeptide is subjected to partial hydrolysis and the following dipeptides are obtained.

  Val · His + His · Val + Val · Phe + Phe · Leu

- The amino acid sequence of the pentapeptide must be:

  Val · His · Val · Phe · Leu
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

- “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
Examples of Polypeptide and Protein Primary Structure

- Oxytocin and Vasopressin
  - Oxytocin stimulates uterine contractions during childbirth
  - Vasopressin causes contraction of peripheral blood vessels and a resultant increase in blood pressure

The two polypeptides are nonapeptides and differ in only 2 amino acid residues

Oxytocin
Vasopressin
Insulin

- Insulin is a hormone which regulates glucose metabolism
  - Insulin deficiency in humans is the major cause of diabetes mellitus
  - The structure of bovine insulin (shown below) was determined in 1953 by Sanger
  - Human insulin differs from bovine insulin at only three amino acids in its sequence

```plaintext
A chain
Gly—Ile—Val—Glu—Gln—Cys
  Cys—Ala

B chain
Phe—Val—Asn—Gln—His—Leu—Cys
  Gly
  Glu—Val—Leu—His—Ser
  Ala
  Leu—Tyr—Leu—Val—Cys
  Gly
  Asn

Ala—Lys—Pro—Thr—Tyr—Phe—Phe—Gly—Arg—Glu
```
Polypeptide and Protein Synthesis

Laboratory synthesis of polypeptides requires orchestration of blocking and activating groups to achieve selective amide bond formation:
- Amino groups must be blocked until their reactivity as a nucleophile is desired
- Carboxylic acid groups must be activated for acyl substitution at the appropriate time

Amino groups are usually blocked using one of the following:
- A benzyloxycarbonyl group (a “Z” group)
- A di-tert-butyloxycarbonyl group (a “Boc” group)
- An 9-fluorenylmethoxycarbonyl group (an “Fmoc” group)

Methods for installing and removing Z, Boc, and Fmoc groups are shown below:

![Diagram of tert-Butyloxycarbonyl Group]

Protection

\[ (\text{CH}_3)_3\text{COOC(CH}_3)_3 + \text{H}_2\text{N} \rightarrow \text{R} \quad \text{base} \quad 25^\circ\text{C} \]

\[ (\text{CH}_3)_3\text{COC} \rightarrow \text{NHR} + (\text{CH}_3)_3\text{COH} \]

Deprotection

\[ (\text{CH}_3)_2\text{C} \equiv \text{CH}_2 + \text{CO}_2 + \text{H}_2\text{N} \rightarrow \text{R} \]

Chapter 24
Methods for installing and removing Z, Boc, and Fmoc groups are shown below:

9-Fluorenylmethoxycarbonyl Group

Protection (introduction of Fmoc group)

Amino acid (side chain protected in advance if necessary)

Fmoc-protected amino acid (stable in acid)

Fmoc =

(9-Fluorenylmethoxycarbonyl group)

Deprotection (removal of Fmoc group)

Fmoc-protected amino acid

(Byproduct)
Carboxylic acid groups are usually activated by conversion to a mixed anhydride:

Ethyl chloroformate can be used.
An Example of Laboratory Peptide Synthesis:

Synthesis of Ala-Leu

\[
\text{CH}_3\text{CHCO}_2^- + \text{C}_6\text{H}_5\text{CH}_2\text{OCl} \xrightarrow{\text{OH}^- 25^\circ\text{C}} \text{CH}_3\text{CH}_2\text{CO}_2\text{H} \quad (1) \quad (\text{C}_6\text{H}_5)_3\text{N} \\
\text{N} \quad \text{NH} \\
\text{C} \quad \text{C} = \text{O} \\
\text{C}_6\text{H}_5\text{CH}_2\text{O} \quad \text{Z-Ala} \\
\text{CH}_2\text{CH} - \text{C} - \text{OCOC}_2\text{H}_5 \xrightarrow{\text{CH}_2\text{CCH}_2\text{CH}_2\text{CO}_2^-} \text{CO}_2 + \text{C}_2\text{H}_5\text{OH} \\
\text{NH} \quad \text{C} = \text{O} \\
\text{C}_6\text{H}_5\text{CH}_2\text{O} \quad \text{Mixed anhydride of Z-Ala} \\
\text{CH}_2\text{CH} - \text{C} - \text{NHCHCO}_2\text{H} \xrightarrow{\text{H}_2/\text{Pd}} \text{Z-Ala} \cdot \text{Leu} \\
\text{C}_6\text{H}_5\text{CH}_2\text{O} \quad \text{Z-Ala} \cdot \text{Leu} \\
\text{CH}_3\text{CHCNHCHCO}_2^- + \text{NH}_3 \xrightarrow{\text{CH}_2} \text{CH}_3 + \text{CO}_2 \\
\text{NH}_3 \quad \text{CH}_2 \\
\text{H}_3\text{C} \quad \text{CH}_3 \\
\text{Ala} \cdot \text{Leu}
\]
Automated Peptide Synthesis

Solid Phase Peptide Synthesis (SPSS) was invented by R. B. Merrifield, for which he earned the Nobel Prize in 1984.

SPSS involves ‘growing’ a peptide on a solid polymer bead by sequential cycles of amide bond formation.

The peptide is cleaved from the bead when the synthesis is complete.

SPSS is used in commercial peptide synthesis machines:
- Peptides dozens of residues in length can be synthesized automatically.
- A landmark example is synthesis of ribonuclease, having 124 amino acid residues.
Chapter 24

Step 1: Attaches C-terminal (Fmoc protected) amino acid residue to resin

Step 2: Purifies resin with attached residue by washing

Step 3: Removes Fmoc protecting group

Step 4: Purifies by washing

Step 5: Adds next (protected) amino acid residue

Step 6: Purifies by washing

Step 7: Removes protecting group

Final Step: Detaches completed polypeptide

etc.
Secondary, Tertiary, and Quaternary Structures of Proteins

- Secondary Structure
  - The secondary structure of a protein is defined by local conformations of its polypeptide backbone
    - These local conformations are specified in terms of regular folding patterns such as helices, pleated sheets, and turns
  - The secondary structure of a protein is determined by the sequence of amino acids in its primary structure
  - Key to secondary structure is that peptide bonds assume a geometry in which all 6 atoms of the amide linkage are trans coplanar

Chapter 24
Coplanarity results from contribution of the second resonance form of amides, in which there is considerable N-C double bond character.

The carbon with attached R groups between the amide nitrogen and the carbonyl group has relatively free rotation and this leads to different conformations of the overall chain.

- Two common secondary structure are the β-pleated sheet and the α-helix.
  - In the β-pleated sheet, a polypeptide chain is in an extended conformation with groups alternating from side to side.
The extended polypeptide chains in β-pleated sheets form hydrogen bonds to adjacent polypeptide chains.

- Slight bond rotations are necessary between amide groups to avoid unfavorable steric interactions between peptide side chains, leading to the pleated structure.
- The β-pleated sheet is the predominant structure in silk fibroin.
The α-helix is the most important protein secondary structure

α-Helices in a polypeptide are right-handed with 3.6 amino acid residues per turn (See figure 24.11 page 1198)

- The amide nitrogen has a hydrogen bond to an amino acid carbonyl oxygen that is three residues away
- The R groups extend away from the axis of the helix

α-Helices comprise the predominant secondary structure of fibrous proteins such as myosin (in muscle) and α-keratin (in hair and nails)

There are other secondary structures that are more difficult to describe

- Examples are coil or loop conformations and reverse turns or β bends
Carbonic Anhydrase

The structure of the enzyme carbonic anhydrase is shown in Figure 24.12, page 1198.

- Alpha helices are in magenta and strands of β-pleated sheets are in yellow.
- The mechanism of carbonic anhydrase reaction was discussed in Chapter 3.
Tertiary Structure

- The tertiary structure of a protein is the three-dimensional shape which results from further folding of its polypeptide chains.
  - This folding is superimposed on the folding caused by its secondary structure.

- In globular proteins, the folding in tertiary structures exposes the maximum number of polar (hydrophilic) side chains to the aqueous environment, making most globular proteins water soluble.
  - The folding also serves to enclose a maximum number of nonpolar (hydrophobic) side chains within the protein interior.

- Tertiary structures are stabilized by forces including hydrogen bonding, disulfide bonds, van der Waals forces, and ionic attractions.
Myoglobin

The globular protein myoglobin transports oxygen within muscle tissues (See Figure 24.13, page 1200)

- Myoglobin has an associated non-polypeptide molecule called heme (shown in gray)
- The heme group is the site of oxygen binding
• Quaternary Structure
  ➔ The overall structure of a protein having multiple subunits is called its quaternary structure
  ✶ Not all proteins have quaternary structure

• Hemoglobin
  ➔ Hemoglobin is a globular protein that transports oxygen in the blood
  ➔ Hemoglobin contains four polypeptide subunits (2 designated α, and 2 designated β) (See Figure 24.21, page 1210)
  ✶ The α subunits are shown in blue and green; β subunits are shown in yellow and cyan
Each of the four protein subunits carries a heme group

- The four heme groups are shown in purple
- Each heme group can bind one oxygen molecule in a reversible complex
● Introduction to Enzymes

- Most enzymes are proteins
- Enzymes can catalyze reactions by a factor of $10^6$-$10^{12}$
- Enzymes have very high specificity for their respective substrates (reactants)
- Enzymatic reactions take place in the active site of each enzyme
  - The structure of the active site facilitates binding and catalysis

- Enzymes sometimes require a cofactor or coenzyme
  - A cofactor can be a metal ion (e.g., Zn$^{+2}$, Mg$^{+2}$) bound at the active site
  - A coenzyme is a small organic molecule bound at the active site that becomes chemically changed during the enzymatic reaction (e.g., NAD$^+$)
Lysozyme

Lysozyme catalyzes hydrolysis of a glycosidic linkage in the polysaccharide cell wall of bacteria

The mechanism of lysozyme involves acid-base reactions and $S_N^1$ reaction

The mechanism of lysozyme is shown in Figure 24.16, page 1204
Serine Proteases

- Proteases hydrolyze amide bonds in proteins
- Chymotrypsin, trypsin, and elastin are serine proteases
- Serine proteases have a serine hydroxyl group that is involved in the mechanism of amide bond hydrolysis
  - A “catalytic triad” involving the side chains of specific aspartic acid, histidine, and serine residues catalyze the amide hydrolysis
  - The serine hydroxyl attacks the amide carbonyl group, forming a tetrahedral intermediate
  - The aspartic acid and histidine side chains form an acid-base relay system to assist with protonation and deprotonation steps
  - The serine tetrahedral intermediate releases the amine, leaving an acylated serine
  - A water molecule attacks the carbonyl group of the acylated serine
  - A new tetrahedral intermediate forms
  - When this tetrahedral intermediate collapses to the carboxylic acid, the serine hydroxyl is released for a new catalytic cycle
  - See the following slide for the mechanism of trypsin

The Active Site Catalytic Triad of Trypsin

- This is shown figure 24.17, page 1205
The Catalytic Mechanism of Trypsin

- Hydrophobic pocket binding site
- Tetrahedral intermediate
- Acylated serine residue
- Regenerated active site
- Carboxylic acid product
• Purification and Analysis of Polypeptides and Proteins
  ➔ Proteins are purified initially by precipitation, column chromatography, and electrophoresis
  ➔ HPLC is the method of choice for final purification of a protein

• Analysis of proteins
  ➔ Molecular weight can be estimated by gel electrophoresis and size exclusion chromatography
  ➔ Mass spectrometry is used to determine protein molecular weights with high accuracy and precision
  ◆ Electrospray ionization (ESI) mass spectrometry is one way to create protein ions for mass spectrometry
  ◆ Matrix-assisted laser desorption ionization (MALDI) mass spectrometry is another technique for generating protein ions for mass spectrometry
  ◆ The 2002 Nobel Prize in Chemistry was awarded in part for development of ESI (by Fenn, et al) and MALDI (by Tanaku) for mass spectrometry
Electrospray Ionization (ESI) Mass Spectrometry (MS)

- Multiply charged ions of the analyte (e.g., a protein sample) are formed by protonation in an acidic solvent
- The protonated analyte may have one, several, or many positive charges
- The charged analyte is sprayed through a high-voltage nozzle into a vacuum chamber
- Molecules of the solvent evaporate, leaving ‘naked’ ions of the multiply charged analyte
- The ions are drawn into a mass analyzer and detected according to mass-to-charge (m/z) ratio
- Quadrupole and time of flight (TOF) mass analyzers are common methods for detecting and separating ions
- The family of detected ions is displayed as a series according to m/z ratio
- Computer deconvolution of the m/z peak series leads to the molecular weight of the analyte
Proteomics

Proteomics involves identification and quantification of all of the proteins expressed in a cell at a given time.

- Proteins expression levels vary in cells over time
- Proteomics involves identification and quantification of all of the proteins expressed in a cell at a given time
- Proteomics data can shed light on the health or life-cycle stage of a cell

Tools for Proteomics

- Polyacrylamide gel electrophoresis (2D-PAGE) is a low resolution technique for separating protein mixtures
- Two-dimensional (2D) microcapillary HPLC coupled with mass spectrometry is a high resolution technique for separating and identifying proteins in a cell extract
● Multidimensional Protein Identification Technology

   MudPIT (Multidimensional protein identification technology) involves:
   - Lysis of intact cells
   - Digestion of the proteins to a mixture of smaller peptides
   - Separation of the peptide mixture by 2D HPLC using a strong cation exchange column in tandem with a reversed-phase (hydrophobic) column
   - Direct introduction of the 2D HPLC eluent into a mass spectrometer
   - Comparison of mass spectra with a database of mass spectral data for known proteins
   - Data matching can lead to identification of >1000 proteins in one integrated analysis