Protein Building Blocks:
Amino Acids, Peptides and Polypeptides

Protein Functions:
- Enzymes – catalysis
- Tissue structure
- Cell structure
- Hormones
- Hormone receptors
- Ion channels
- Transporters
- Transcription factors
- Contraction
- etc.

Protein Synthesis

- Transcription
- Translation

DNA

genes

protein

mRNA (codon sequence for amino acids)

Protein (specific amino acid sequence)
How Many Proteins?
“Proteome” = a catalog of all proteins

Amino Acids

R-groups in Amino Acids

Hydrophobic
Non-polar; hydrocarbon
Doesn’t interact with H₂O

Hydrophilic
Polar/charged
Interacts with H₂O; hydrogen bonds to N, O

In General:
Hydrophobic amino acids are in the interior of a folded protein
Hydrophilic amino acids are on the surface in contact with water.
Common Amino Acids (20)

Common Amino Acids (cont'd)

<table>
<thead>
<tr>
<th>Type</th>
<th>Amino Acid</th>
<th>Charge</th>
<th>Synthesis</th>
<th>Soluble</th>
<th>Cyclic</th>
<th>All</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>Glutamic</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Aspartic</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Valine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Leucine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Isoleucine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Methionine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Phenylalanine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Tyrosine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Histidine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Lysine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Proline</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Alanine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Threonine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Serine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Glutamine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Asparagine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

*Note: Amino acids in the table are generally considered "all." Some may be "soluble" or "cyclic." The "protein" column indicates their occurrence in this context.*
Peptides and the Peptide Bond

\[
\text{glycine} + \text{alanine} \rightarrow \text{Glycylalanine} (\text{a dipeptide})
\]

Methionine Enkephalin

1. Tyr-Gly-Gly-Phe-Met
2. \(H_2N\text{Tyr-Gly-Gly-Phe-Met-COO}^{-}\)
3. \(H\text{Tyr-Gly-Gly-Phe-Met-OH}\)

N-terminus; COOH-terminus

(An amino acid in a peptide is often called a “residue”)

A Derived Amino Acid—Cystine

Oxidation of two cysteine side chain thiol groups

Two cysteines

Cystine

(a disulfide bond)
Antidiuretic Hormone—Vasopressin
(a nonapeptide)

\[
\begin{array}{cccccccc}
1 & H & Cys & Tyr & Phe & Gln & Asn & Cys & Pro & Arg & Gly(NH_2)
\end{array}
\]

Insulin

Peptide or Protein?

Peptide: <50 amino acids
Protein: >50 amino acids

Median protein size: ≈ 400 amino acids
Largest protein (titin): ≈ 30,000 amino acids
Acid - Base Review

Definitions:
Acid: proton donor
Base: proton acceptor

Acid dissociation equilibrium constant $K_a$:

$$\text{HA} \leftrightarrow \text{H}^+ + \text{A}^-$$

The concentrations [ ] at equilibrium:

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

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

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

Henderson-Hasselbalch Equation

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

<table>
<thead>
<tr>
<th>$pH - pK_a$</th>
<th>$\frac{[\text{A}^-]}{[\text{HA}]}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>1000</td>
</tr>
<tr>
<td>-1</td>
<td>0.1</td>
</tr>
<tr>
<td>-2</td>
<td>0.01</td>
</tr>
<tr>
<td>-3</td>
<td>0.001</td>
</tr>
</tbody>
</table>
**pKₐ Values of Some Amino Acid Side Chains**

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Uncharged Form</th>
<th>Charged Form</th>
<th>pKₐ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>CH₃CH(NH₂)CO₂⁻</td>
<td>CH₃CH(NH₃)⁺</td>
<td>4.25</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>H₂NCH(COO⁻)CH₃</td>
<td>CH₂(NH₂)CO₂⁻</td>
<td>2.24</td>
</tr>
<tr>
<td>Lysine</td>
<td>CH₁₂H₂₁N⁺</td>
<td>CH₁₂H₂₁N⁺</td>
<td>10.5</td>
</tr>
</tbody>
</table>

**Isoelectric point (pI): pH where equal number of + and – charges (zwitterion).**

(The pI is half way between the pKₐ’s that border the zwitterion.)
Proteins also have a pI
- pI (isoelectric point)
- Equal + and – charges (no net charge)
- Doesn’t migrate in an electric field
- Depends on the amino acid composition

Human serum albumin (585 amino acids)
- 36 Aspartates
- 61 Glutamates
- 57 Lysines
- 24 Arginines
- 16 Histidines

\[ pI = 5.9 \]

Clinical Use of Protein Electrophoresis

Agarose gel electrophoresis of plasma proteins carried out at pH 8.6:

- Normal Pattern
- Paraprotein ("Monoclonal Gammapathy")
  - e.g., multiple myeloma
The Proteome: 2-Dimensional Electrophoresis

The method:
Separation of a mixture of proteins by pi:
Isoelectric focusing

Sodium dodecyl
sulfate gel
electrophoresis
(separation of
proteins by size)

Then stain gel for protein

[Image of 2D electrophoresis gel]

[Image of molecular mass KOA]
Cut out spot
Digest protein with an enzyme:
- trypsin (cleaves after a Lys or Arg)
Separate fragments by HPLC*, measure mass

Cancer Proteomics:
Determine protein/peptide biomarkers
- Early Diagnosis
- Staging (severity)
- Molecular-targeted therapy
- Monitoring therapy

http://proteomics.cancer.gov/proteomics/proteogenomics