PROTEIN BUILDING BLOCKS: AMINO ACIDS, PEPTIDES, AND POLYPEPTIDES

Date: August 5, 2019 (9:00)

KEY CONCEPTS AND LEARNING OBJECTIVES

1. DNA contains regions (genes) that code for the structures of proteins. A gene is transcribed into an RNA molecule, which is then modified to an mRNA, transported to the cytosol and translated into a protein.
   a. Explain the overall principle of this scheme.
   b. Indicate the approximate number of genes and how their transcription can ultimately lead to >1 million proteins.

2. The general structure of a common amino acid is comprised of a central carbon atom to which an amino group, a carboxylic acid group, a hydrogen atom, and a side chain group are attached.
   a. Illustrate the general structure of an amino acid.
   b. Indicate which optical isomer of amino acids is found in proteins.
   c. Define hydrophobic and hydrophilic as it pertains to amino acid side chains.

3. The 20 common amino acids are defined by the nature of their side chain groups.
   a. Name a common amino acid upon viewing the structure of its side chain group.
   b. Reproduce the three-letter abbreviations for the common amino acids.
   c. Recognize the existence of the single letter abbreviations for the common amino acids.
   d. Given the name or structure of an amino acid, describe its unique physical and chemical properties.
   e. Describe the unique structural feature of proline.

4. Amino acids are joined by a peptide bond that connects the α-carboxylic acid group of one amino acid with the α-amino group of a second amino acid.
   a. Write the reaction for the joining of two amino acids to form a peptide bond.

5. Polypeptide chains have an NH₂-terminal end and a COOH-terminal end. The numbering of the amino acids in a polypeptide chain proceeds from the NH₂-terminal towards the COOH-terminal.
   a. Identify the residue number of an amino acid within a polypeptide chain.
   b. Interpret the different ways that a free N-terminal or C-terminal amino acid can be written.
6. A cystine disulfide bond is formed from two cysteines within the primary structure (i.e., the amino acid sequence) of a protein. Cystine bonds act to stabilize the native structure of a folded protein.
   a. Explain the relationship between cysteines and cystine in protein structures.
   b. Define “derived amino acids.”
   c. Recognize that the hormone vasopressin has a disulfide bond and a C-terminal, uncharged amide group.
   d. Describe the basic structure of insulin.

7. Peptides and proteins have a large range of sizes.
   a. Provide an approximate definition of “peptide” and “protein.”

8. Weak acids are characterized by an acidic dissociation constant and the corresponding pK\textsubscript{a} value. The Henderson-Hasselbalch Equation calculates the relationship between 1) the ratio of the concentration of a conjugate acid to its conjugate base, 2) the acid pK\textsubscript{a}, and 3) the solution pH.
   a. Calculate the ratio of the concentration of a base to its conjugate acid when the solution pH and acid pK\textsubscript{a} differ by integer units.
   b. Calculate the pH of a solution when the concentration of an acid and its conjugate base are equal.
   c. Calculate the percent activity at pH 7.0 of an enzyme that requires its active site histidine (pK\textsubscript{a} 6.0) to be in the form of its conjugate base (small group problem).

9. The side chain groups of the amino acids Asp, Glu, His, Lys and Arg determine the charge characteristics of a peptide or protein.
   a. Indicate the approximate side-chain pK\textsubscript{a} values for the carboxylic acid groups of Asp and Glu, the ammonium group of Lys, the imidazolium group of His, and the guanidinium group of Arg.
   b. Predict the contribution of each of these amino acids to the charge of a protein at physiological pH, using the approximate pK\textsubscript{a} values.
   c. Draw a titration curve for alanine (from lecture) and histidine (from the small group problem).
   d. Draw the charged forms that appear at each stage of a titration for alanine, aspartic acid, and lysine.

10. At the isoelectric point, the number of positive and negative charges in a protein are equal, and the protein cannot migrate in an electric field.
    a. Predict the relative pI values for two proteins based on their amino acid compositions.

11. Agarose gel electrophoresis of plasma proteins can be used clinically to diagnose certain diseases.
    a. Describe the method of agarose gel electrophoresis.
    b. Compare the plasma protein patterns of normal and multiple myeloma patients.

12. Modern techniques of proteomics can analyze the expression of thousands of proteins within a cell or tissue in a single assay.
    a. Describe the method of 2-D electrophoresis and how it could be used to identify protein biomarkers for the diagnosis of cancers.
Sequences in DNA that code for RNA and proteins (genes) are transcribed in the nucleus to give an RNA copy. The RNA copy is modified to give mRNA, which is then transported into the cytoplasm. The mRNA is translated into a protein (polymer of amino acids) based on the sequence of codons in the mRNA. The protein can undergo modification after translation.

The Proteome (catalog of all proteins): The ~20,000 genes can give rise to ~100,000 mRNAs by use of alternative promoters (transcription start sites) and modifications of the initial RNA transcript. The proteins translated from the mRNAs can be further modified by addition of groups such as sugars, phosphate, acetyl, hydroxyl, etc. and by cleavage by a protease enzyme to give >1 million distinct proteins.
**Amino Acid—Basic Structure:**

An amino acid consists of an amino group \((\text{-NH}_2)\), an acid group (carboxylic acid, \(-\text{COOH}\)), a hydrogen, and an R-side chain, all attached to one carbon (\(\alpha\)). At physiological pH, the amino group is protonated (ammonium) and the carboxylic acid is deprotonated (carboxylate), to give a “zwitterion.” \(\alpha\) is a chiral center, so two optical isomers are possible. Only L-amino acids are found in proteins.

---

**Hydrophobic**

Non-polar; hydrocarbon

Doesn’t interact with \(\text{H}_2\text{O}\)

**Hydrophilic**

Polar/charged

Interacts with \(\text{H}_2\text{O}\); 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.
Some side chain group names:
- Tryptophan: indole
- Lysine: amino (ammonium +)
- Arginine: guanidine (guanidinium +)
- Histidine: imidazole (imidazolium +)
Abbreviations and Summary of Chemical Properties

<table>
<thead>
<tr>
<th>Type</th>
<th>Amino acid</th>
<th>Abbreviation*</th>
<th>Hydrophobic interactions</th>
<th>Hydrogen bonds†</th>
<th>Salt bonds‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
<td>(+)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Branched chain</td>
<td>Valine</td>
<td>Val</td>
<td>V</td>
<td>++</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
<td>+++</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
<td>+++</td>
<td>–</td>
</tr>
<tr>
<td>Hydroxy</td>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Sulfur</td>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
<td>++</td>
<td>– (+)</td>
</tr>
<tr>
<td></td>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
<td>+++</td>
<td>–</td>
</tr>
<tr>
<td>Aromatic</td>
<td>Phenyllalanine</td>
<td>Phe</td>
<td>F</td>
<td>+++</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
<td>++</td>
<td>–</td>
</tr>
<tr>
<td>Acidic and derivatives</td>
<td>Aspartate</td>
<td>Asp</td>
<td>D</td>
<td>–</td>
<td>++ ++</td>
</tr>
<tr>
<td></td>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
<td>–</td>
<td>+++</td>
</tr>
<tr>
<td>Basic</td>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>Histidine</td>
<td>His</td>
<td>H</td>
<td>–</td>
<td>+++</td>
</tr>
<tr>
<td>Imino</td>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
<td>++</td>
<td>–</td>
</tr>
</tbody>
</table>

* The amino acids are abbreviated by either a three-letter code or a one-letter code.
† Hydrogen bonds formed by the side chain in the uncharged state.
‡ In the pH 6 to 8 range.

The underlined amino acids are charged at or near physiological pH.

Peptide Bond: Amino acids can be linked by a peptide bond between the α-carboxyl group of the first amino acid and the α-ammonium group of the second. The elements of water are removed in the process. The peptide bond is chemically an amide and is planar, with the O and H oriented opposite to one another (trans). The figure shows a “dipeptide” formed from two amino acids, glycine and alanine, which is named glycylalanine.
The figure shows a pentapeptide (five amino acids joined by four peptide bonds). The amino acid sequence of a polypeptide is written such that the amino acid having the free α-ammonium group is first, and the amino acid with the free α-carboxylate group is at the end to the right. The two ends are called the amino (NH$_2$- or N-) terminus and the carboxyl (COOH or C-) terminus, resp. Three different ways of denoting the free N- and C-termini are shown. The numbering of a polypeptide chain starts with the N-terminal residue as 1.

**A Derived Amino Acid—Cystine**

Oxidation of two cysteine side chain thiol groups

Cystine is a “derived amino acid” formed by oxidation of two cysteine thiol groups, resulting in a covalent disulfide bond. There is no mRNA codon for cystine. Instead, the disulfide bond is formed after the peptide or protein is synthesized. The original cysteines, which are separated from each other in the primary sequence, are brought together when the protein folds. A disulfide bond can stabilize the folded structure. (There are many other derived amino acids formed by covalent modifications of the common amino acids after the protein is synthesized.)

**Vasopressin (antidiuretic hormone)**

is a nonapeptide (9-amino acids). It contains two cysteines that form a disulfide bond, resulting in a cyclic structure. The C-terminal glycine carboxylate is modified to form an amide. The glycinamide does not have a charge.
**Insulin** consists of two polypeptide chains (A and B; 21 and 30 amino acids, resp.). The two chains are connected by two covalent disulfide bridges. The A chain has an intra-chain disulfide bond. Insulin is initially made as a single chain “proinsulin” that has a C-(connecting) peptide between the A and B chains (between A-chain Gly-1 and B-chain Thr-30). This structure allows proper alignment of the cysteine residues to form each cystine. The C-peptide is removed later.

---

**Peptide or Protein?**

<table>
<thead>
<tr>
<th>Description</th>
<th>Peptide</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤50 amino acids</td>
<td></td>
<td>&gt;50 amino acids</td>
</tr>
</tbody>
</table>

**Definitions:**

- **Acid:** proton donor
- **Base:** proton acceptor

**Acid dissociation equilibrium constant** $K_a$

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

**Logarithmic forms:**

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

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

**Henderson-Hasselbalch Equation**

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

---

**Acid - Base Review**

**Definitions:**

- **Acid:** proton donor
- **Base:** proton acceptor

**Acid dissociation equilibrium constant** $K_a$

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

**Logarithmic forms:**

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

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

**Henderson-Hasselbalch Equation**

$$pH = pK_a + \log \frac{[A^-]}{[HA]}$$
The Henderson-Hasselbalch equation allows one to calculate pH, pKₐ, or the ratio of conjugate base to conjugate acid if any two of the three parameters is known. The table demonstrates their logarithmic relationship. When the pH is equal to pKₐ, the log [A⁻]/[HA] is zero. The antilog of zero is 1. This means that [A⁻] = [HA]. For each unit change in pH - pKₐ, the [A⁻]/[HA] changes by a factor of 10. For example, if pH - pKₐ is 2, there will be 100 times more A⁻ than HA in solution. If the pH - pKₐ is -2, there will be 100 times more HA than A⁻.

### pKₐ Values of Some Amino Acid Side Chains

(low pKₐ, acidic; high pKₐ basic)

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Protonated Form</th>
<th>Deprotonated Form</th>
<th>pKₐ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glutamate</td>
<td>(CH₂)₂ COOH</td>
<td>(CH₂)₂ COO⁻</td>
<td>4.3</td>
</tr>
<tr>
<td>Aspartate</td>
<td>(CH₂)₂ COOH</td>
<td>(CH₂)₂ COO⁻</td>
<td>3.9</td>
</tr>
<tr>
<td>Cysteine</td>
<td>CH₂ SH</td>
<td>CH₂ S⁻</td>
<td>8.3</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>CH₂ O OH</td>
<td>CH₂ O⁻</td>
<td>10.1</td>
</tr>
<tr>
<td>Lysine</td>
<td>(CH₄)₄ NH⁺</td>
<td>(CH₄)₄ NH⁻</td>
<td>10.8</td>
</tr>
<tr>
<td>Arginine</td>
<td>(CH₃)₃ NH⁺</td>
<td>(CH₃)₃ NH⁻</td>
<td>12.5</td>
</tr>
<tr>
<td>Histidine</td>
<td>CH₂ NH⁺</td>
<td>CH₂ NH⁻</td>
<td>6.0</td>
</tr>
</tbody>
</table>

*In proteins, the side chain pKₐ values may differ by more than one pH unit from those in the free amino acids.

The Table above shows the pKₐ values for ionizable side chains and for Cα amino and carboxyl groups in free amino acids and peptides. The “protonated form” is considered the conjugate acid and the “deprotonated form” the conjugate base. The pKₐ values for side chains are an average, as they can vary depending on their environment in a protein. Based on the pKₐ value, and assuming a physiological pH of 7.4, the Henderson-Hasselbalch equation can be used to determine the degree of ionization of the side chain. For example, the glutamate side chain carboxyl group (pKₐ = 4.3) is almost completely in the –COO⁻ deprotonated form (1260 to 1). Similarly, the basic amino acids lysine and arginine are almost completely in the protonated form.
Titration of Alanine. In this titration, the pH of a solution of alanine was monitored during addition of an increasing amount of NaOH. The titration was started at pH 1.0, where both the Cα amino group and carboxyl group are fully protonated and the charge is +1. The addition of NaOH initially pulled off a proton from the carboxylic acid group to give the carboxylate and an overall net charge of 0. Addition of more NaOH pulled off the proton from the ammonium group to give an uncharged amino group and a net charge of -1. Note that the pH did not change much when the carboxylic acid was being titrated. This is because alanine was acting as a buffer. When the carboxylic acid was completely titrated, the pH rose rapidly, as NaOH was now pulling H+ from water. Alanine also served as a buffer when the ammonium group was being titrated, such that the pH did not change much.

We know that the pKa of a group is equivalent to the pH where the conjugate acid and conjugate base are present in equal concentrations. Therefore the middle of the flat portion of the titration curve indicates its pKa, as shown in the figure.

The isoelectric point (pI) is the pH at which there is an equal number of + and – charges, such that the molecule would not migrate in an electric field. The pI can be calculated from the average of the pKa’s values that border the zwitterion form. For aspartic acid, this is (2.0 + 3.9)/2 = 2.95. For lysine, this is (9.2 + 10.8)/2 = 10.0.

The charged forms of aspartic acid and lysine during titration. The pKa’s of each group are indicated. They are equivalent to the pH where the conjugate acid and conjugate base are present in equal concentrations. (The titration curves—not shown—do not clearly show the pKa’s of these side chains because they overlap with the Cα groups.)
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

The Isoelectric Point (pI) of Human Serum Albumin.
Albumin has equal + and – charges at pH 5.9. At pH values below 5.9, the protein is positively charged. At pH values above 5.9, the protein is negatively charged.

Plasma proteins can be separated by agarose gel electrophoresis. At pH 8.6, a majority of proteins are negatively charged (pH above their pI’s) and migrate towards the anode at different speeds based on charge, size, and shape. The gel is stained to visualize the proteins and scanned to quantitate them. Compared with the normal pattern (left), a patient with multiple myeloma has decreased albumin and a major gamma-globulin spike that is an antibody produced by a single cancerous clone of “plasma” cells in the bone marrow (right). (Fig. 3.20, 3.21)
The Proteome: 2-Dimensional Electrophoresis

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

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

2-Dimensional electrophoresis can separate thousands of proteins. A mixture of proteins (e.g. from a cell extract or tumor sample) is applied to an isoelectric focusing gel, which has a pH gradient across its length. When put into an electric field, the negatively charged proteins will migrate to the anode (+). As they move through the gel and encounter lower pH’s, the proteins become protonated and less negatively charged until they finally reach their isoelectric points and stop moving. Similarly, positively charged proteins will move to the cathode until they reach their isoelectric points. Proteins therefore get separated based on their pI’s. But many proteins will have the same pI and will not get separated from each other. The gel is then put on top of a different type of gel that separates proteins by size rather than charge. As they move down the second gel in a perpendicular direction, proteins with the same pI will separate according to size. Finally, the gel can be stained (e.g., with a silver-containing stain) to visualize the proteins.
Proteins were separated horizontally by isoelectric focusing (pH 4.9-5.5) and then further separated vertically by SDS-gel electrophoresis according to molecular weight. (Fig. 3.67)

Are there spots that are different between normal tissue and cancer tissue? Identify them (next page).
Cut out spot
Digest protein with an enzyme: trypsin (cleaves after a Lys or Arg)
Separate fragments by HPLC*, measure mass

*High Performance Liquid Chromatography

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

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