Lecture 5 (9/17/21)

OUTLINE
Amino Acids
- Definition
- The 4 S’s
- Common Properties
- Five Classes
  - Hydrophobic–aliphatic [6]
  - Hydrophobic–aromatic [3]
  - Special–sulfur [2]
  - Hydrophilic–charged [5]
- Other amino acids
- Linking amino acids
- Acid/base properties
  - Titrations
  - Isoelectric point
- Electrophoresis

• Reading: Ch3; 82–87, 92–94
• Problems: Ch3; 3,5,8,13,14, Ch2; 34

NEXT
• Reading & Problems announced after exam

Hydrophilic, Charged

Amino Acids: Classification

<table>
<thead>
<tr>
<th>Name</th>
<th>3-letter</th>
<th>1-letter</th>
<th>Year discovered</th>
<th>% abundance in proteins</th>
<th>NOTES</th>
<th>(pK_a)</th>
<th>Structure mnemonic device</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspartate</td>
<td>Asp</td>
<td>D</td>
<td>1868</td>
<td>5</td>
<td>(\alpha)-amino-succinate; Most acidic(\alpha)-amino-glutarate</td>
<td>3.7</td>
<td>Ala+carboxyl</td>
</tr>
<tr>
<td>Glutamate</td>
<td>Glu</td>
<td>E</td>
<td>1866</td>
<td>7</td>
<td>(\alpha)-amino-glutarate</td>
<td>4.3</td>
<td>Ala+acetate</td>
</tr>
</tbody>
</table>
# Amino Acids: Naming Dicarboxylic Acids

<table>
<thead>
<tr>
<th>Rubric</th>
<th>Name (conjugate base)</th>
<th>Name of Acid</th>
<th>Structure (conjugate base)</th>
<th>X = (CH₂)ₓ in “OOOC-X-COO”&lt;sup&gt;−&lt;/sup&gt;</th>
<th>z=?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oh</td>
<td>oxalate</td>
<td>oxalic</td>
<td>OOOC-COO&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>My</td>
<td>malate</td>
<td>malic</td>
<td>OOOC-CH₂-COO&lt;sup&gt;−&lt;/sup&gt;</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Such</td>
<td>succinate</td>
<td>succinic</td>
<td>OOOC-CH₃-C₄H₉-COO&lt;sup&gt;−&lt;/sup&gt;</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>glutarate</td>
<td>glutamic</td>
<td>OOOC-CH₃-C₆H₁₃-COO&lt;sup&gt;−&lt;/sup&gt;</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Apple</td>
<td>adipate</td>
<td>adipic</td>
<td>OOOC-CH₃-C₈H₁₇-COO&lt;sup&gt;−&lt;/sup&gt;</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Pia</td>
<td>pimelate</td>
<td>pimelic</td>
<td>OOOC-CH₃-C₁₀H₂₁-COO&lt;sup&gt;−&lt;/sup&gt;</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
Hydrophilic, Charged Amino Acids: Classification

As the electrons creating the double bond become delocalized, the positive charge also becomes delocalized.

<table>
<thead>
<tr>
<th>Name</th>
<th>3-letter</th>
<th>1-letter</th>
<th>Year discovered</th>
<th>% abundance in proteins</th>
<th>NOTES</th>
<th>Structure mnemonic</th>
<th>pK_a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspartate</td>
<td>Asp</td>
<td>D</td>
<td>1868</td>
<td>5</td>
<td>α-amino-succinate; Most acidic</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>Glutamate</td>
<td>Glu</td>
<td>E</td>
<td>1866</td>
<td>7</td>
<td>α-amino-glutarate</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
<td>1889</td>
<td>6</td>
<td>Only “bis” amino acid</td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
<td>1886</td>
<td>5</td>
<td>Most basic</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
<td>1896</td>
<td>2</td>
<td>Only physiological ionization</td>
<td>6.0</td>
<td></td>
</tr>
</tbody>
</table>
### Amino Acids: Classification

- The 20 amino acids found in proteins can be placed in five families based on the physical and chemical properties of their R groups:
  - Hydrophobic, aliphatic (6)  
    - Gavlip family
  - Hydrophobic, aromatic (3)  
    - PTT family
  - Special (hydrophobic/hydrophilic)(2)  
    - MC family
  - Hydrophilic, polar (4)  
    - Qnst family
  - Hydrophilic, charged (5)  
    - Dekrh family

<table>
<thead>
<tr>
<th>Name</th>
<th>3-letter</th>
<th>1-letter</th>
<th>Year discovered</th>
<th>% abundance in proteins</th>
<th>NOTES</th>
<th>pKa</th>
<th>Structure mnemonic device</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
<td>1820</td>
<td>7</td>
<td>Smallest, not chiral</td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
<td>1888</td>
<td>8</td>
<td>Foundational for ~10 other AA</td>
<td></td>
<td>methyl</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
<td>V</td>
<td>1856</td>
<td>7</td>
<td>Isopropyl</td>
<td></td>
<td>V-shaped</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
<td>1819</td>
<td>10</td>
<td>Most abundant, dominant</td>
<td></td>
<td>Ala + Val</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
<td>1904</td>
<td>6</td>
<td>Two chiral centers (L &amp; D)</td>
<td></td>
<td>Val + Me, 5-membered ring; same # as Val; 3C</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
<td>1901</td>
<td>5</td>
<td>Only imino acid (2° amine); special bonds in proteins; is modified by OH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
<td>1879</td>
<td>4</td>
<td>Aromatic, not chiral, most acidic</td>
<td></td>
<td>Phenyl+Ala, p-phenol+Ala</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
<td>1846</td>
<td>3</td>
<td>Aromatic, can ionize, amphipathic</td>
<td>10.6</td>
<td></td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
<td>1901</td>
<td>1</td>
<td>Aromatic &amp; fluorescent, least abundant</td>
<td></td>
<td>Indole+Ala</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
<td>1922</td>
<td>2</td>
<td>Most like straight-chain aliphatic can ionize; nucleophile</td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
<td>1899</td>
<td>2</td>
<td>Most like straight-chain aliphatic can ionize; nucleophile</td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td>Glutamine</td>
<td>Glu</td>
<td>Q</td>
<td>1883</td>
<td>4</td>
<td>Gets hydrolyzed to Glu, first isolated from asparagus</td>
<td></td>
<td>Amide of Glu</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
<td>1806</td>
<td>4</td>
<td>Gets hydrolyzed to Asp, isolated from serine</td>
<td></td>
<td>Amide of Asp</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
<td>1865</td>
<td>7</td>
<td>Gets hydrolyzed to Serin, polar cousin of Ala</td>
<td></td>
<td>hydroxy+Ala</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
<td>1935</td>
<td>6</td>
<td>Two chiral centers (L &amp; D)</td>
<td></td>
<td>Me+Ser</td>
</tr>
<tr>
<td>Aspartate</td>
<td>Glu</td>
<td>D</td>
<td>1868</td>
<td>5</td>
<td>Alpha-amino-succinate, most acidic</td>
<td>3.7</td>
<td>Ala+carboxyl</td>
</tr>
<tr>
<td>Glutamate</td>
<td>Asp</td>
<td>E</td>
<td>1866</td>
<td>7</td>
<td>Alpha-amino-glutamate</td>
<td>4.3</td>
<td>Ala-acetate</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
<td>1889</td>
<td>6</td>
<td>Only “bis” amino acid</td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
<td>1886</td>
<td>5</td>
<td>Most basic</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
<td>1896</td>
<td>2</td>
<td>Only physiological ionization</td>
<td>6.0</td>
<td>Ala+imidazole</td>
</tr>
</tbody>
</table>
Amino Acids Polymerize by Peptide Bonds

- Using full amino acid names:
  - seryglyctyrolylaamylleucine
- Using the three-letter code abbreviation:
  - Ser-Gly-Tyr-Ala-Leu
- For longer peptides (like proteins) the one-letter is used:
  - SGYAL

Peptide Bond

In a protein, the Ionizable Side Chains have altered pKₐ values

| Group                        | Acid       | Typical pKₐ | as AA | ΔpKₐ
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminal carboxyl</td>
<td>COOH → COO⁻ + H⁺</td>
<td>3.1</td>
<td>2.3</td>
<td>+0.8</td>
</tr>
<tr>
<td>Aspartic and glutamic acid</td>
<td>COOH → COO⁻ + H⁺</td>
<td>4.4</td>
<td>3.7</td>
<td>+0.7</td>
</tr>
<tr>
<td>Histidine</td>
<td>CH₃N⁺ + H⁺</td>
<td>6.5</td>
<td>6.0</td>
<td>+0.5</td>
</tr>
<tr>
<td>Terminal amino</td>
<td>NH₃⁺ → NH₃₂⁺ + H⁺</td>
<td>8.0</td>
<td>9.6</td>
<td>−1.6</td>
</tr>
<tr>
<td>Cysteine</td>
<td>SH → S⁻ + H⁺</td>
<td>8.5</td>
<td>10.5</td>
<td>−2.0</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>OH⁻ → O⁻ + H⁺</td>
<td>10.0</td>
<td>10.1</td>
<td>−0.1</td>
</tr>
<tr>
<td>Lysine</td>
<td>NH₂⁺ → NH₃⁺ + H⁺</td>
<td>10.0</td>
<td>10.5</td>
<td>−0.5</td>
</tr>
<tr>
<td>Arginine</td>
<td>HNN⁺ → NN₂⁺ + H⁺</td>
<td>12.0</td>
<td>12.5</td>
<td>−0.5</td>
</tr>
</tbody>
</table>

*pKₐ values depend on temperature, ionic strength, and the microenvironment of the ionizable group.
Chemical Environment Affects $pK_a$ Values

EXAMPLE: $\alpha$-carboxy group is much more acidic than in carboxylic acids.

IN PROTEINS, the environment can be much different than in bulk solution, and $pK_a$ values can change by several orders of magnitude.

Amino Acids Can Act as Buffers

**Question:** For one mole of completely protonated Gly, how many moles of base are required to titrate all the protons?

2

**Cation $\rightarrow$ Zwitterion $\rightarrow$ Anion**

**Question:** What does the curve of pH versus moles of base look like (i.e., the titration curve)?

**Curved**

Amino acids with uncharged side chains, such as glycine, have two $pK_a$ values:

- The $pK_a$ of the $\alpha$-carboxyl group is 2.34.
- The $pK_a$ of the $\alpha$-amino group is 9.6.

As buffers prevent change in pH close to the $pK_a$, glycine can act as a buffer in two pH ranges.

**Question:** Why is this not linear?
Question: For one mole of completely protonated Gly, how many moles of base are required to titrate all the protons?

2

Cation $\rightarrow$ Zwitterion $\rightarrow$ Anion

Question: What does the curve of pH versus moles of base look like (i.e., the titration curve)?

Curved

Amino acids with uncharged side chains, such as glycine, have two $pK_a$ values:

- The $pK_a$ of the $\alpha$-carboxyl group is 2.34.
- The $pK_a$ of the $\alpha$-amino group is 9.6.

As buffers prevent change in pH close to the $pK_a$, glycine can act as a buffer in two pH ranges.

Question: Why is this not linear?

### Henderson-Hasselbalch Eqn!

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

<table>
<thead>
<tr>
<th>% dissociated</th>
<th>[A−]/[HA]</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>1/1</td>
<td>$pK_a$</td>
</tr>
</tbody>
</table>

$pK_1 = 2.34$
Amino Acids Can Act as Buffers

- Those amino acids with ionizable side chains can be also titrated.

- Titration curves are now more complex, as each $pK_a$ has a buffering zone of 2 pH units.

Amino Acids Can Act as Buffers

- Those amino acids with ionizable side chains can be also titrated.

- Titration curves are now more complex, as each $pK_a$ has a buffering zone of 2 pH units.
Amino Acids Carry a Net Charge of Zero at a Specific pH value (the pI)

• The Isoelectric Point (equivalence point, pI) is the pH value where the net charge is ZERO.

• Zwitterions predominate at pH values between the pKₐ values of the amino and carboxyl groups.

• The exact value is the average of the two pKₐ values forming or titrating the zwitterion.

• At the pH equal to the pI:
  – AA is least soluble in water.
  – AA does not migrate in electric field.
  – AA does not bind well to other charged media/compounds

\[
pI = \frac{pK_{a1} + pK_{a2}}{2}
\]

How to Calculate pI

What is the pI of glycine?

• Identify the zwitterion (species that carries a net charge of zero).

• Identify the pKₐ value for the reaction that protonates the zwitterion.

• Identify the pKₐ value for the reaction that titrates a proton from the zwitterion.

• Take the average of these two pKₐ values.

\[
pI = \frac{pK_{a1} + pK_{a2}}{2}
\]

\[
pI = \frac{2.3 + 9.6}{2}
\]

\[
pI = 5.95
\]
How to Calculate the pI When the Side Chain Is Ionizable

- Identify species that carries a net zero charge.
- Identify the $pK_a$ value for the reaction that protonates the zwitterion. For His this occurs on the R-group ($pK_R$).
- Identify the $pK_a$ value for the reaction that titrates the next proton from the zwitterion. For His this is the $\alpha$-amino group ($pK_{NH_2}$).
- Take the average of these two $pK_a$ values.

$\text{pI} = \frac{pK_1 + pK_2}{2}$

What is the pI of histidine?

$pI = \frac{6.0 + 9.2}{2} = 7.6$

How to Calculate the pI of a peptide

Estimate the pI value of the following hexapeptide:

Phe-Lys-Asp-Cys-Thr-Tyr

Step 1: Determine the total positive charge on the peptide when all acidic and basic groups are fully protonated (at low pH).
Step 2: Determine the total negative charge on the peptide when all the groups are titrated (at high pH).
Step 3: List the $pK_a$ values of all acidic and basic groups in order from lowest ($pK_{\text{acid}}$) to highest.
Step 4: Calculate the pI as the average of the values for $pK_a$ value of the proton dissociation forming a neutral species from a $+1$ species, and $pK_a$ value of the proton dissociation forming a $-1$ species from the neutral species.

So for this peptide

Step 1: charge when fully protonated $+2$
Step 2: charge when fully de-protonated $-4$
Step 3: $pK_a$ values are:

9.0(N-term), 10.5(Lys), 3.9(Asp), 8.4(Cys), 10.5(Tyr), 3.5(C-term)

List from lowest to highest

<table>
<thead>
<tr>
<th>$pK_a$</th>
<th>3.5</th>
<th>3.9</th>
<th>8.4</th>
<th>9.0</th>
<th>10.5</th>
<th>10.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charges</td>
<td>$+2$</td>
<td>$+1$</td>
<td>$0$</td>
<td>$-1$</td>
<td>$-2$</td>
<td>$-3$</td>
</tr>
</tbody>
</table>

Step 4: The pI is $(3.9 + 8.4)/2 = 6.2$
Proteins and their pI Values

- **IN GENERAL**, if you take the % abundance of acidic and basic residues (Glu+Asp) and (Lys+Arg), you have 12 and 11%.
- So, there are more acidic residues than there are basic residues.
- Half way between the most basic of these acidic residues lies the pI values for most proteins; most are below the average of 4.3 (Glu) + 10.5 (Lys) ÷ 2 = 7.4.
- Therefore, given that there is 1% more (Glu+Asp) than (Lys+Arg), most proteins are slightly more acidic than physiological pH.
  - That doesn’t mean there are not many proteins that are very acidic (pI values <<7.4; negatively charged at pH 7.4):
    - Transcription factors
    - Pepsin
    - Ovalbumin
    - Serum albumin
  - Or, very basic (pI values >>7.4; positively charged at pH 7.4):
    - Cytochrome c
    - Lysozyme
    - Histones
    - Ribosomal proteins

Electrophoresis for Protein Analysis

- **Electrophoresis** is the migration of molecules in an electric field.
- **Electrophoresis** is one of the most commonly used analytical scale **separation** techniques
  - The electric field pulls proteins according to their charge.
- **Gel electrophoresis** adds a solid support in which the separation occurs. The gel matrix hinders mobility of proteins according to their size and shape.
  - The commonly used gels are either polyacrylamide (proteins) or agarose (nucleic acids).
  - Separation of proteins via electrophoresis is often called polyacrylamide gel electrophoresis, or PAGE.
- **For proteins to separate, they have to have a charge.**
Electrophoresis for Protein Analysis

Gels

Agarose

Polyacrylamide

Electrophoresis for Protein Analysis

Gels

Polyacrylamide
**Electrophoresis for Protein Analysis**

- For proteins to separate, they have to have a charge.
- The charge on a protein will depend on the pH.

But once they are moving, what does the velocity depend on?

\[
\text{Velocity} = \frac{E \cdot z}{f}
\]

where:

\[
f \propto \text{mass, shape, viscosity of media}
\]

\[
f = 6\pi \eta \cdot r
\]

where:

- \(\eta\) = coefficient of viscosity
- \(r\) = Stokes radius (mass and shape) [this is from actual radius and specific volume (cm³/g)]

This is essentially the charge:mass ratio if all proteins are roughly the same shape (globular).

**Electrophoresis for Protein Analysis**

- For proteins to separate, they have to have a charge.

- pH and pI dependence:
  - At pH near the pI, not much movement
  - At pH below the pI, proton concentration is higher, so charge becomes positive \(\oplus\)
  - At pH above the pI, protons will be titrated off, so the charge will become \(\ominus\)
Isoelectric Focusing Takes advantage of the $pI$ differences in Proteins for Separation

Can also be used to determine the $pI$ values

**Isoelectric Focusing**

A protein sample may be applied to one end of a gel strip with an immobilized $pH$ gradient. Or, a protein sample in a solution of ampholytes may be used to rehydrate a dehydrated gel strip.

**Ampholytes** are highly charged small MW polymer with variable $pI$ values. Due to their high $\frac{z}{r}$, they migrate rapidly setting up a buffered $pH$ gradient. Once they reach the $pH$ equal to their individual $pI$ values, they STOP migrating, thus creating an immobilized $pH$ gradient.

**Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS PAGE) Separates Proteins by Molecular Weight**

- SDS – sodium dodecyl sulfate – a detergent

  ![Sodium dodecyl sulfate (SDS)](image)

  Velocity = $\frac{E}{6\eta r}$

- SDS micelles bind to proteins and facilitate unfolding.
  - SDS gives all proteins a uniformly negative charge and shape (micelle)
  - The native shape is perturbed; de-natured.
  - SDS binds proteins at a constant ratio of mass (1.4g SDS/g protein), coating them with a negative charge.
  - So much charge is added that all proteins have the same charge:mass ratio, and the rate of movement will only depend on the sieving properties of the gel: small proteins will move farther.
SDS PAGE Separates Proteins by Molecular Weight

SDS-PAGE

SDS PAGE Can Be Used to Calculate the Molecular Weight of a Protein

- Myosin: 200,000
- β-Galactosidase: 116,250
- Glycogen phosphorylase b: 97,400
- Bovine serum albumin: 66,200
- Ovalbumin: 45,000
- Carbonic anhydrase: 31,000
- Soybean trypsin inhibitor: 21,500
- Lysozyme: 14,400

Unknown protein
SDS-PAGE + Isoelectric Focusing Can Separate nearly ALL the Proteins in *E. coli*