Lecture 5 (9/15/25)

OUTLINE

Amino Acids

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

Protein Purification

TODAY

- Reading: Ch3; 76–82, 87–89
- Homework #5

NEXT

- Reading: Ch3; 83–87, 89–90
  Ch1; Fig 7
  Ch9; 313–314
- Homework #6

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)
  - Hydrophilic, polar (4)
  - Hydrophilic, charged (5)

Special (Sulfur)

Amino Acids: Classification

Hydrophobic
/Hydrophilic (Cys)

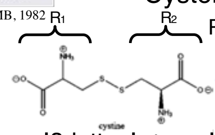
AA	Hydropathy index (ΔG)
Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8
Tyr	-1.3
Met	1.9
Cys	2.5
Trp	-0.9

Kyte and Doolittle, JMB, 1982

- Cysteine can ionize:

  Cys–SH ⇌ Cys–S⁻ (thiolate anion) + H⁺

- Cysteine can form disulfide bonds:



R_1–S–S–R_2 ⇌ R_1–SH + R_2–SH

Oxidized — Cystine (1810)

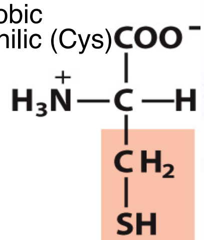
Reduced — Cysteine

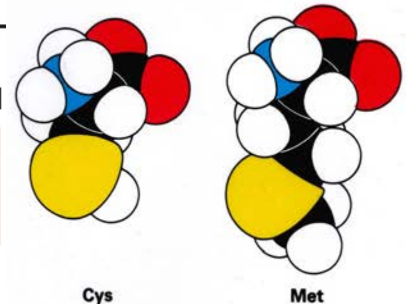
Name	3-letter	1-letter	Year discovered	% abundance in proteins	NOTES	pK_a	Structure mnemonic device
Methionine ✓	Met	M	1922	2	*Most like straight-chain aliphatic*		*Ala+Me/thioether*
Cysteine	Cys	C	1899	2	*can ionize; nucleophile*	*10.5*	*Ala+SH (thiol)*

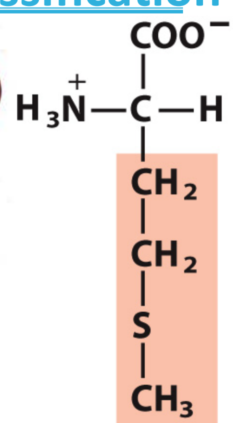
Special (Sulfur)

Amino Acids: Classification

Hydrophobic
/Hydrophilic (Cys)







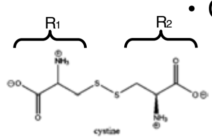
MC family

(master of ceremony)

- Cysteine can ionize:

  Cys–SH ⇌ Cys–S⁻ (thiolate anion) + H⁺

- Cysteine can form disulfide bonds:



R_1–S–S–R_2 ⇌ R_1–SH + R_2–SH

Oxidized — Cystine (1810)

Reduced — Cysteine

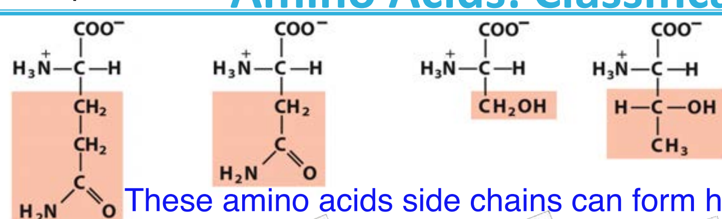
Name	3-letter	1-letter	Year discovered	% abundance in proteins	NOTES	pK_a	Structure mnemonic device
Methionine ✓	Met	M	1922	2	*Most like straight-chain aliphatic*		*Ala+Me/ether*
Cysteine	Cys	C	1899	2	*can ionize; nucleophile*	*10.5*	*Ala+SH*

Amino Acids: Classification

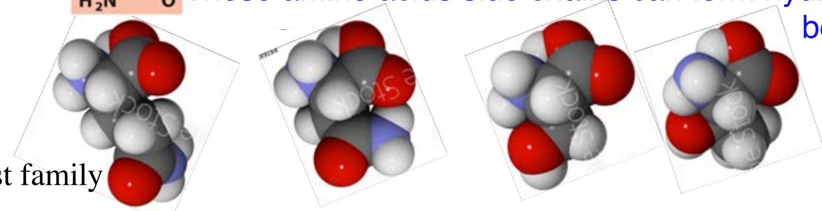
- The 20 amino acids found in proteins can be placed in five families based on the physical and chemical properties of their R groups:
  - Hydrophobic, aliphatic (6) Gavlip family
  - Hydrophobic, aromatic (3) PTT family
  - Special (hydrophobic/hydrophilic)(2) MC family
  - Hydrophilic, polar (4)
 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

Hydrophilic, polar

Amino Acids: Classification



These amino acids side chains can form hydrogen bonds.



Qnst family

Gln Amides Asn

Ser Alcohols Thr

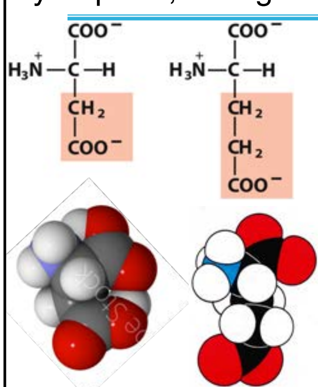
Name	3-letter	1-letter	Year discovered	% abundance in proteins	NOTES	Structure mnemonic device
Glutamine	Gln	Q	1883	4	*Glx; gets hydrolyzed to Glu*	*Amide of Glu*
Asparagine	Asn	N	1806	4	*First isolated from asparagus Asx; gets hydrolyzed to Asp*	*Amide of Asp*
Serine	Ser	S	1865	7	*Isolated from* Sericin, *polar cousin of Ala*	*hydroxyl+Ala*
Threonine ✓	Thr	T	1935	6	*Two chiral centers (L & D)*	*Me+Ser*

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)

Amino Acids: Classification

Hydrophilic, Charged



Negative

Name	3-letter	1-letter	Year discovered	% abundance in proteins	NOTES	pK_a	Structure mnemonic device
Aspartate	Asp	D	1868	5	*α-amino-succinate; Most acidic*	3.7	*Ala+carboxyl*
Glutamate	Glu	E	1866	7	*α-amino-glutarate*	4.3	*Ala+acetate*

Amino Acids: Naming Dicarboxylic Acids

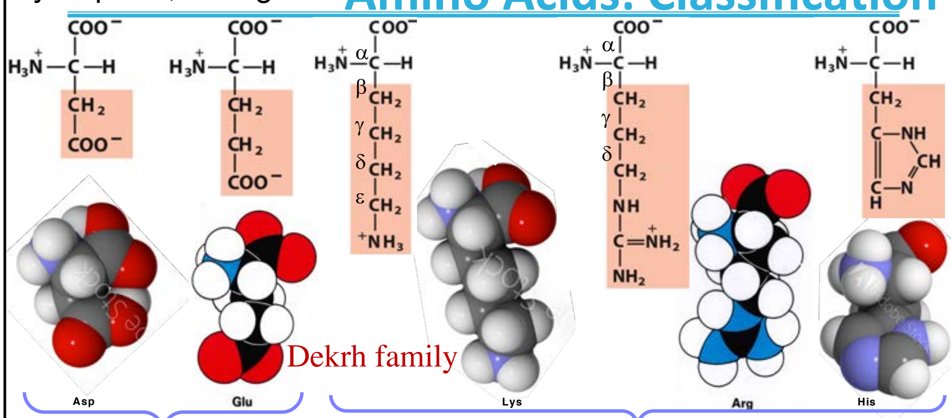
Oh My Such Good Apple Pie!!

Dicarboxylic acids: OMSGAP

Rubric	Name (conjugate base)	Name of Acid	Structure (conjugate base)	X = $(CH_2)_z$ in ^-OOC-X-COO^- ; z=?
Oh	oxalate	oxalic	^-OOC-COO^-	0
My	malate	malic	^-OOC-CH_2-COO^-	1
Such	succinate	succinic	^-OOC-CH_2-CH_2-COO^-	2
Good	glutarate	glutaric	^-OOC-CH_2-CH_2-CH_2-COO^-	3
Apple	adipate	adipic	^-OOC-CH_2-CH_2-CH_2-CH_2-COO^-	4
Pie	pimelate	pimelic	^-OOC-CH_2-CH_2-CH_2-CH_2-CH_2-COO^-	5

Hydrophilic, Charged

Amino Acids: Classification



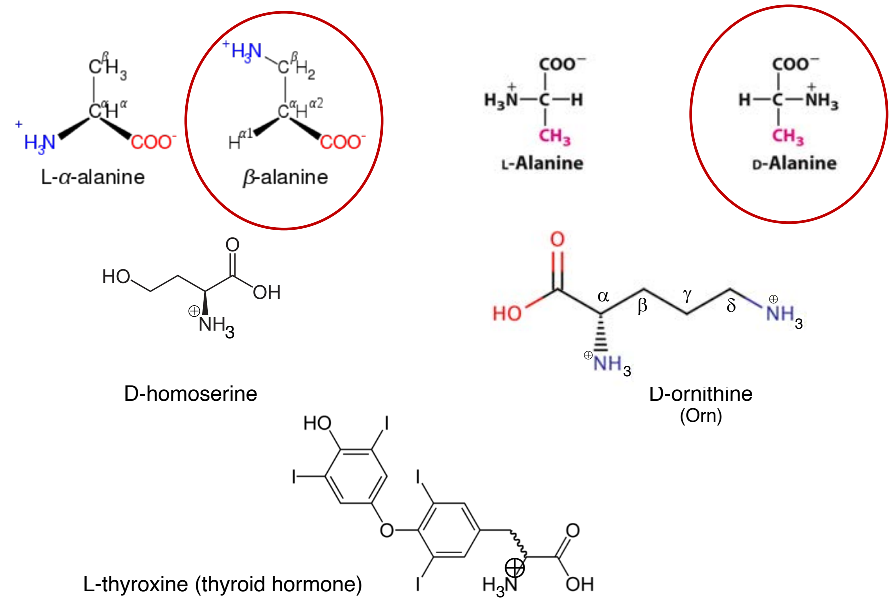
Name	3-letter	1-letter	Year discovered	% abundance in proteins	NOTES	pK_a	Structure mnemonic device
Aspartate	Asp	D	1868	5	*α-amino-succinate; Most acidic*	3.7	*Ala+carboxyl*
Glutamate	Glu	E	1866	7	*α-amino-glutarate*	4.3	*Ala+acetate*
Lysine ✓	Lys	K	1889	6	*Only "bis" amino acid*	10.5	*ε-amino*
Arginine ✓	Arg	R	1886	5	*Most basic*	12.5	*δ-guanidino*
Histidine ✓	His	H	1896	1	*Only physiological ionization*	6.0	*Ala+imidazole*

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

Name	3-letter	1-letter	Year discovered	% abundance in proteins	NOTES	pK_a	Structure mnemonic device
Glycine	Gly	G	1820	7	*Smallest, not chiral*		*H*
Alanine	Ala	A	1888	8	*Foundational for ~10 other AA*		*Methyl*
Valine ✓	Val	V	1856	7	*isopropyl*		*V-shaped*
Leucine ✓	Leu	L	1819	10	*Most abundant, dominant*		*Ala + Val*
Isoleucine ✓	Ile	I	1904	6	*Two chiral centers (L & D)*		*Val + Me*
Proline	Pro	P	1901	5	*Only imino acid (2° amine); special bonds in proteins; is modified by OH*		*5-membered ring; same # as Val; 3C*
Phenylalanine ✓	Phe	F	1879	4	*aromatic*		*Phenyl+Ala*
Tyrosine	Tyr	Y	1846	3	*aromatic, can ionize; amphipathic*	10.1	*p-phenol+Ala*
Tryptophan ✓	Trp	W	1901	1	*aromatic & fluorescent; least abundant*		*Indole+Ala*
Methionine ✓	Met	M	1922	2	*Most like straight-chain aliphatic*		*Ala+Me/ether*
Cysteine	Cys	C	1899	2	*can ionize; nucleophile*	10.5	*Ala+SH*
Glutamine	Gln	Q	1883	4	*Glx; gets hydrolyzed to Glu*		*Amide of Glu*
Asparagine	Asn	N	1806	4	*First isolated from asparagus Asx; gets hydrolyzed to Asp*		*Amide of Asp*
Serine	Ser	S	1865	7	*Isolated from Sericin, polar cousin of Ala*		*hydroxyl+Ala*
Threonine ✓	Thr	T	1935	6	*Two chiral centers (L & D)*		*Me+Ser*
Aspartate	Glu	D	1868	5	*α-amino-succinate; Most acidic*	3.7	*Ala+carboxyl*
Glutamate	Asp	E	1866	7	*α-amino-glutarate*	4.3	*Ala+acetate*
Lysine ✓	Lys	K	1889	6	*Only "bis" amino acid*	10.5	*ε-amino*
Arginine ✓	Arg	R	1886	5	*Most basic*	12.5	*δ-guanidino*
Histidine ✓	His	H	1896	1	*Only physiological ionization*	6.0	*Ala+imidazole*

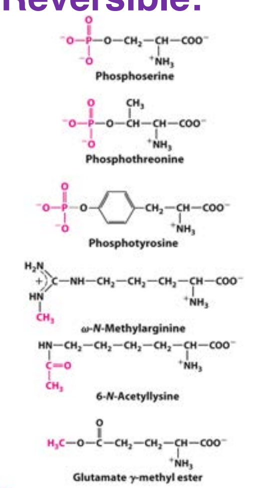
Other Cool Amino Acids

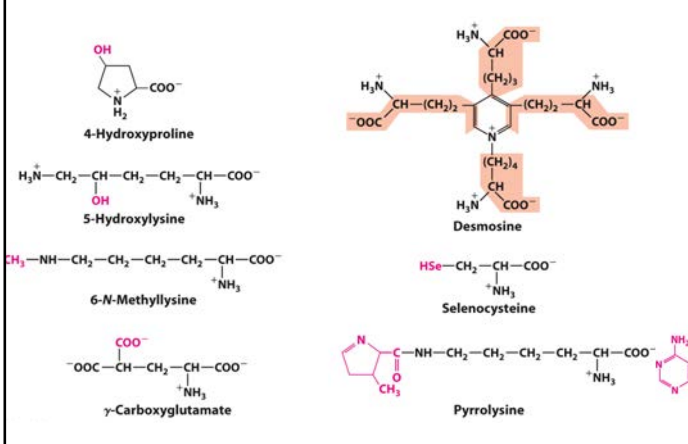


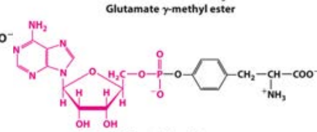
Modified Amino Acids Found in Proteins

- Not incorporated by ribosomes
  - –except for selenocysteine
- Arise by posttranslational modifications of proteins
- Reversible modifications, especially phosphorylation, are important in regulation and signaling.

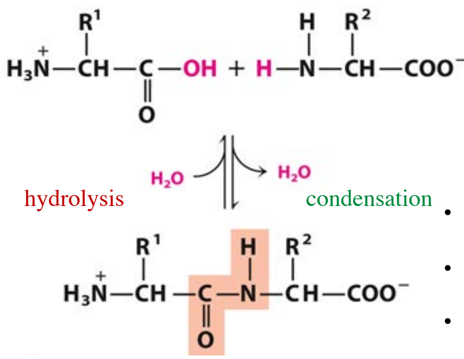
Reversible:

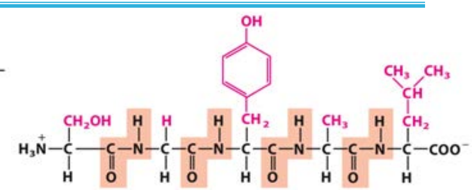




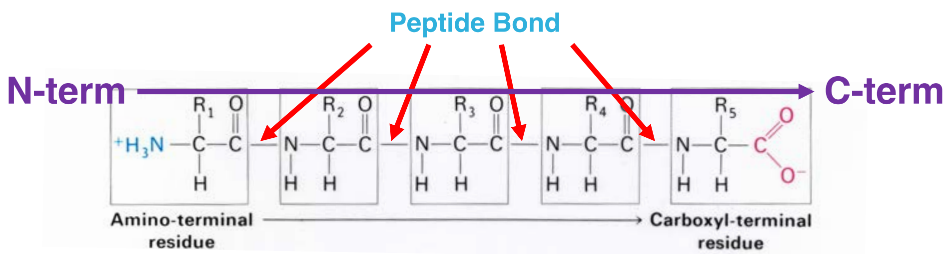


Amino Acids Polymerize by Peptide Bonds





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



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

pK values of ionizable groups in proteins

Group	Acid ⇌ base + H⁺	Typical pK*	as AA	ΔpKa
Terminal carboxyl	$-COOH \rightleftharpoons -COO^- + H^+$	3.1	2.3	+0.8
Aspartic and glutamic acid	$-COOH \rightleftharpoons -COO^- + H^+$	4.4	3.7 4.3	+0.7 +0.1
Histidine	$-CH_2$-imidazolium (^+HN, NH) ⇌ $-CH_2$-imidazole (N, NH) + H^+	6.5	6.0	+0.5
Terminal amino	$-NH_3^+ \rightleftharpoons -NH_2 + H^+$	8.0	9.6	−1.6
Cysteine	$-SH \rightleftharpoons -S^- + H^+$	8.5	10.5	−2.0
Tyrosine	phenyl–OH ⇌ phenyl–O^- + H^+	10.0	10.1	−0.1
Lysine	$-NH_3^+ \rightleftharpoons -NH_2 + H^+$	10.0	10.5	−0.5
Arginine	$-N(H)-C(=NH_2^+)NH_2 \rightleftharpoons -N(H)-C(=NH)NH_2 + H^+$	12.0	12.5	−0.5

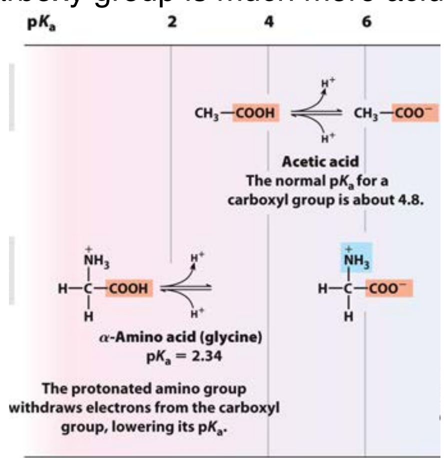
*pK values depend on temperature, ionic strength, and the microenvironment of the ionizable group.

Become more basic (like H⁺ more)

Become more acidic (like H⁺ less)

Chemical Environment Affects pK_a Values

EXAMPLE: α-carboxy group is much more acidic than in carboxylic acids.



In proteins, the α-amino group is in the peptide bond and loses it's charge effect so that the carboxylate becomes more like acetic acid (pK_a = 3.1)

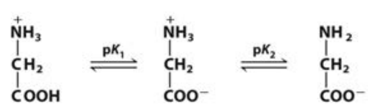
IN PROTEINS, the environment can be much different than in bulk solution, and pK_a values can change by several orders of magnitude.

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 → Zwitterion → Anion



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

Curved

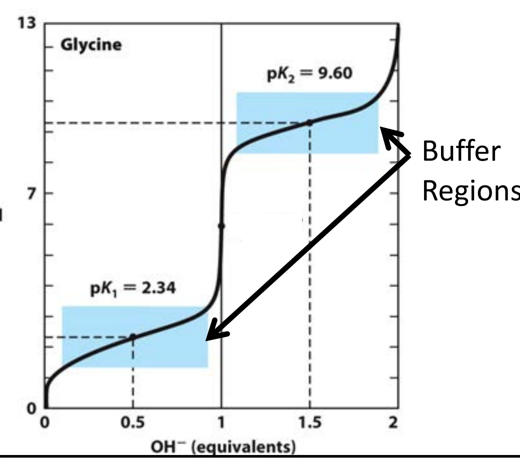
Amino acids with uncharged side chains, such as glycine, have two pK_a values:

•The pK_a of the α-carboxyl group is 2.34.

•The pK_a of the α-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?

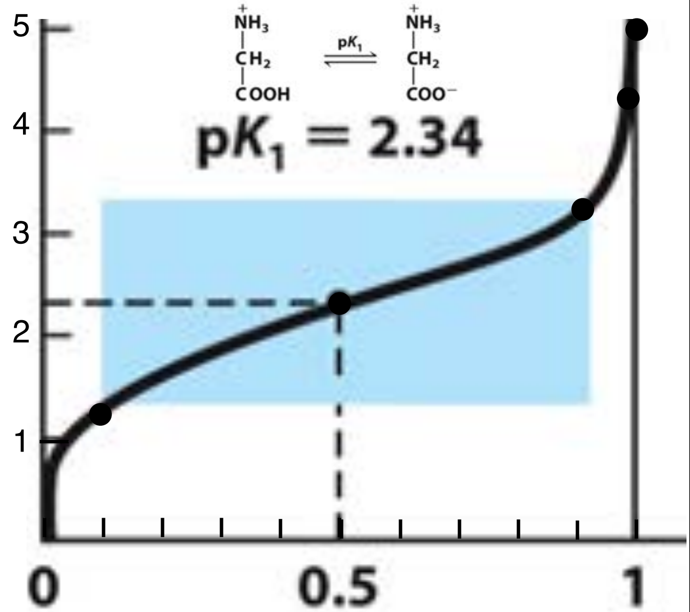


Amino Acids Can Act as Buffers

Question: Why is this not linear? **Henderson-Hasselbalch Eqn!**

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

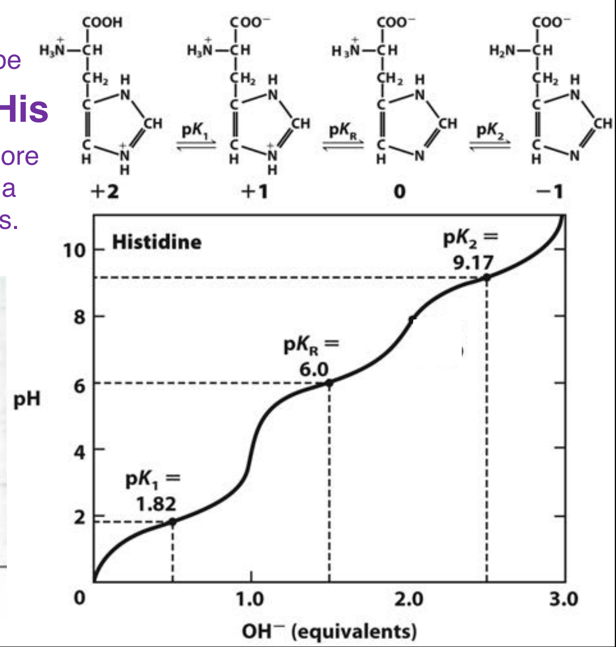
% dissociated	$[A^-]/[HA]$	pH
50	1/1	pK_a
10	10/90	pK_a– 0.95
1	1/99	pK_a–2
0.1	1/999	pK_a–3
90	90/10	pK_a+ 0.95
99	99/1	pK_a+2
99.9	999/1	pK_a+3



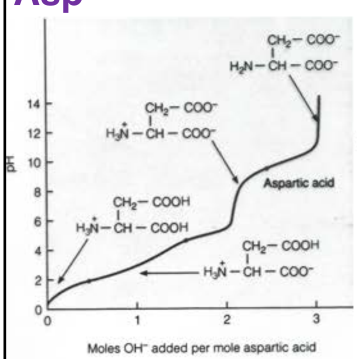
Amino Acids Can Act as Buffers

- 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.

His



Asp



Amino Acids Carry a Net Charge of Zero at a Specific pH value (the pI)

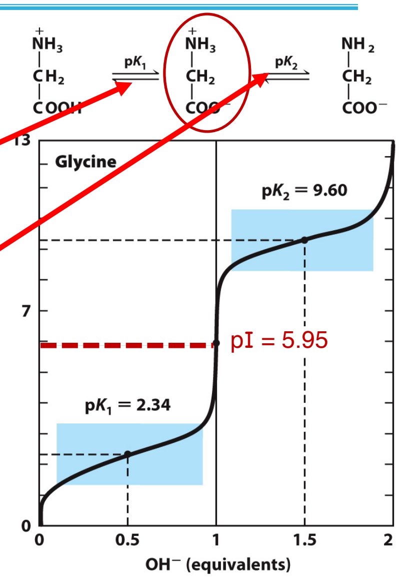
- The Isoelectric Point (equivalence point, pI) is the pH value where the net charge is ZERO.
- Zwitterions predominate at pH values between the pK_a values of the amino and carboxyl groups.
- The exact value is the average of the two pK_a values forming or titrating the zwitterion.
- At the pH equal 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

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

How to Calculate pI

What is the pI of glycine?

- Identify the zwitterion (species that carries a net charge of zero).
- Identify the pK_a value for the reaction that protonates the zwitterion.
- Identify the pK_a value for the reaction that titrates a proton from the zwitterion.
- Take the average of these two pK_a values.



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

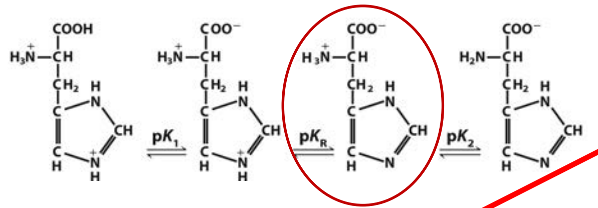
$$pI = (2.3 + 9.6)/2$$

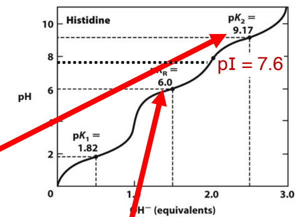
$$pI = 5.95$$

How to Calculate the pI When the Side Chain Is Ionizable

His

- 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 α-amino group (pK_{NH_2}).
- Take the average of these two pK_a values.

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

$$pI = (6.0 + 9.2)/2$$

What is the pI of histidine?

$$pI = 7.6$$

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_{a1}) 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.

<u>So for this peptide</u>
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

pKa		3.5		3.9		8.4		9.0		10.5		10.5	
Charges	+2	⇆	+1	⇆	0	⇆	-1	⇆	-2	⇆	-3	⇆	-4

Step 4: The pI is (3.9 + 8.4)/2 = 6.2

Proteins and their pI Values

	Cation →	Zwitterion →	Anion
pH at:	<pI	pI	>pI

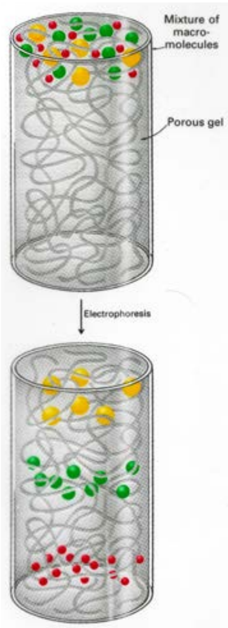
- IN GENERAL, if you take the % abundance of acidic and basic residues (Glu+Asp) and (Lys+Arg), you have ~12 and ~11%.
- So, there are slightly more acidic residues than there are basic residues.
- Half way between the most basic of these acidic residues lies the pI values for most proteins; most are 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

- Like AA, for proteins, at the pH equal to the pI:
  - Protein is least soluble in water, could precipitate.
  - Protein does not migrate in electric field.

Electrophoresis for Protein Analysis

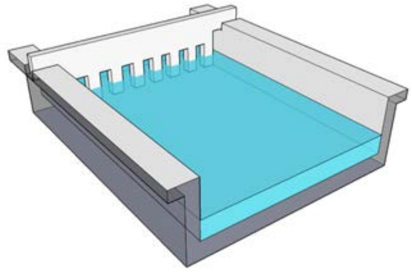
- Electrophoresis is the migration of molecules in an electric field.
- Electrophoresis is one of the most commonly used analytical scale **separation** techniques
  - The electric field pulls proteins according to their charge.
- Gel electrophoresis adds a solid support in which the separation occurs. 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

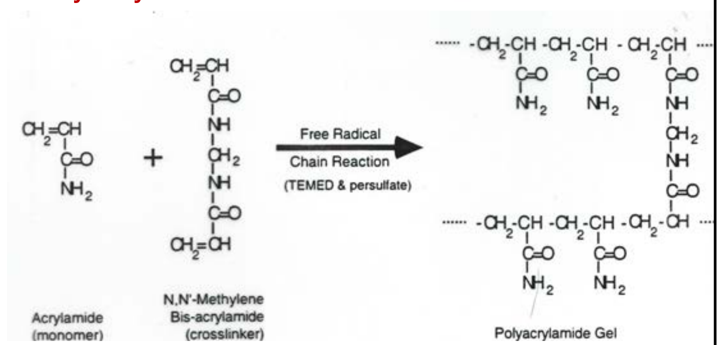


Gels

Agarose

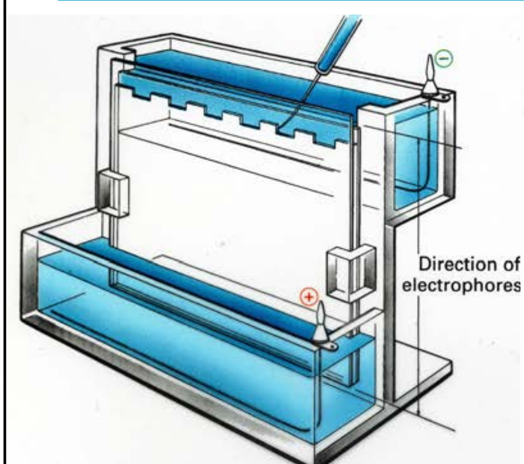


Polyacrylamide

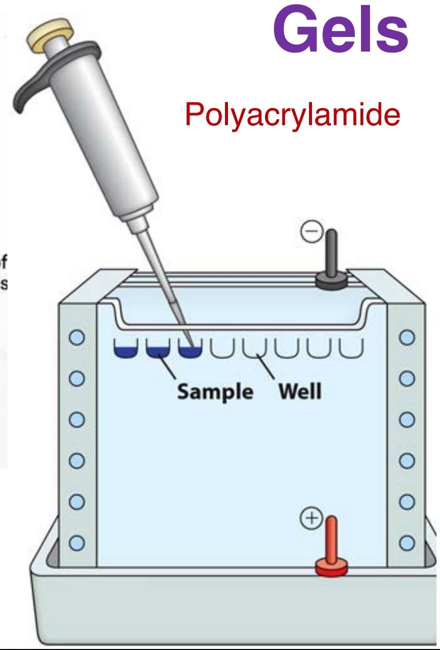


Electrophoresis for Protein Analysis

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?

$$\text{Velocity} = \frac{E \bullet z}{f}$$

Electric field

charge on the protein

friction

$f \propto$ mass, shape, viscosity of media

$f = 6\Pi \bullet \boldsymbol{\eta} \bullet r$

where:

$\boldsymbol{\eta}$ = coefficient of viscosity

r = Stokes radius (mass and shape)

[this is from actual radius and specific volume (cm^3/g)]

$$\text{Velocity} = \frac{E}{6\Pi \bullet \boldsymbol{\eta}} \bullet \frac{z}{r}$$

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 ⊕
  - At pH above the pI, protons will be titrated off, so the charge will become ⊖

Recall:

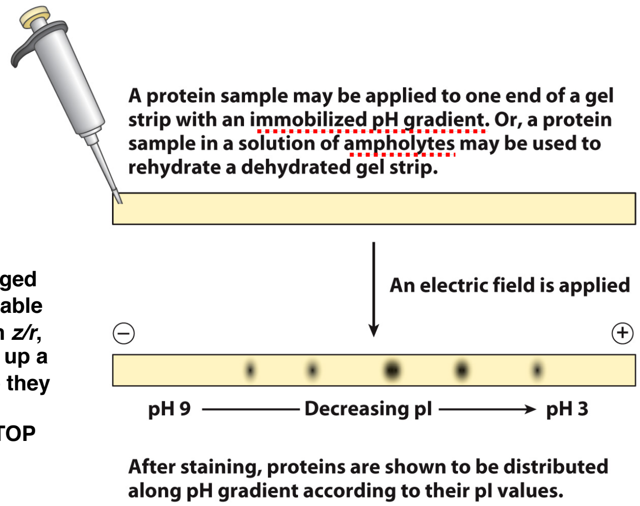
	Cation →	Zwitterion →	Anion
pH at:	<pI	pI	>pI

Isoelectric Focusing Takes advantage of the pI differences in Proteins for Separation

Can also be used to determine the pI values

Isoelectric Focusing

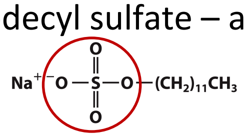
$$\text{Velocity} = \frac{E}{6\Pi \cdot \eta} \cdot \frac{z}{r}$$



Ampholytes are highly charged small MW polymer with variable pI values. Due to their high *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



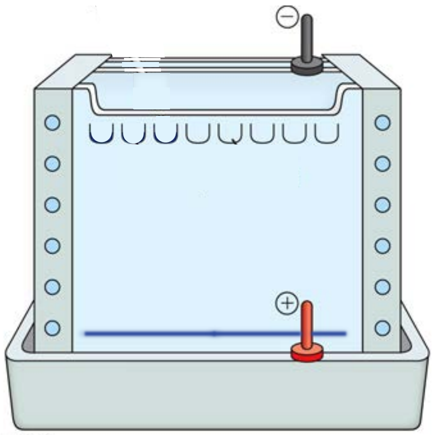
Sodium dodecyl sulfate (SDS)

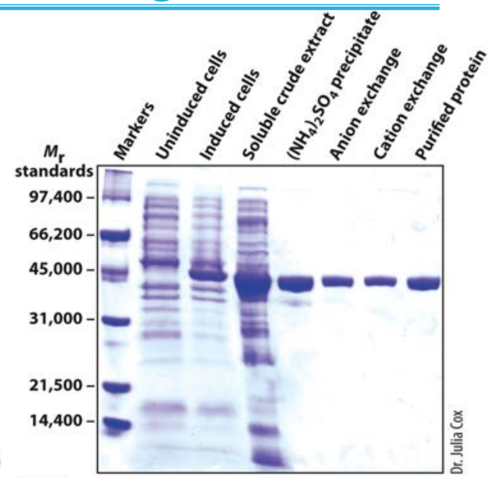
$$\text{Velocity} = \frac{E}{6\Pi \cdot \eta} \cdot \frac{z}{r}$$

- 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

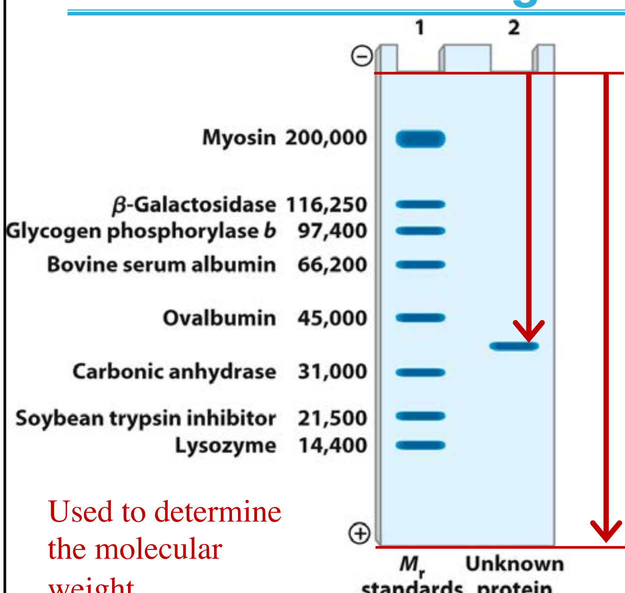
SDS-PAGE

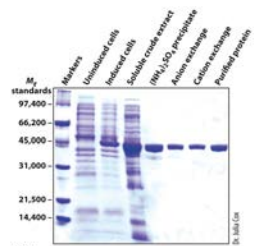


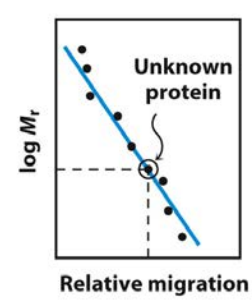


Used to detect purity or purification

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

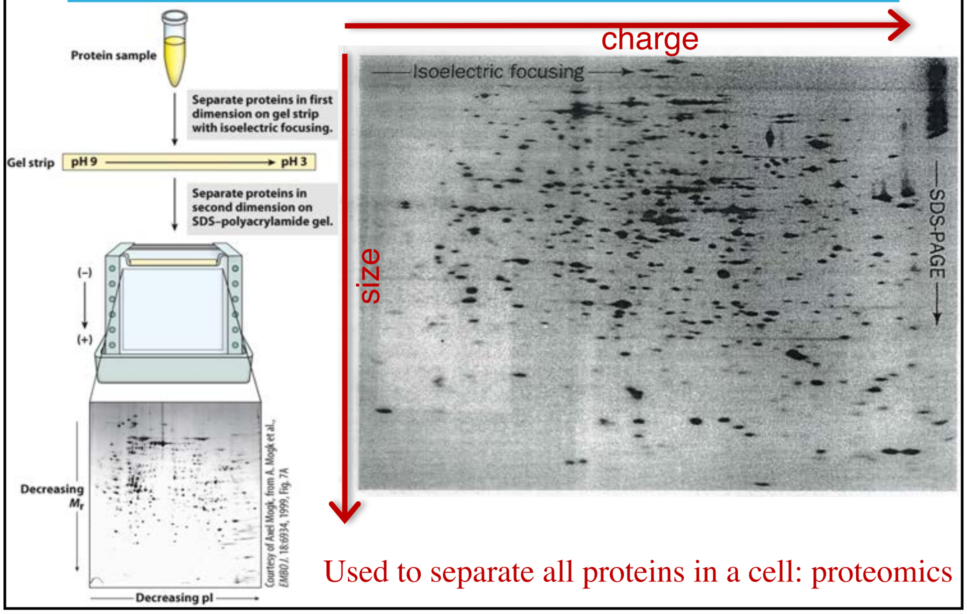






Used to determine the molecular weight

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



Used to separate all proteins in a cell: proteomics

Protein Purification

Protein Purification

Proteins are separated from each other (along with other macromolecules) due to the vast variability they have. The basis of the separation can be put into 4 categories:

- Size, shape, density } Hydrodynamic properties
- Charge
- Solubility } Chemical properties
- Binding characteristics } Biological properties

Protein Purification Procedures

TABLE 5-2 Protein Purification Procedures

Protein Characteristic	Purification Procedure	
Solubility	Salting out	
Ionic Charge	Ion exchange chromatography Electrophoresis Isoelectric focusing	Lab
Polarity	Hydrophobic interaction chromatography	
Size	Gel filtration chromatography SDS-PAGE Centrifugation	Lab Lab Lab
Binding Specificity	Affinity chromatography	Lab

Before you can separate "your favorite protein (YFP)" from all the thousands of others, you need a way to "see" it........An Assay!

- enzymes ➞ activity assay
- binding proteins ➞ binding assay
- other proteins ➞ immunoblot or ELISA

How do you monitor the purification?

- Specific Activity = $\dfrac{\text{Activity of YFP}}{\text{Total protein}}$