Lecture 5 (9/16/20)

- Reading: Ch3; 82–87, 92–94
- Problems: Ch3 (text); 2,3,4,7,11,12
  Ch3 (study guide); **4,5,6,8,9**
  Ch2 (text); 34

NEXT

- Reading: Ch3; 89–93, 95-96
  Ch1; Fig 1-9
  Ch9, 332-333
- Problems: Ch3 (text); 9,10,14,15,22
  Ch1 (text); 9

Lecture 5 (9/16/20)

OUTLINE

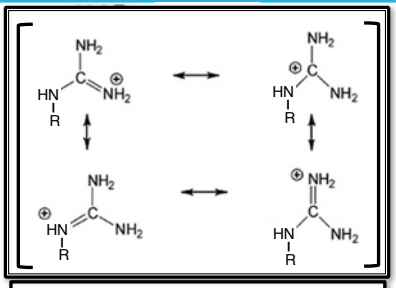
Amino Acids
- Definition
- The 4 S's
- Common Properties
- Five Classes
  - Hydrophobic–aliphatic [6]
  - Hydrophobic–aromatic [3]
  - Special–sulfur [2]
  - Hydrophilic–polar [4]
  - Hydrophilic–charged [5]
- Other amino acids
- Linking amino acids
- Acid/base properties
  - Titrations
  - Isoelectric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

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)

Hydrophilic, Charged

Amino Acids: Classification

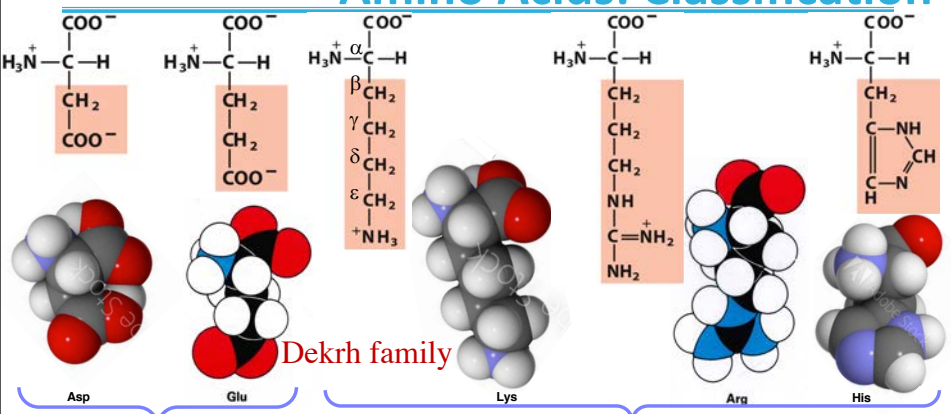


As the electrons creating the double bond become de-localized, the positive charge also becomes delocalized

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	2	*Only physiological ionization*	6.0	*Ala+imidazole*

Hydrophilic, Charged

Amino Acids: Classification



Dekrh family

Negative

Positive

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	2	*Only physiological ionization*	6.0	*Ala+imidazole*

Amino Acids: Naming Dicarboxylic Acids

Dicarboxylic acids: OMSGAP

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

Amino Acids: Naming Dicarboxylic Acids

Dicarboxylic acids: OMSGAP

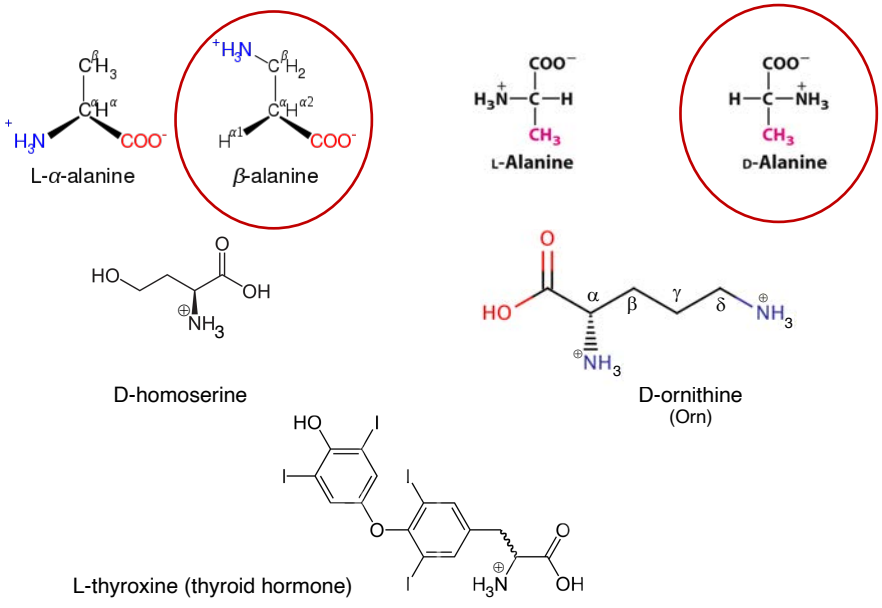
Rubric	Name (conjugate base)	Name of Acid	Structure (conjugate base)	$X = (CH_2)_z$ in $^-OOC\text{-}X\text{-}COO^-$; z=?
Oh	oxalate	oxalic	$^-OOC\text{-}COO^-$	0
My	malate	malic	$^-OOC\text{-}CH_2\text{-}COO^-$	1
Such	succinate	succinic	$^-OOC\text{-}CH_2\text{-}CH_2\text{-}COO^-$	2
Good	glutarate	glutaric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

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

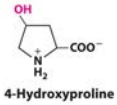
Name	3-letter	1-letter	Year discovered	% abundance in proteins	NOTES	pK_a	Structure mnemonic device
Glycine	Gly	G	1820	7	*Smallest, not chiral*		*H*
Alanine	Ala	A	1888	8	*Foundational for ~10 other AA*		*Methyl*
Valine ✓	Val	V	1856	7	*isopropyl*		*V-shaped*
Leucine ✓	Leu	L	1819	10	*Most abundant, dominant*		*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	2	*Only physiological ionization*	6.0	*Ala+imidazole*

Other Cool Amino Acids

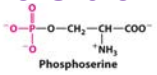


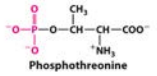
Modified Amino Acids Found in Proteins

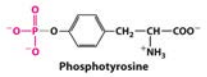
- Not incorporated by ribosomes
  - –except for selenocysteine
- Arise by posttranslational modifications of proteins
- Reversible modifications, especially phosphorylation, are important in regulation and signaling.

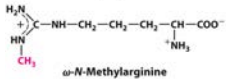

4-Hydroxyproline

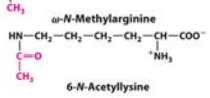
Reversible:

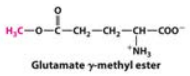

Phosphoserine

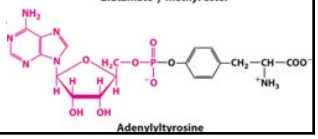

Phosphothreonine


Phosphotyrosine


ω-*N*-Methylarginine

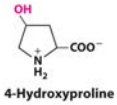

6-*N*-Acetyllysine

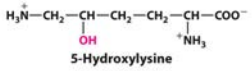

Glutamate γ-methyl ester

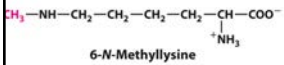

Adenylyltyrosine

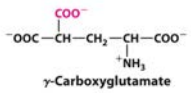
Modified Amino Acids Found in Proteins

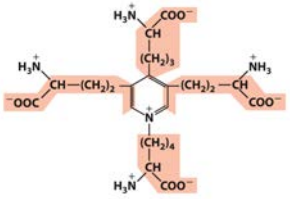
- Not incorporated by ribosomes
  - –except for selenocysteine
- Arise by posttranslational modifications of proteins
- Reversible modifications, especially phosphorylation, are important in regulation and signaling.

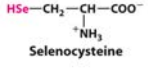

4-Hydroxyproline

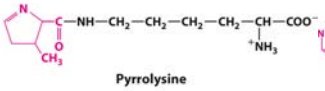

5-Hydroxylysine


6-*N*-Methyllysine

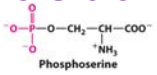

γ-Carboxyglutamate

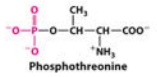

Desmosine

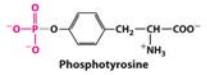

Selenocysteine

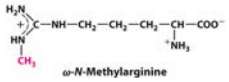

Pyrrolysine

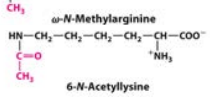
Reversible:

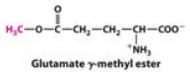

Phosphoserine

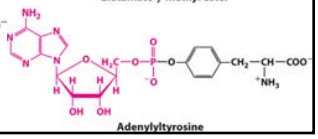

Phosphothreonine


Phosphotyrosine

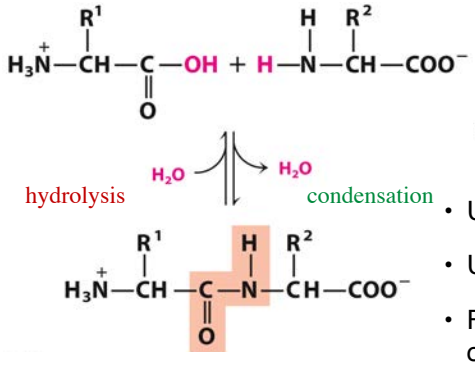

ω-*N*-Methylarginine

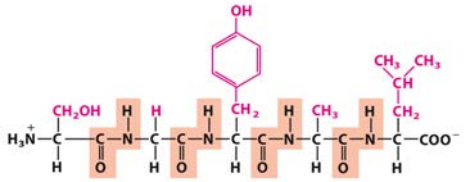

6-*N*-Acetyllysine


Glutamate γ-methyl ester

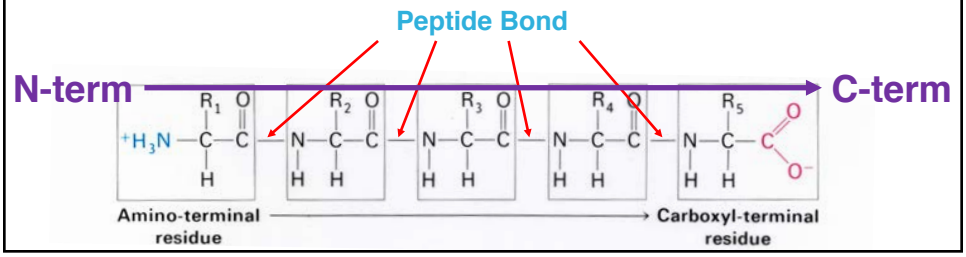

Adenylyltyrosine

Amino Acids Polymerize by Peptide Bonds



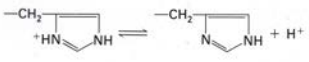
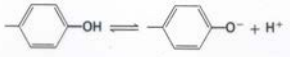
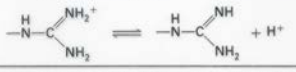


- Using full amino acid names:
  - serylglycyltyrosylalanylleucine
- Using the three-letter code abbreviation:
  - Ser-Gly-Tyr-Ala-Leu
- For longer peptides (like proteins) the one-letter code can be used:
  - SGYAL



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

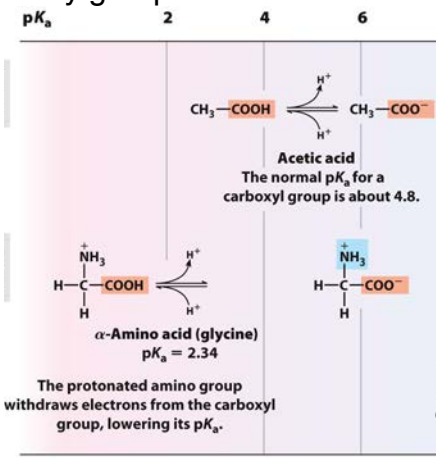
pK values of ionizable groups in proteins

Group	Acid ⇌ base + H^+	Typical pK*	as AA	ΔpK_a
Terminal carboxyl	—COOH ⇌ —COO^- + H^+	3.1	2.3	+0.8
Aspartic and glutamic acid	—COOH ⇌ —COO^- + H^+	4.4	3.7 4.3	+0.7 +0.1
Histidine		6.5	6.0	+0.5
Terminal amino	—NH_3^+ ⇌ —NH_2 + H^+	8.0	9.6	−1.6
Cysteine	—SH ⇌ —S^- + H^+	8.5	10.5	−2.0
Tyrosine		10.0	10.1	−0.1
Lysine	—NH_3^+ ⇌ —NH_2 + H^+	10.0	10.5	−0.5
Arginine		12.0	12.5	−0.5

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

Chemical Environment Affects pK_a Values

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



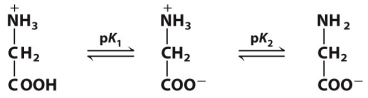
IN PROTEINS, the environment can be much different than in bulk solution, and pK_a values can change by several orders of magnitude.

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 → Zwitterion → Anion



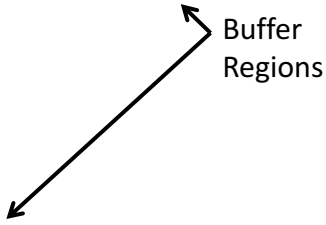
Question: What does the curve of pH *versus* moles of base look like (i.e., the titration curve)?

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.



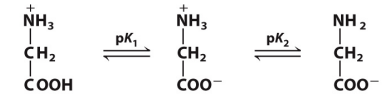
Question: Why is this not linear?

Amino Acids Can Act as Buffers

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

2

Cation → Zwitterion → Anion



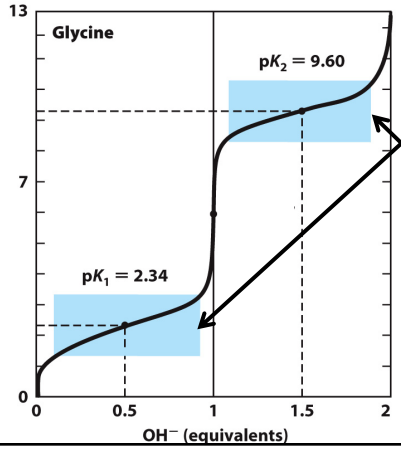
Question: What does the curve of pH *versus* moles of base look like (i.e., the titration curve)?

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.



Buffer Regions

Question: Why is this not linear?

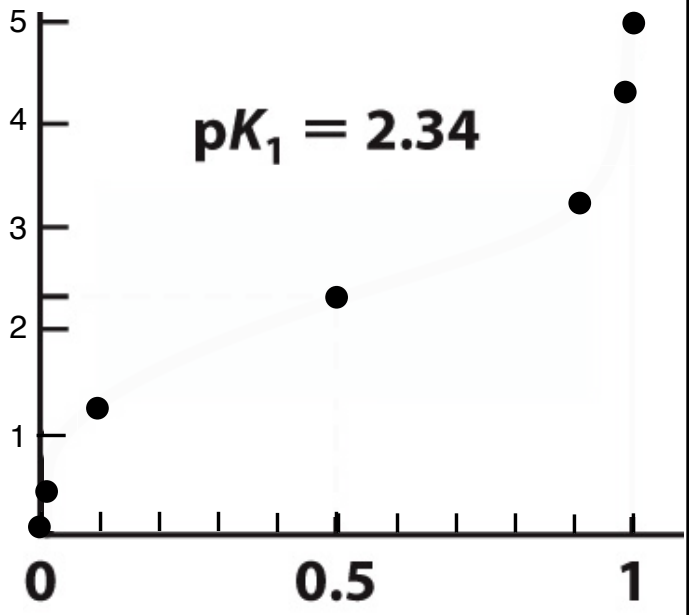
Amino Acids Can Act as Buffers

Question: Why is this not linear?

Henderson-Hasselbalch Eqn!

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

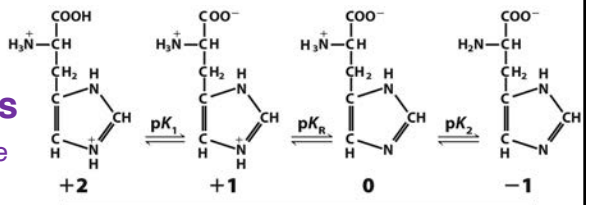
% dissociated	[A⁻]/[HA]	pH
50	1/1	pK_a



Amino Acids Can Act as Buffers

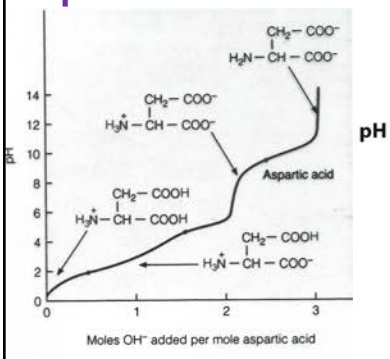
- Those amino acids with Ionizable side chains can be also titrated.

His



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

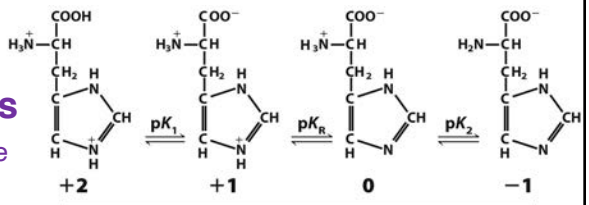
Asp



Amino Acids Can Act as Buffers

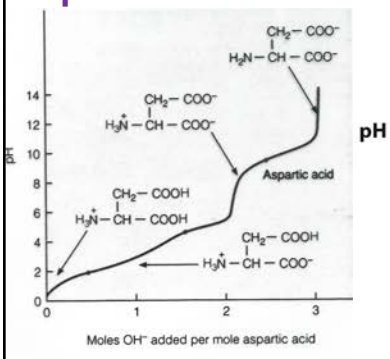
- Those amino acids with Ionizable side chains can be also titrated.

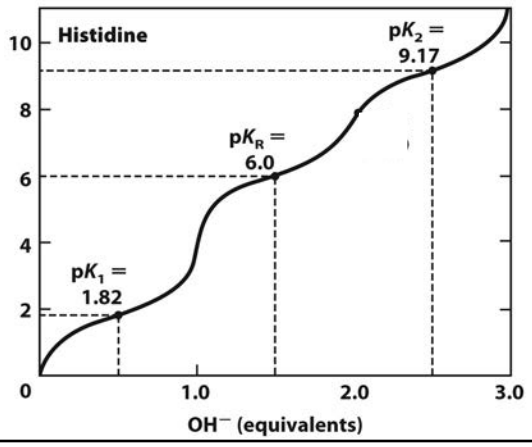
His



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

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.

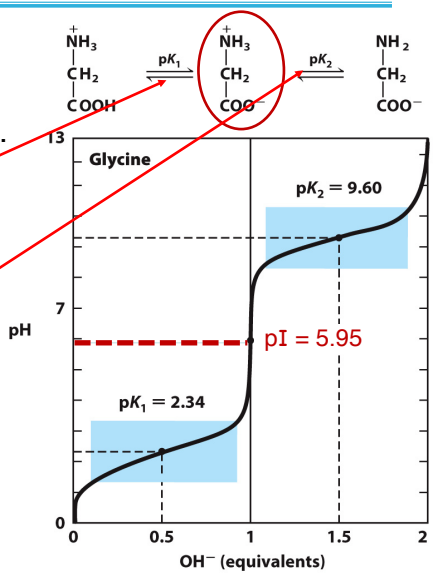
$$pI = \frac{pK_1 + pK_2}{2}$$

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 the next proton from the zwitterion.
- Take the average of these two pK_a values.

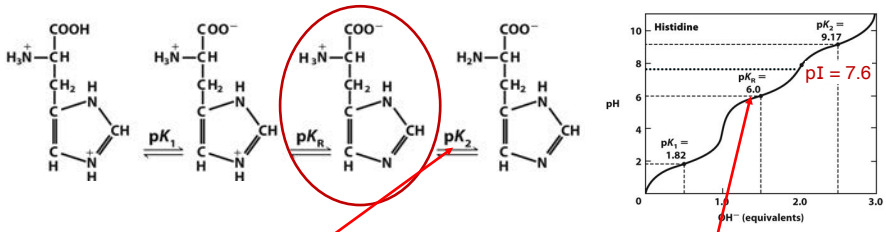
$$pI = \frac{pK_1 + pK_2}{2}$$
$$pI = (2.3 + 9.6)/2$$
$$pI = 5.95$$



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.

What is the pI of histidine?

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

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

$$pI = 7.6$$

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 more acidic residues than there are basic residues.
- Half way between 4.3 (Glu) and 10.5 (Lys) is 7.4, so its not the differences in pK_a values.
- 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:

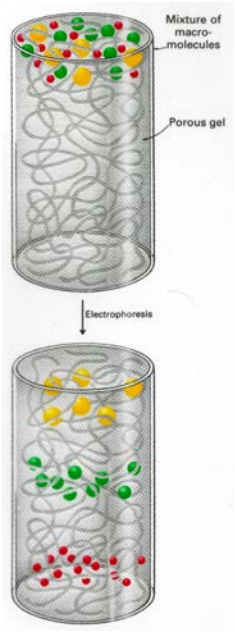
– Protein is least soluble in water, could precipitate.

– Protein does not migrate in electric field.

Electrophoresis for Protein Analysis

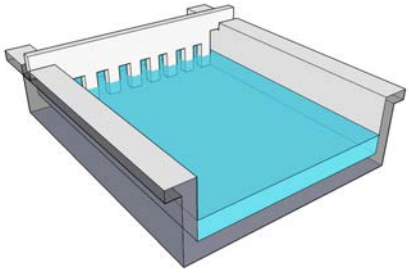
- Electrophoresis is the migration of molecules in an electric field.
- Electrophoresis is one of the most commonly used analytical scale **separation** techniques
  - The electric field pulls proteins according to their charge.
- Gel electrophoresis adds a solid support in which the separation occurs.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

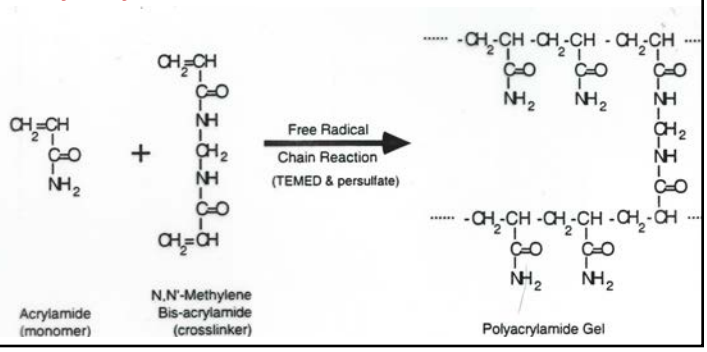


Gels

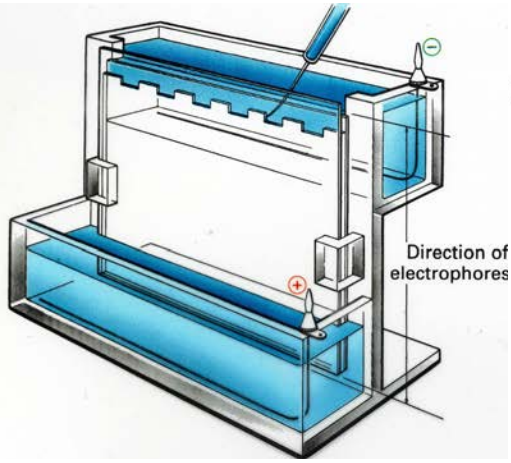
Agarose



Polyacrylamide

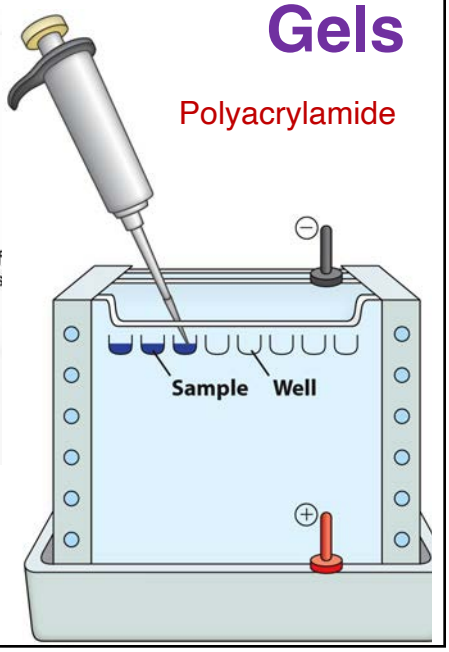


Electrophoresis for Protein Analysis



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

$f = 6\Pi \bullet \eta \bullet r$

where:

η = coefficient of viscosity

r = Stokes radius (mass and shape)

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

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

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