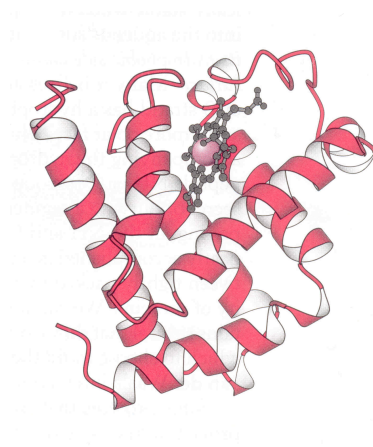
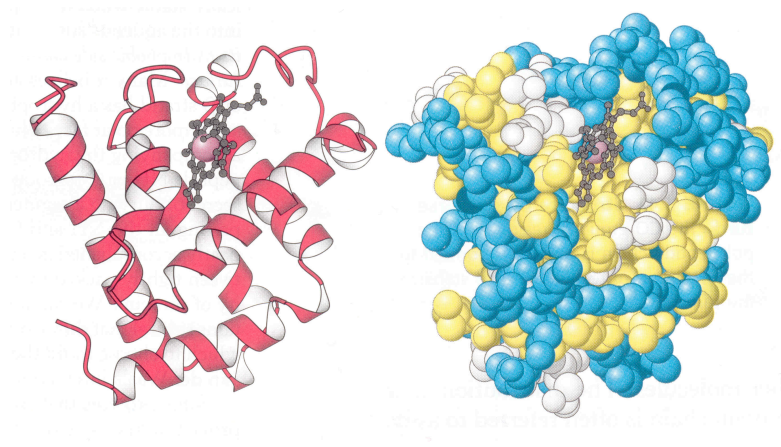


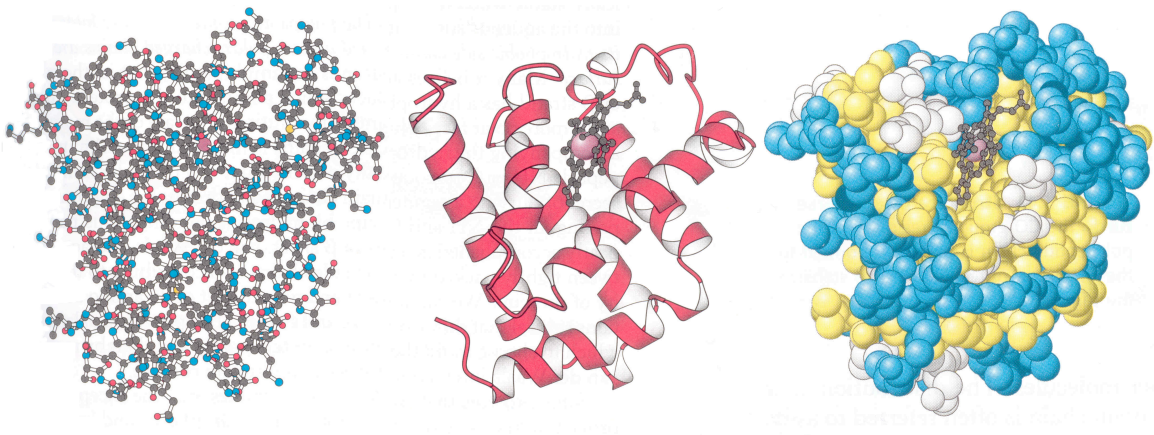
# Proteins



# Proteins



# Proteins (Myoglobin)



154 amino acids  
 17,266 Daltons  
 ~1,900 atoms (1,264 atoms shown)

## Structure/Sequence Relationship

Equus_caballus	-GLSDGEWQQVLNVWGKVEADIAGHGQEVLRIRLFTGHPETLEKFDKFKHLKTEAEMKASE	59
Elephas_maximus	-GLSDGEWELVLTWGWKVEADIPGHGEFVLRVLRFTGHPETLEKFDKFKHLKTEGEMKASE	59
Homo_sapiens	MGLSDGEWQLVNLVWGKVEADIPGHGQEVLRIRLFKGPETLEKFDKFKHLKSEDEMKASE	60
Physeter_catodon	MVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEMKASE	60
Physeter_catodon_HbA	-VLSPADKTNVKAAWAKVGNHAADFGEALERMFMSFPSTKTYFSHF-DLG-----HNST	53
Physeter_catodon_HbB	VHLTGEESGLTALWAKVN--VEEIGGEALGRLLVVYPWTQRFEFHFGDLSTADAVMKNP	58
	*: : : *.* * **:: .* * *:* *	.
Equus_caballus	DLKKHGTVVLTALGGILKKKGHHEAELKPLAQDHATKHKIPIKYLEFISDAIIHVLHSKH	119
Elephas_maximus	DLKKQGVTVLTALGGILKKKGHHEAEIQPLAQSHATKHKIPIKYLEFISDAIIHVLQSKH	119
Homo_sapiens	DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH	120
Physeter_catodon	DLKKHGTVVLTALGAILKKKGHHEAELKPWAQSHATKHKIPIKYLEFISEAIIHVLHSRH	120
Physeter_catodon_HbA	QVKGHGKKVADALTKAVGHLDLTPDALSDLSDLHAKLRVDPVNFKLLSCHLLVTLAAHL	113
Physeter_catodon_HbB	KVKKHGQKVLASFGEGLKHLNLDLTKGTATLSELHCDKLVDPENFRLLGNVLLVVLARHF	118
	..* :* * :: : . . : :: * . * :: : ..: . : . * :	
Equus_caballus	PGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	153
Elephas_maximus	PAEFGADAQGAMKALELFRNDIAAKYKELGFQG	153
Homo_sapiens	PGDFGADAQGAMNKALELFRKDMSNYKELGFQG	154
Physeter_catodon	PGNFGADAQGAMNKALELFRKDIAAKYKELGYQG	154
Physeter_catodon_HbA	PGDFTPSVHASLDKFLASVSTVLTSKYR-----	141
Physeter_catodon_HbB	GKEFTPELQTAYQKVVAGVANALAHKYH-----	146
	:* . . : : * : . . : : *:	

# Structure/Sequence Relationship



# Structure/Sequence Relationship

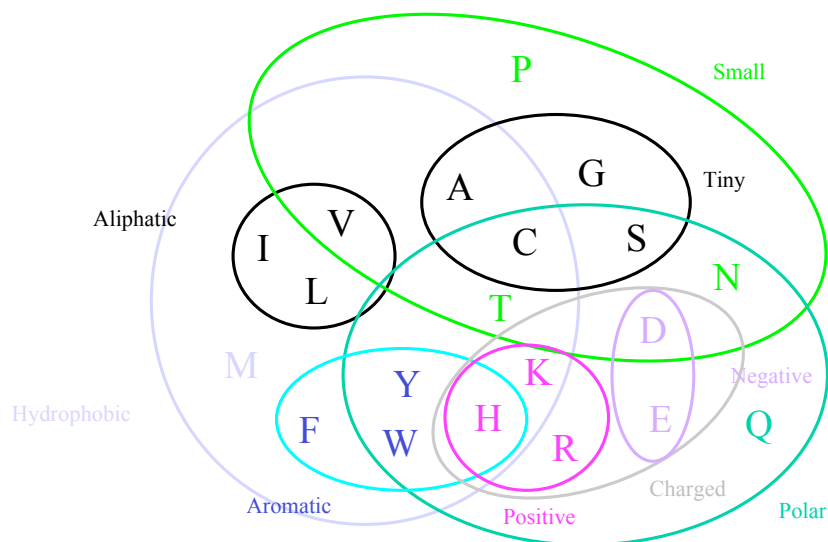


# Structure/Sequence Relationship

To understand and manipulate protein sequence and structure we have to know:

- \_ The 20 'naturally occurring' amino acids
- \_ Their physical properties
- \_ The nature of the peptide bond connecting them
- \_ Possible modifications that can occur after synthesis

## Amino Acid Properties

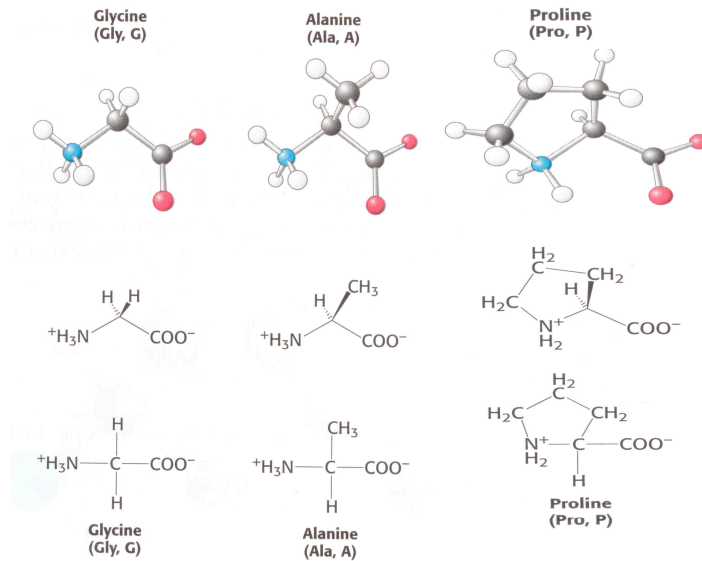


# The Amino Acids

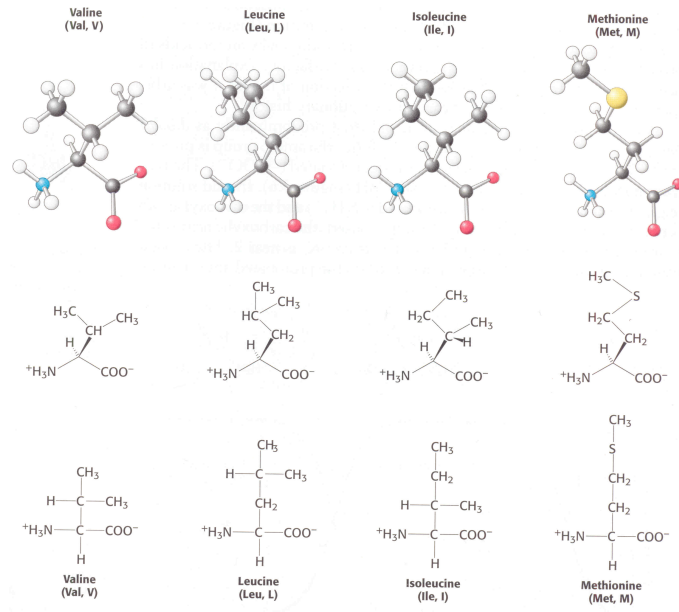
Amino acid	Three-letter abbreviation	One-letter abbreviation	Amino acid	Three-letter abbreviation	One-letter abbreviation
Alanine	Ala	A	Methionine	Met	M
Arginine	Arg	R	Phenylalanine	Phe	F
Asparagine	Asn	N	Proline	Pro	P
Aspartic Acid	Asp	D	Serine	Ser	S
Cysteine	Cys	C	Threonine	Thr	T
Glutamine	Gln	Q	Tryptophan	Trp	W
Glutamic Acid	Glu	E	Tyrosine	Tyr	Y
Glycine	Gly	G	Valine	Val	V
Histidine	His	H	Asparagine or aspartic acid	Asx	B
Isoleucine	Ile	I	Glutamine or glutamic acid	Glx	Z
Leucine	Leu	L			
Lysine	Lys	K			

\_ Know the code – Learn **name**, **abbreviation** and **structure**

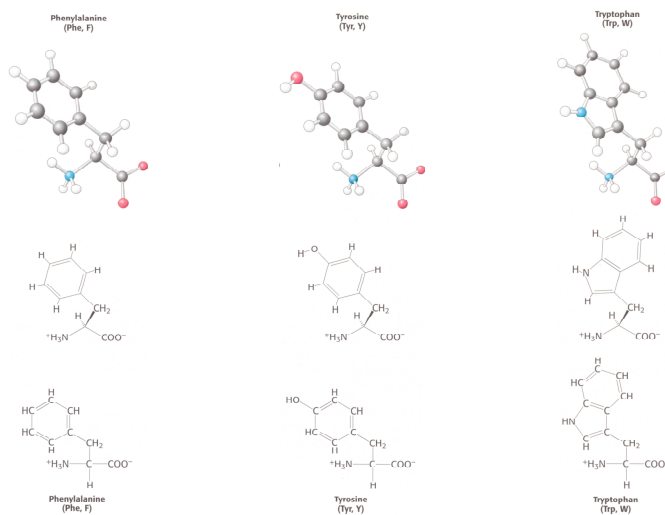
## Aliphatic Amino Acids 1



# Aliphatic Amino Acids 2

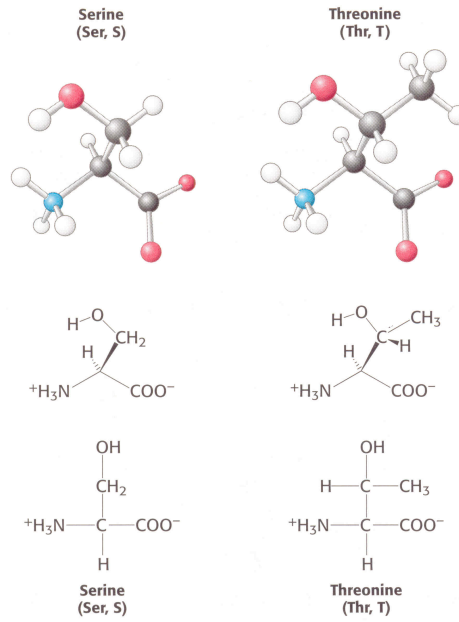


# Aromatic Amino Acids



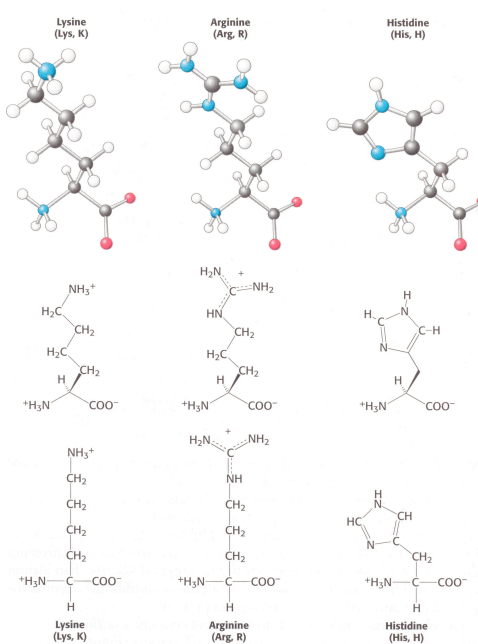
Histidine?

# Hydroxyl Containing Amino Acids

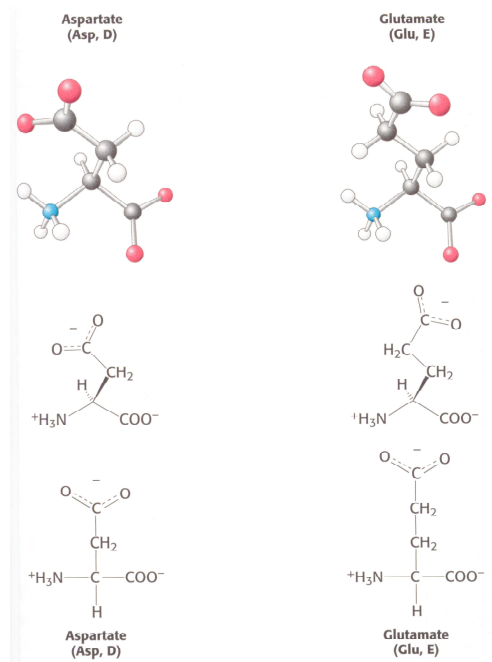


Tyrosine?

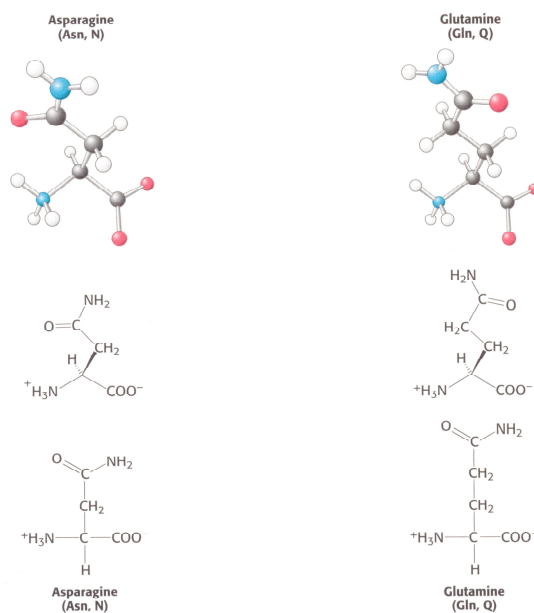
# Basic Amino Acids



# Acidic Amino Acids

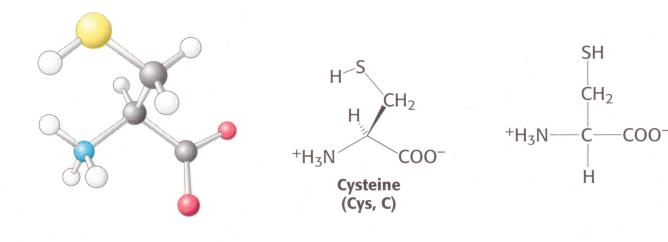


# Amide Containing Amino Acids





# Cysteine



## Properties of AAs

Main chain chirality

Side chain chirality

Main chain ionization

Side chain ionization

Mass

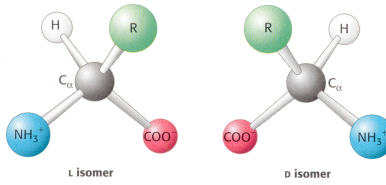
Absorbance

Hydrogen bond interactions

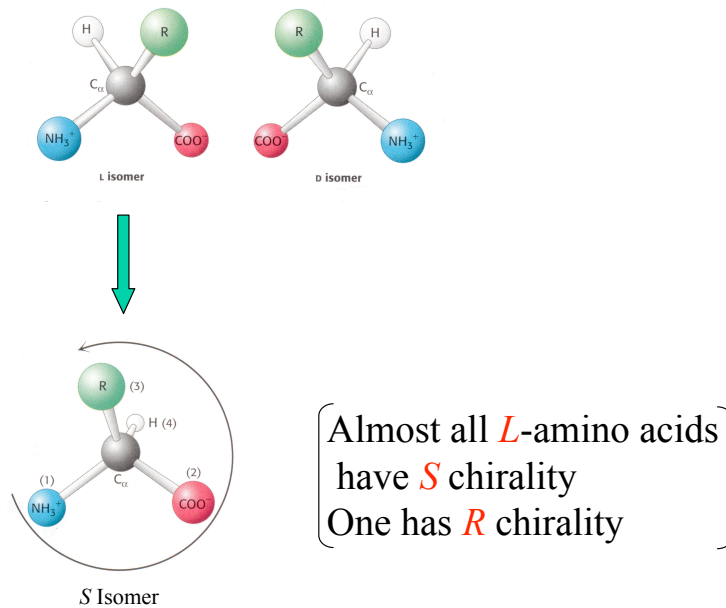
Salt bridge interactions

Structural implications - Thursday

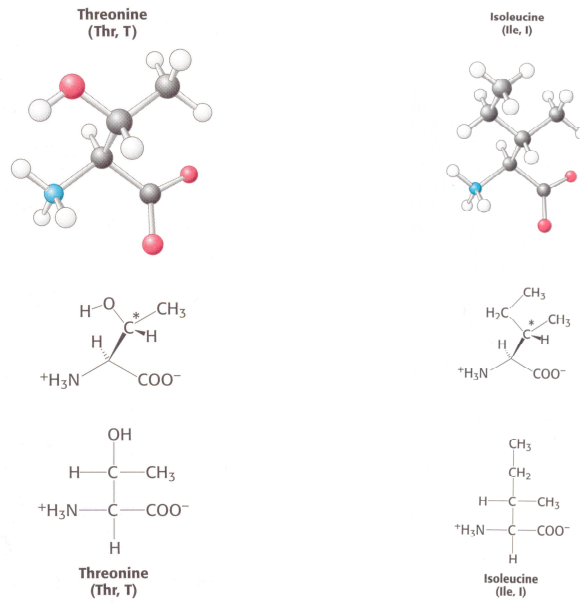
# Main Chain Chirality



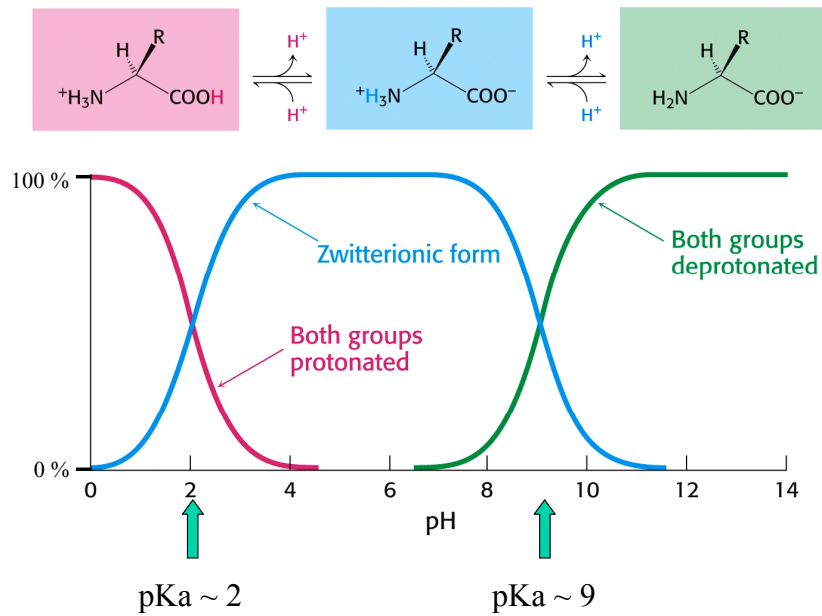
# Main Chain Chirality



# Side Chain Chirality



# Main Chain Ionization



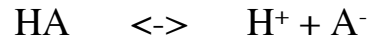
## Evaluating ionization state with pH & pK<sub>a</sub>

### What is pH?

$$\text{pH} = -\log_{10}[\text{H}^+] = \log_{10}(1/[\text{H}^+])$$

### What is pK?

Consider a weak acid,



The equilibrium constant, K<sub>a</sub>, for this rxn is:

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

$$\text{pK}_a = -\log K_a = \log (1/K_a) = \log ([\text{HA}]/[\text{H}^+][\text{A}^-])$$

## Evaluating ionization state with pH and pK

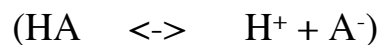
$$\text{pH} = \log (1/[\text{H}^+])$$

$$\text{pK}_a = \log ([\text{HA}]/[\text{H}^+][\text{A}^-])$$

$$= \log (1/[\text{H}^+]) + \log ([\text{HA}]/[\text{A}^-])$$

$$\text{pK}_a = \text{pH} + \log ([\text{HA}]/[\text{A}^-])$$

when  $[\text{HA}] = [\text{A}^-]$  i.e. the acid is 50% protonated



$$\text{pK}_a = \text{pH}$$

in other words...

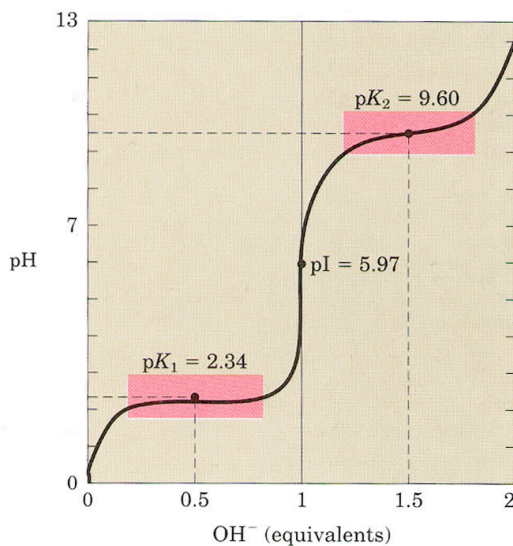
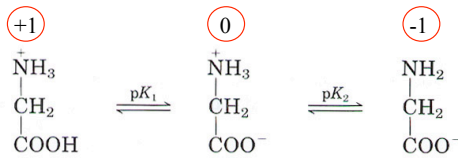
**pK<sub>a</sub> is the pH at which the acid is 50% ionized**

**TABLE 3.4**  $pK_a$  values of some amino acids

Amino acid	$pK_a$ values (25°C)		
	$\alpha$ -COOH group	$\alpha$ -NH <sub>3</sub> <sup>+</sup> group	Side chain
Alanine	2.3	9.9	
Glycine	2.4	9.8	
Phenylalanine	1.8	9.1	
Serine	2.1	9.2	
Valine	2.3	9.6	
Aspartic acid	2.0	10.0	3.9
Glutamic acid	2.2	9.7	4.3
Histidine	1.8	9.2	6.0
Cysteine	1.8	10.8	8.3
Tyrosine	2.2	9.1	10.9
Lysine	2.2	9.2	10.8
Arginine	1.8	9.0	12.5

After J. T. Edsall and J. Wyman, *Biophysical Chemistry* (Academic Press, 1958), Chapter 8.

### The titration curve of a dibasic amino acid



The titration curve of 0.1 M glycine at 25 °C.

Isoelectric Point (pI):

pH at which an amino acid has no NET charge

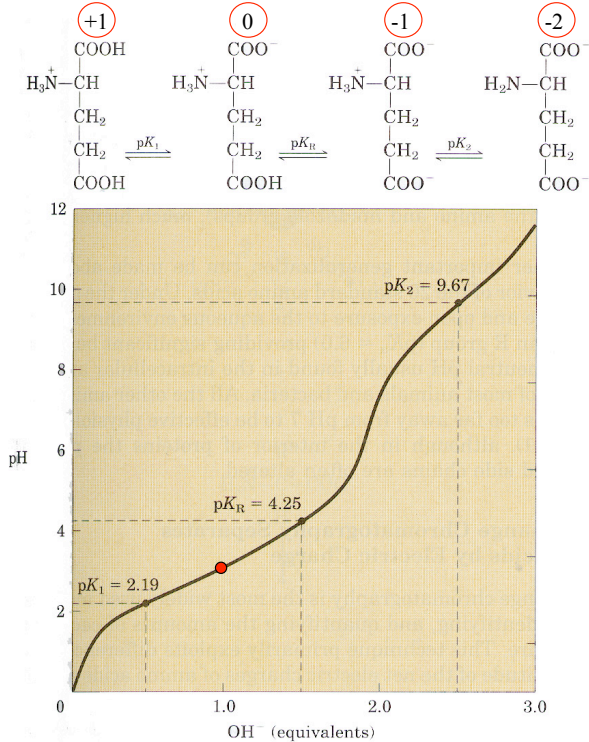
For an amino acid with an uncharged R group:

$$pI = (pK_1 + pK_2) / 2$$

Thus, for Gly:

$$\begin{aligned}
 pI &= (2.34 + 9.60) / 2 \\
 &= 5.97
 \end{aligned}$$

## The titration curve of a tribasic amino acid



The titration curves of glutamate

Isoelectric Point (pI):

pH at which an amino acid has no NET charge

For an amino acid with a charged R group:

pI = the average of the pKs "flanking" the neutral ionization state

Thus, for Glu:

$$\begin{aligned} \text{pI} &= (2.19 + 4.25) / 2 \\ &= 3.22 \end{aligned}$$

Why do we have to learn all this pI stuff?

- Remember: a protein's function depends on its structure
- A protein's structure, as you will see in the next class, depends on its charge (pI)
- A protein's pI depends on the  $pK_a$  of all its amino acids

# Side Chain Ionization

Group	Acid	⇌	Base	Typical pK <sub>a</sub> *
Terminal α-carboxyl group		⇌		3.1
Aspartic acid Glutamic acid		⇌		4.1
Histidine		⇌		6.0
Terminal α-amino group		⇌		8.0
Cysteine		⇌		8.3
Tyrosine		⇌		10.9
Lysine		⇌		10.8
Arginine		⇌		12.5

## Amino Acid pKa Values

pKa values depend on the environment:

A nearby **positive** charge will shift pKa values **down**  
(resulting in less positive charge)

A nearby **negative** charge will shift pKa values **up**  
(resulting in less negative charge)

A low dielectric environment (hydrophobic protein core) will:

Shift the pKa value of a **basic** group **down** (less charge)

Shift the pKa value of an **acidic** group **up** (less charge)

Calculating pKa shifts in a protein is difficult

$$\text{Coulomb's Law } (F=q_1q_2/4\pi\epsilon R^2)$$

$$\epsilon (\text{protein core}) \approx 4, \quad \epsilon (\text{water}) \approx 80$$

# Protein Isoelectric Point

The average of all of the pKa values in a protein is the **Isoelectric point (pI)**

At this pH value the net charge of the protein will be **zero**

The pI can be **estimated** by using the standard pKa values for all of the ionizable groups in a protein

See [http://ca.expasy.org/tools/pi\\_tool.html](http://ca.expasy.org/tools/pi_tool.html)

## Amino Acid Masses

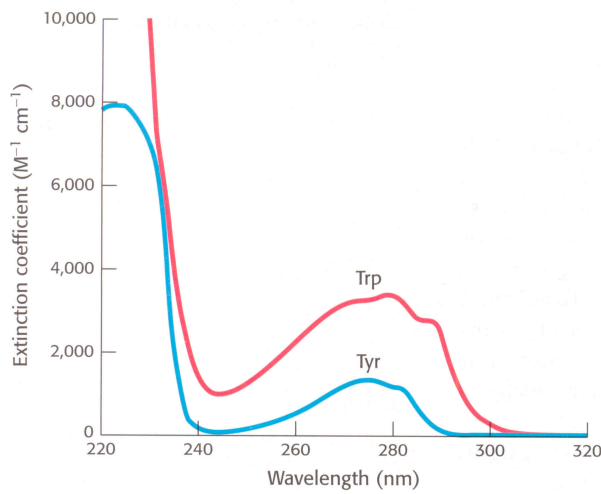
			Monoisotopic	Average Mass
A	Ala	C <sub>3</sub> H <sub>5</sub> ON	71.03711	71.0788
R	Arg	C <sub>6</sub> H <sub>12</sub> ON <sub>4</sub>	156.10111	156.1875
N	Asn	C <sub>4</sub> H <sub>6</sub> O <sub>2</sub> N <sub>2</sub>	114.04293	114.1038
D	Asp	C <sub>4</sub> H <sub>5</sub> O <sub>3</sub> N	115.02694	115.0886
C	Cys	C <sub>3</sub> H <sub>5</sub> ONS	103.00919	103.1388
E	Glu	C <sub>5</sub> H <sub>7</sub> O <sub>3</sub> N	129.04259	129.1155
Q	Gln	C <sub>5</sub> H <sub>8</sub> O <sub>2</sub> N <sub>2</sub>	128.05858	128.1307
G	Gly	C <sub>2</sub> H <sub>3</sub> ON	57.02146	57.0519
H	His	C <sub>6</sub> H <sub>7</sub> ON <sub>3</sub>	137.05891	137.1411
I	Ile	C <sub>6</sub> H <sub>11</sub> ON	113.08406	113.1594
L	Leu	C <sub>6</sub> H <sub>11</sub> ON	113.08406	113.1594
K	Lys	C <sub>6</sub> H <sub>12</sub> ON <sub>2</sub>	128.09496	128.1741
M	Met	C <sub>5</sub> H <sub>9</sub> ONS	131.04049	131.1926
F	Phe	C <sub>9</sub> H <sub>9</sub> ON	147.06841	147.1766
P	Pro	C <sub>5</sub> H <sub>7</sub> ON	97.05276	97.1167
S	Ser	C <sub>3</sub> H <sub>5</sub> O <sub>2</sub> N	87.03203	87.0782
T	Thr	C <sub>4</sub> H <sub>7</sub> O <sub>2</sub> N	101.04768	101.1051
W	Trp	C <sub>11</sub> H <sub>10</sub> ON <sub>2</sub>	186.07931	186.2132
Y	Tyr	C <sub>9</sub> H <sub>8</sub> O <sub>2</sub> N	163.06333	163.1760
V	Val	C <sub>3</sub> H <sub>5</sub> ON	99.06841	99.1326

Assumes amino acids are linked by peptide bonds (H<sub>2</sub>O removed)

See [http://ca.expasy.org/tools/pi\\_tool.html](http://ca.expasy.org/tools/pi_tool.html)



# Absorbance



Beer's Law

$$A = \epsilon c l$$

A = absorbance

$\epsilon$  = extinction coefficient ( $M^{-1} \text{ cm}^{-1}$ )

c = concentration (M)

l = pathlength (cm)

Absorbance at 280 nm usually used for protein quantification

Note: Cysteine absorbs at 280 nm if disulfide bonded ( $\epsilon = 60 M^{-1} \text{ cm}^{-1}$ )

$$\epsilon (\text{Trp}) \approx 5690 M^{-1} \text{ cm}^{-1}$$

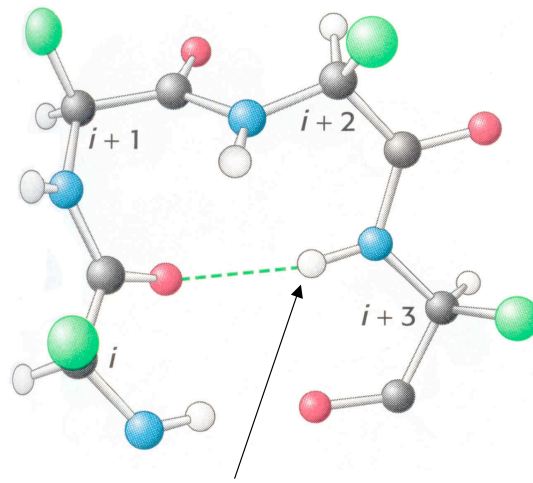
$$\epsilon (\text{Tyr}) \approx 1280 M^{-1} \text{ cm}^{-1}$$

See <http://ca.expasy.org/cgi-bin/protparam.html>

Parameters for the Unmodified Physiological L- $\alpha$ -Amino Acids			
Amino Acid	3-Letter Code	1-Letter Code	Hydrophobicity
Alanine	Ala	A	0.616
Cysteine	Cys	C	0.680
Aspartate	Asp	D	0.028
Glutamate	Glu	E	0.043
Phenylalanine	Phe	F	1.00
Glycine	Gly	G	0.501
Histidine	His	H	0.165
Isoleucine	Ile	I	0.943
Lysine	Lys	K	0.283
Leucine	Leu	L	0.943
Methionine	Met	M	0.738
Asparagine	Asn	N	0.236
Proline	Pro	P	0.711
Glutamine	Gln	Q	0.251
Arginine	Arg	R	0.000
Serine	Ser	S	0.359
Threonine	The	T	0.450
Valine	Val	V	0.825
Tryptophan	Trp	W	0.878
Tyrosine	Tyr	Y	0.880

<http://psyche.uthct.edu/shaun/SBlack/aagrease.html>

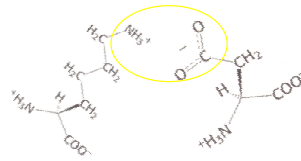
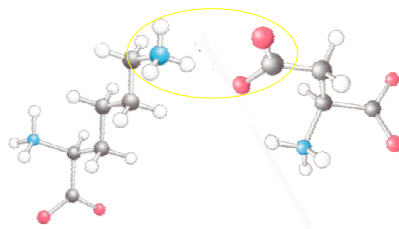
# Hydrogen Bond Interactions



Sharing of hydrogen

Side chains of W, Y, S, T, N, Q, D, E, K, R

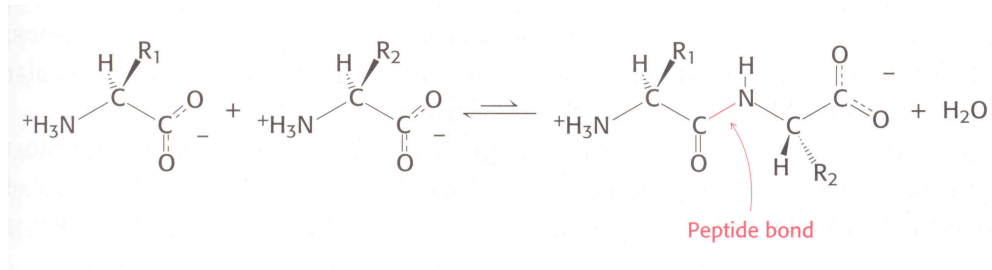
# Salt Bridge Interactions



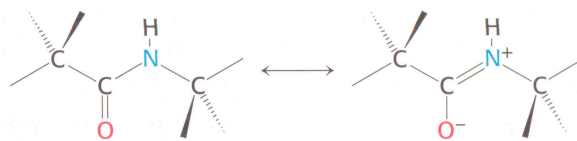
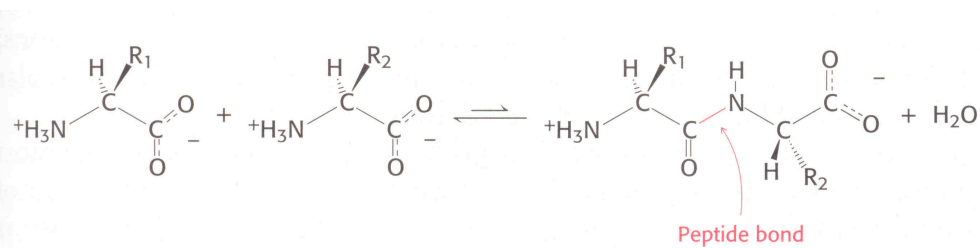
Charge – charge interactions

Typically between (D or E) and (K, R or H)

# Peptide Bond



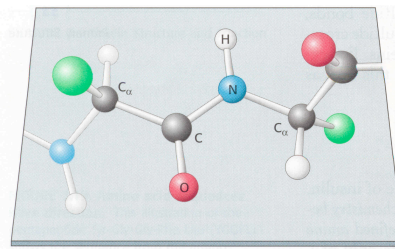
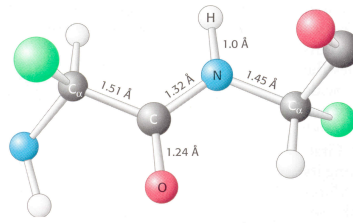
# Peptide Bond



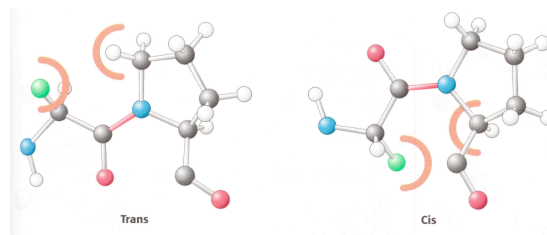
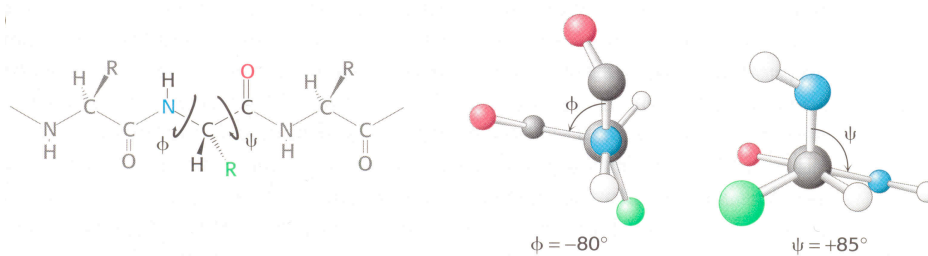
Peptide bond resonance structures

Additional stability due to resonance

# Peptide Bond Planarity

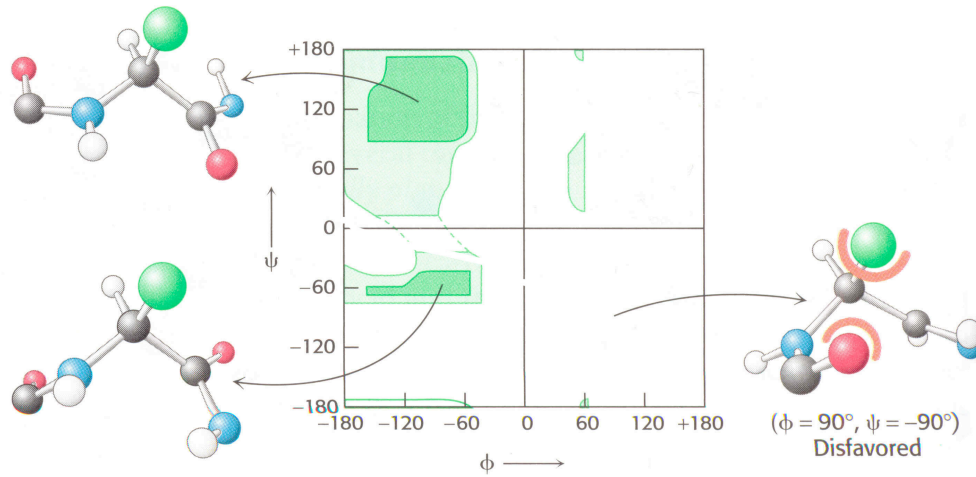


# Main Chain Geometry (phi and psi)

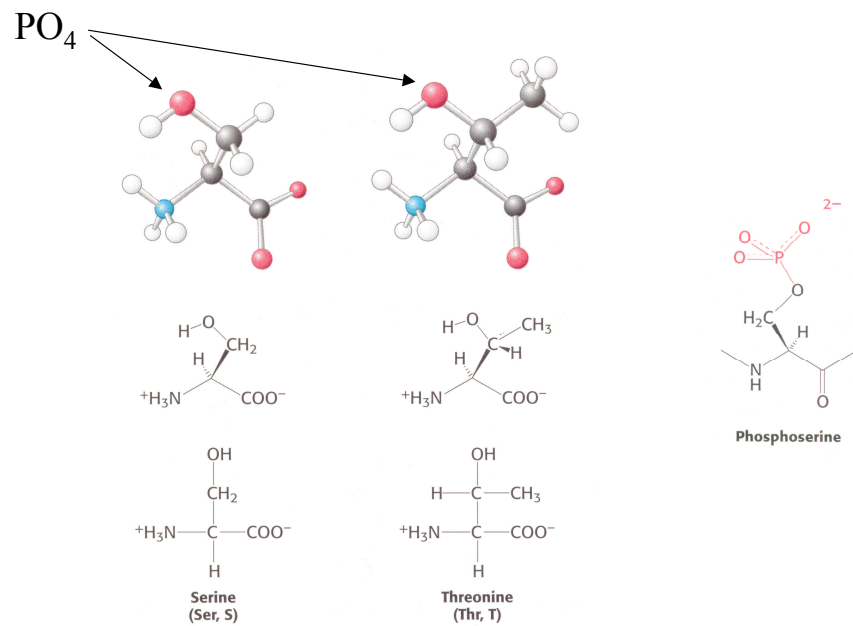


Omega can be  $180^\circ$  (trans) or  $0^\circ$  (cis) in proline

# Ramachandran Plot

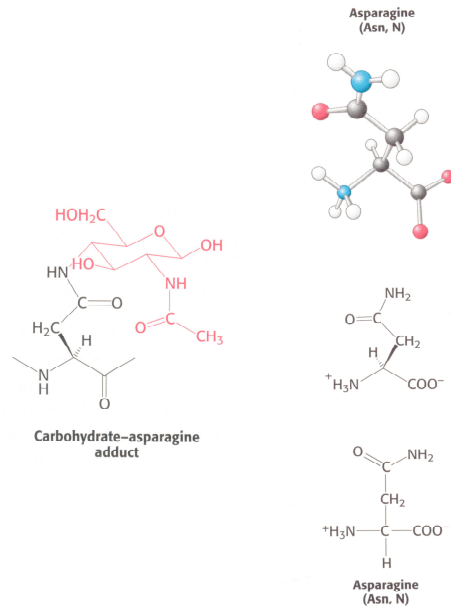


# Posttranslational Modifications



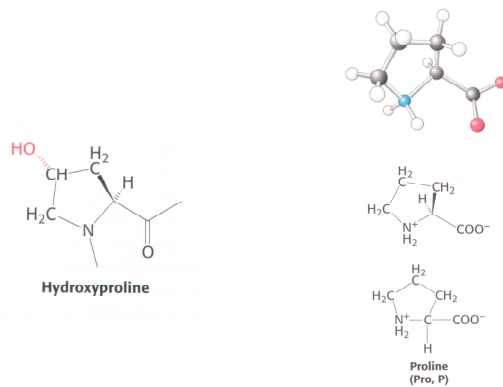
## Phosphorylation

# Posttranslational Modifications



## Glycosylation

# Posttranslational Modifications



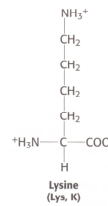
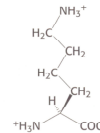
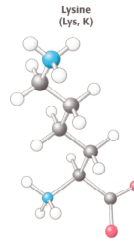
## Hydroxylation

# Posttranslational Modifications

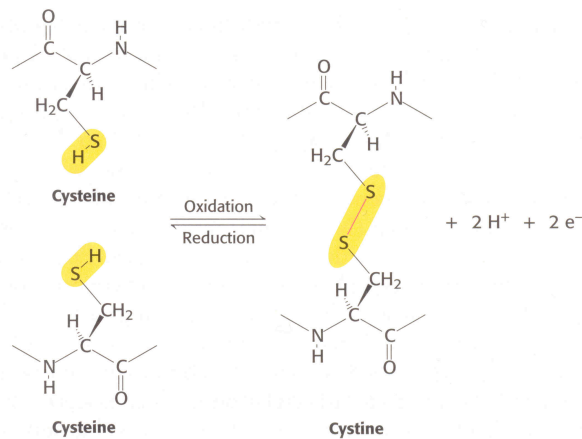
Methylation

Hydroxylation

Acetylation (“histone code”)

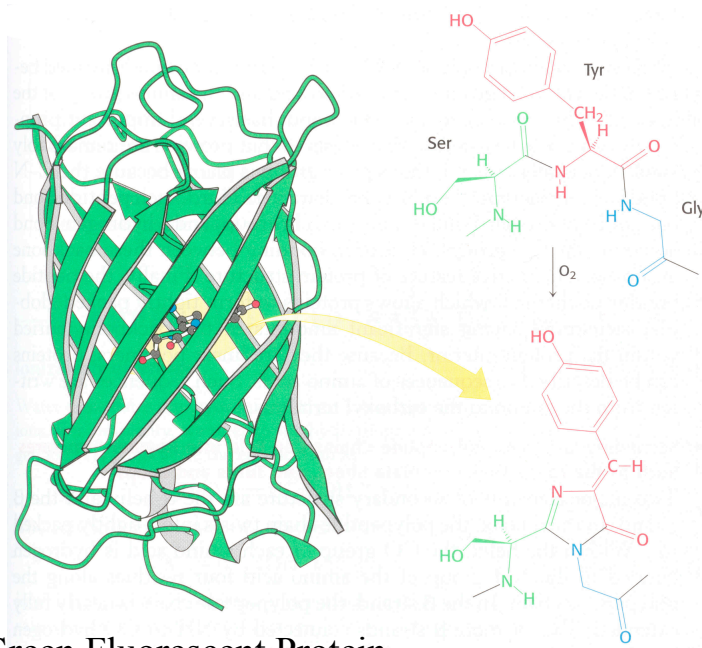


# Posttranslational Modifications



Disulfide formation

# Posttranslational Modifications



Green Fluorescent Protein

Next Class  
Protein Structure