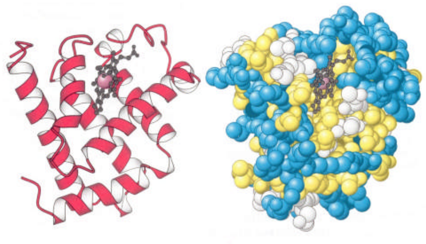
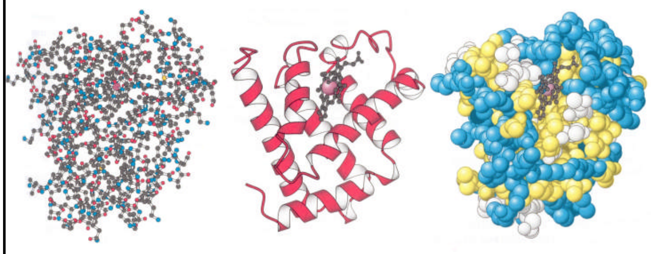
Proteins

Proteins



Proteins (Myoglobin)



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

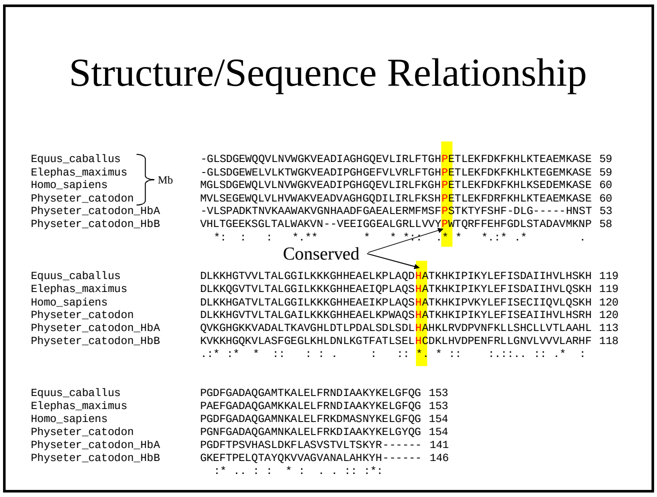
Structure/Sequence Relationship

```
Equus_caballus            -GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASE 59
Elephas_maximus           -GLSDGEWELVLKTWGKVEADIPGHGEFVLVRLFTGHPETLEKFDKFKHLKTEGEMKASE 59
Homo_sapiens              MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASE 60
Physeter_catodon          MVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASE 60
Physeter_catodon_HbA      -VLSPADKTNVKAAWAKVGNHAADFGAEALERMFMSFPSTKTYFSHF-DLG-----HNST 53
Physeter_catodon_HbB      VHLTGEEKSGLTALWAKVN--VEEIGGEALGRLLVVYPWTQRFFEHFGDLSTADAVMKNP 58
                            *:  :   :   *.**       *   * *::  .* *   *.:* .*        .

Equus_caballus            DLKKHGTVVLTALGGILKKKGHHEAELKPLAQDHATKHKIPIKYLEFISDAIIHVLHSKH 119
Elephas_maximus           DLKKQGVTVLTALGGILKKKGHHEAEIQPLAQSHATKHKIPIKYLEFISDAIIHVLQSKH 119
Homo_sapiens              DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH 120
Physeter_catodon          DLKKHGVTVLTALGAILKKKGHHEAELKPWAQSHATKHKIPIKYLEFISEAIIHVLHSRH 120
Physeter_catodon_HbA      QVKGHGKKVADALTKAVGHLDTLPDALSDLSDLHAHKLRVDPVNFKLLSHCLLVTLAAHL 113
Physeter_catodon_HbB      KVKKHGQKVLASFGEGLKHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVVVLARHF 118
                          .:* :*  *  ::   : : .     :   :: *. * ::    :.::.. :: .*  :

Equus_caballus            PGDFGADAQGAMTKALELFRNDIAAKYKELGFQG 153
Elephas_maximus           PAEFGADAQGAMKKALELFRNDIAAKYKELGFQG 153
Homo_sapiens              PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154
Physeter_catodon          PGNFGADAQGAMNKALELFRKDIAAKYKELGYQG 154
Physeter_catodon_HbA      PGDFTPSVHASLDKFLASVSTVLTSKYR------ 141
Physeter_catodon_HbB      GKEFTPELQTAYQKVVAGVANALAHKYH------ 146
                            :* .. : :  * :  . . :: :*:
```

Structure/Sequence Relationship



Mb

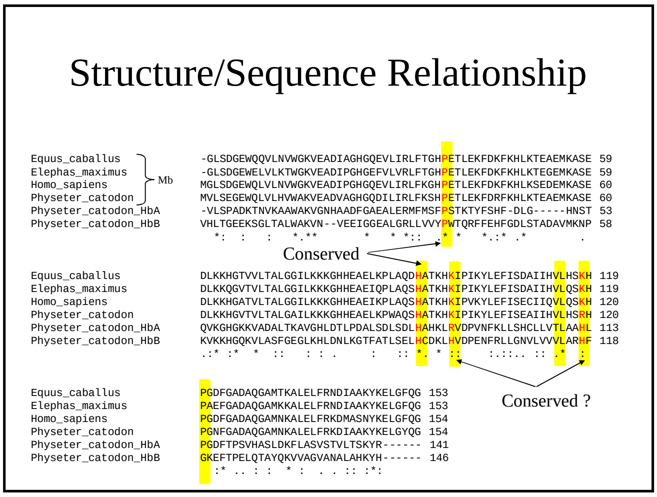
```
Equus_caballus            -GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASE 59
Elephas_maximus           -GLSDGEWELVLKTWGKVEADIPGHGEFVLVRLFTGHPETLEKFDKFKHLKTEGEMKASE 59
Homo_sapiens              MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASE 60
Physeter_catodon          MVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASE 60
Physeter_catodon_HbA      -VLSPADKTNVKAAWAKVGNHAADFGAEALERMFMSFPSTKTYFSHF-DLG-----HNST 53
Physeter_catodon_HbB      VHLTGEEKSGLTALWAKVN--VEEIGGEALGRLLVVYPWTQRFFEHFGDLSTADAVMKNP 58
                            *:  :   :   *.**       *   * *::  .* *   *.:* .*        .
```

Conserved

```
Equus_caballus            DLKKHGTVVLTALGGILKKKGHHEAELKPLAQDHATKHKIPIKYLEFISDAIIHVLHSKH 119
Elephas_maximus           DLKKQGVTVLTALGGILKKKGHHEAEIQPLAQSHATKHKIPIKYLEFISDAIIHVLQSKH 119
Homo_sapiens              DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH 120
Physeter_catodon          DLKKHGVTVLTALGAILKKKGHHEAELKPWAQSHATKHKIPIKYLEFISEAIIHVLHSRH 120
Physeter_catodon_HbA      QVKGHGKKVADALTKAVGHLDTLPDALSDLSDLHAHKLRVDPVNFKLLSHCLLVTLAAHL 113
Physeter_catodon_HbB      KVKKHGQKVLASFGEGLKHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVVVLARHF 118
                          .:* :*  *  ::   : : .     :   :: *. * ::    :.::.. :: .*  :

Equus_caballus            PGDFGADAQGAMTKALELFRNDIAAKYKELGFQG 153
Elephas_maximus           PAEFGADAQGAMKKALELFRNDIAAKYKELGFQG 153
Homo_sapiens              PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154
Physeter_catodon          PGNFGADAQGAMNKALELFRKDIAAKYKELGYQG 154
Physeter_catodon_HbA      PGDFTPSVHASLDKFLASVSTVLTSKYR------ 141
Physeter_catodon_HbB      GKEFTPELQTAYQKVVAGVANALAHKYH------ 146
                            :* .. : :  * :  . . :: :*:
```

Structure/Sequence Relationship



Mb

```
Equus_caballus            -GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASE 59
Elephas_maximus           -GLSDGEWELVLKTWGKVEADIPGHGEFVLVRLFTGHPETLEKFDKFKHLKTEGEMKASE 59
Homo_sapiens              MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASE 60
Physeter_catodon          MVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASE 60
Physeter_catodon_HbA      -VLSPADKTNVKAAWAKVGNHAADFGAEALERMFMSFPSTKTYFSHF-DLG-----HNST 53
Physeter_catodon_HbB      VHLTGEEKSGLTALWAKVN--VEEIGGEALGRLLVVYPWTQRFFEHFGDLSTADAVMKNP 58
                            *:  :   :   *.**       *   * *::  .* *   *.:* .*        .
```

Conserved

```
Equus_caballus            DLKKHGTVVLTALGGILKKKGHHEAELKPLAQDHATKHKIPIKYLEFISDAIIHVLHSKH 119
Elephas_maximus           DLKKQGVTVLTALGGILKKKGHHEAEIQPLAQSHATKHKIPIKYLEFISDAIIHVLQSKH 119
Homo_sapiens              DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH 120
Physeter_catodon          DLKKHGVTVLTALGAILKKKGHHEAELKPWAQSHATKHKIPIKYLEFISEAIIHVLHSRH 120
Physeter_catodon_HbA      QVKGHGKKVADALTKAVGHLDTLPDALSDLSDLHAHKLRVDPVNFKLLSHCLLVTLAAHL 113
Physeter_catodon_HbB      KVKKHGQKVLASFGEGLKHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVVVLARHF 118
                          .:* :*  *  ::   : : .     :   :: *. * ::    :.::.. :: .*  :
```

Conserved ?

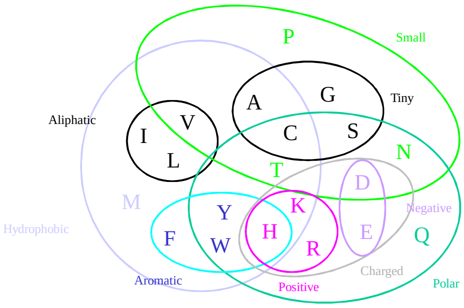
```
Equus_caballus            PGDFGADAQGAMTKALELFRNDIAAKYKELGFQG 153
Elephas_maximus           PAEFGADAQGAMKKALELFRNDIAAKYKELGFQG 153
Homo_sapiens              PGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154
Physeter_catodon          PGNFGADAQGAMNKALELFRKDIAAKYKELGYQG 154
Physeter_catodon_HbA      PGDFTPSVHASLDKFLASVSTVLTSKYR------ 141
Physeter_catodon_HbB      GKEFTPELQTAYQKVVAGVANALAHKYH------ 146
                            :* .. : :  * :  . . :: :*:
```

Structure/Sequence Relationship

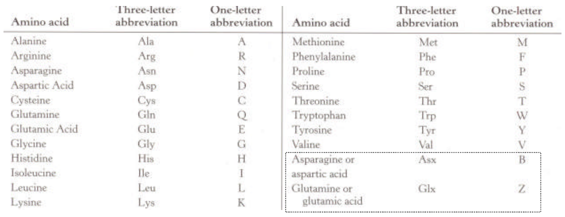
To understand and manipulate protein sequence and structure we have to know:

- 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



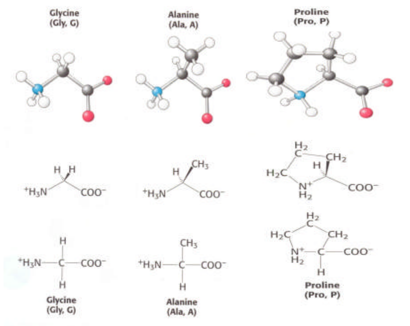
The Amino Acids



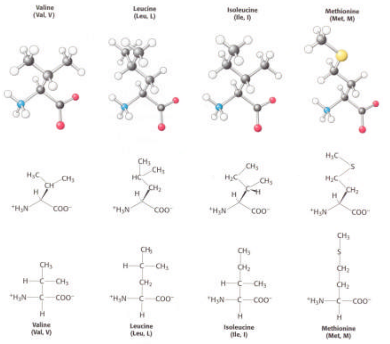
Amino acid	Three-letter abbreviation	One-letter abbreviation	Amino acid	Three-letter abbreviation	One-letter abbreviation
Alanine	Ala	A	Methionine	Met	M
Arginine	Arg	R	Phenylalanine	Phe	F
Asparagine	Asn	N	Proline	Pro	P
Aspartic Acid	Asp	D	Serine	Ser	S
Cysteine	Cys	C	Threonine	Thr	T
Glutamine	Gln	Q	Tryptophan	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			
Leucine	Leu	L	Glutamine or glutamic acid	Glx	Z
Lysine	Lys	K			

☼ Know the code – Learn name, abbreviation and structure

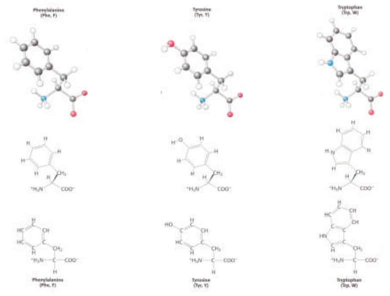
Aliphatic Amino Acids 1



Aliphatic Amino Acids 2

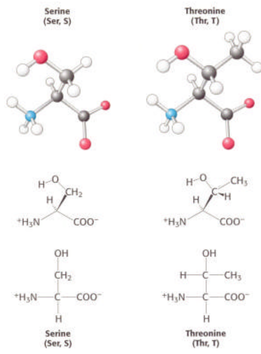


Aromatic Amino Acids



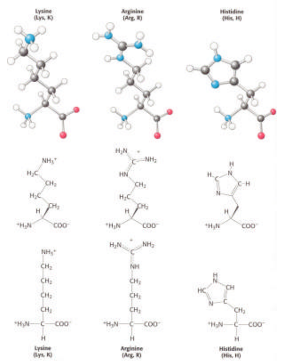
Histidine?

Hydroxyl Containing Amino Acids

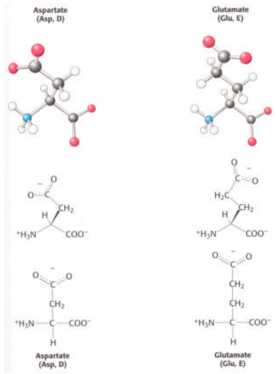


Tyrosine?

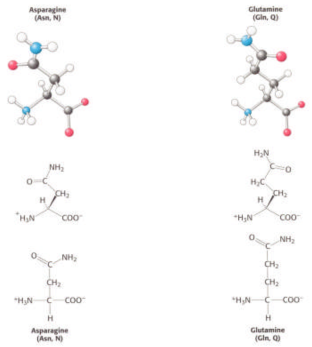
Basic Amino Acids



Acidic Amino Acids



Amide Containing Amino Acids



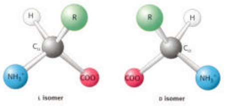
Cysteine

Properties of AAs

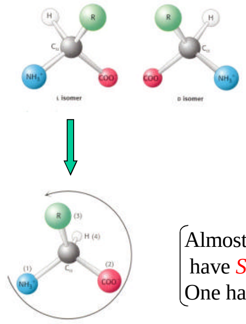
Main chain chirality
Side chain chirality
Main chain ionization
Side chain ionization
Mass
Absorbance
Hydrogen bond interactions
Salt bridge interactions

Structural implications - Thursday

Main Chain Chirality



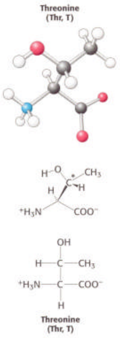
Main Chain Chirality



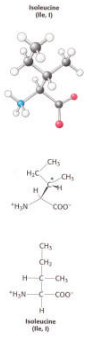
S Isomer

Almost all *L*-amino acids
have *S* chirality
One has *R* chirality

Side Chain Chirality

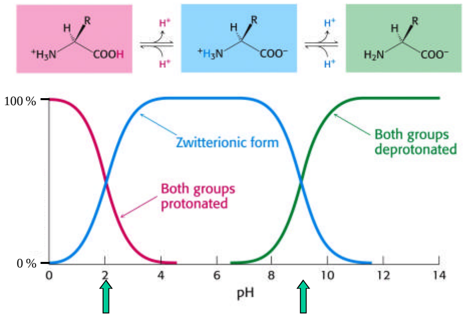


Threonine
(Thr, T)



Isoleucine
(Ile, I)

Main Chain Ionization



pKa ~ 2

pKa ~ 9

Evaluating ionization state with pH & pK_a

What is pH?

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

What is pK?

Consider a weak acid,

$HA \ \text{<->} \ H^+ + A^-$

The equilibrium constant, K_a, for this rxn is:

$K_a = [H^+][A^-]/[HA]$

$pK_a = -\log K_a = \log (1/K_a) = \log ([HA]/[H^+][A^-])$

Evaluating ionization state with pH and pK

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

$pK_a = \log ([HA]/[H^+][A^-])$

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

$pK_a = pH + \log ([HA]/[A^-])$

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

$(HA \ \text{<->} \ H^+ + A^-)$

$pK_a = pH$

in other words…

pK_a is the pH at which the acid is 50% ionized

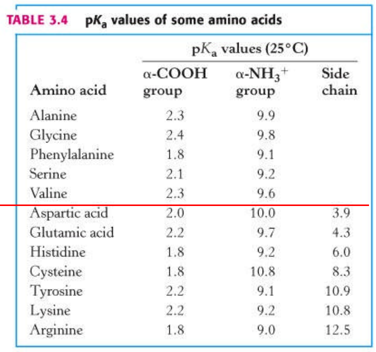
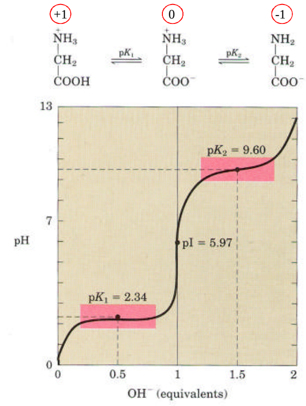


TABLE 3.4 pK_a values of some amino acids

Amino acid	α-COOH group	α-NH_3^+ group	Side chain
	pK_a values (25°C)		
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



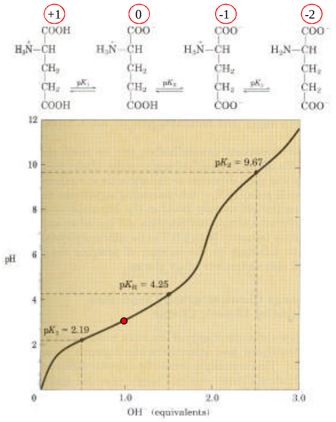
The titration curve of 0.1 M glycine at 25 °C.

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:
pI = (2.34 + 9.60) / 2
= 5.97

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

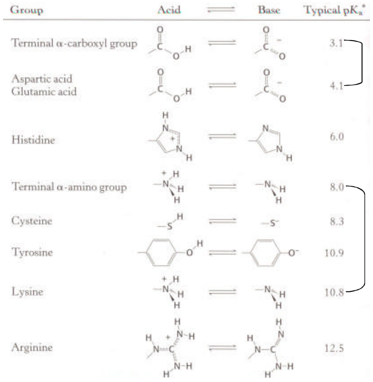
For an amino acid with a charged R group:
pI = the average of the pKs "flanking" the neutral ionization state

Thus, for Glu:
pI = (2.19 + 4.25) / 2
= 3.22

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_a*
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

Coulomb's Law ($F=q_1q_2/4\pi\varepsilon R^2$) ε (protein core) ≈ 4, ε (water) ≈ 80

Protein Isoelectric Point

The Isoelectric point (pI) is the pH value at which the net charge of the protein will be zero

The pI can be measured using isoelectric focusing gels

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

See http://ca.expasy.org/tools/pi_tool.html

*Remember: only one amino- and one carboxy- terminus

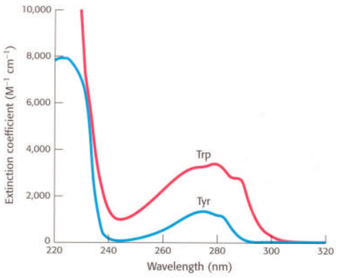
Amino Acid Masses

			Monoisotopic	Average Mass
A	Ala	C_3H_5ON	71.03711	71.0788
R	Arg	$C_6H_{12}ON_4$	156.10111	156.1875
N	Asn	$C_4H_6O_2N_2$	114.04293	114.1038
D	Asp	$C_4H_5O_3N$	115.02694	115.0886
C	Cys	C_3H_5ONS	103.00919	103.1388
E	Glu	$C_5H_7O_3N$	129.04259	129.1155
Q	Gln	$C_5H_8O_2N_2$	128.05858	128.1307
G	Gly	C_2H_3ON	57.02146	57.0519
H	His	$C_6H_7ON_3$	137.05891	137.1411
I	Ile	$C_6H_{11}ON$	113.08406	113.1594
L	Leu	$C_6H_{11}ON$	113.08406	113.1594
K	Lys	$C_6H_{12}ON_2$	128.09496	128.1741
M	Met	C_5H_9ONS	131.04049	131.1926
F	Phe	C_9H_9ON	147.06841	147.1766
P	Pro	C_5H_7ON	97.05276	97.1167
S	Ser	$C_3H_5O_2N$	87.03203	87.0782
T	Thr	$C_4H_7O_2N$	101.04768	101.1051
W	Trp	$C_{11}H_{10}ON_2$	186.07931	186.2132
Y	Tyr	$C_9H_9O_2N$	163.06333	163.1760
V	Val	C_5H_9ON	99.06841	99.1326

Assumes amino acids are linked by peptide bonds (H_2O removed)

See http://ca.expasy.org/tools/pi_tool.html

Absorbance



Beer's Law

$A = \varepsilon\ c\ l$

A = absorbance
ε = extinction coefficient (M^{-1} 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 (ε = 60 M^{-1} cm^{-1})

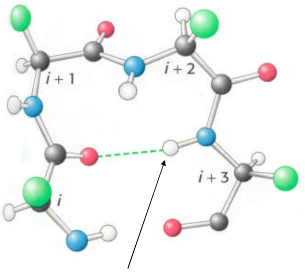
ε (Trp) ≈ 5690 M^{-1} cm^{-1}
ε (Tyr) ≈ 1280 M^{-1} cm^{-1}

See http://ca.expasy.org/tools/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

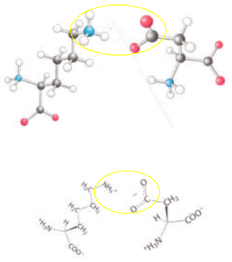
Hydrogen Bond Interactions



Sharing of hydrogen

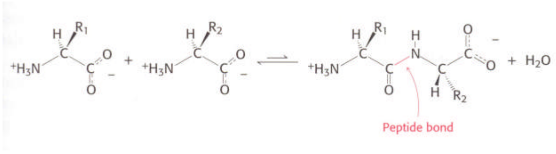
Side chains of W, Y, S, T, N, Q, D, E, K, R

Salt Bridge Interactions

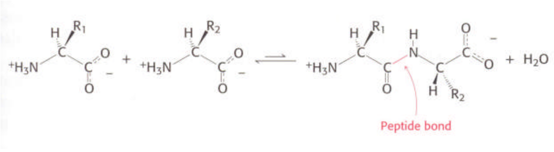


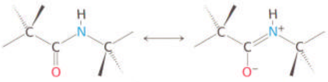
Charge – charge interactions
Typically between (D or E) and (K, R or H)

Peptide Bond



Peptide Bond

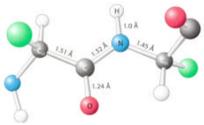


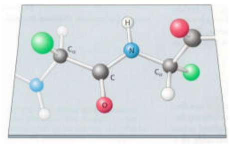


Peptide bond resonance structures

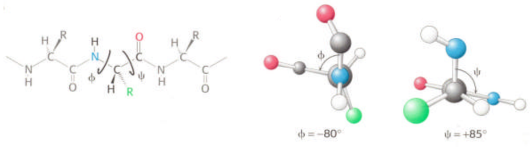
Additional stability due to resonance

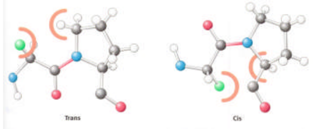
Peptide Bond Planarity





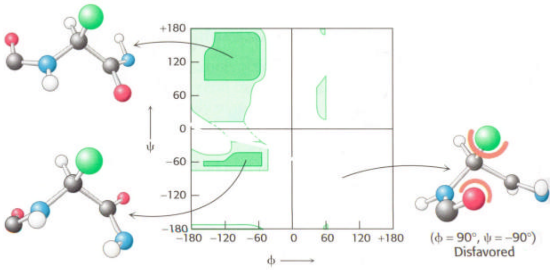
Main Chain Geometry (phi and psi)





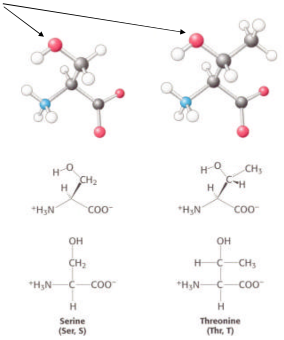
Omega can be 180° (trans) or 0° (cis) in proline

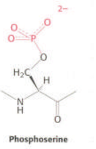
Ramachandran Plot



Posttranslational Modifications

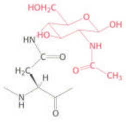
PO_4



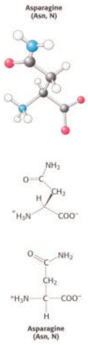


Phosphorylation

Posttranslational Modifications

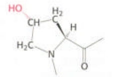


Carbohydrate–asparagine adduct

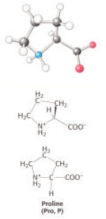


Glycosylation

Posttranslational Modifications



Hydroxyproline



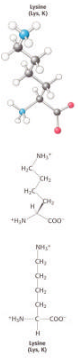
Hydroxylation

Posttranslational Modifications

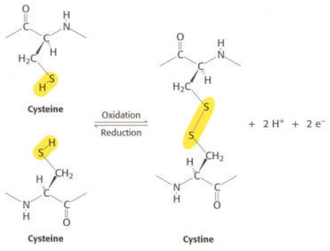
Methylation

Hydroxylation

Acetylation (“histone code”)

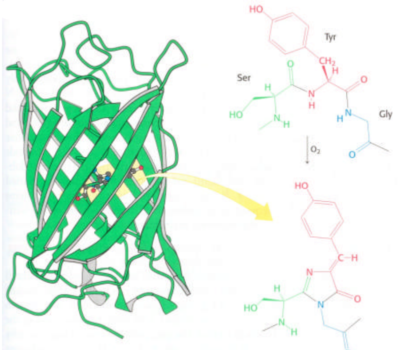


Posttranslational Modifications



Disulfide formation

Posttranslational Modifications



Green Fluorescent Protein

Next Class
Protein Structure