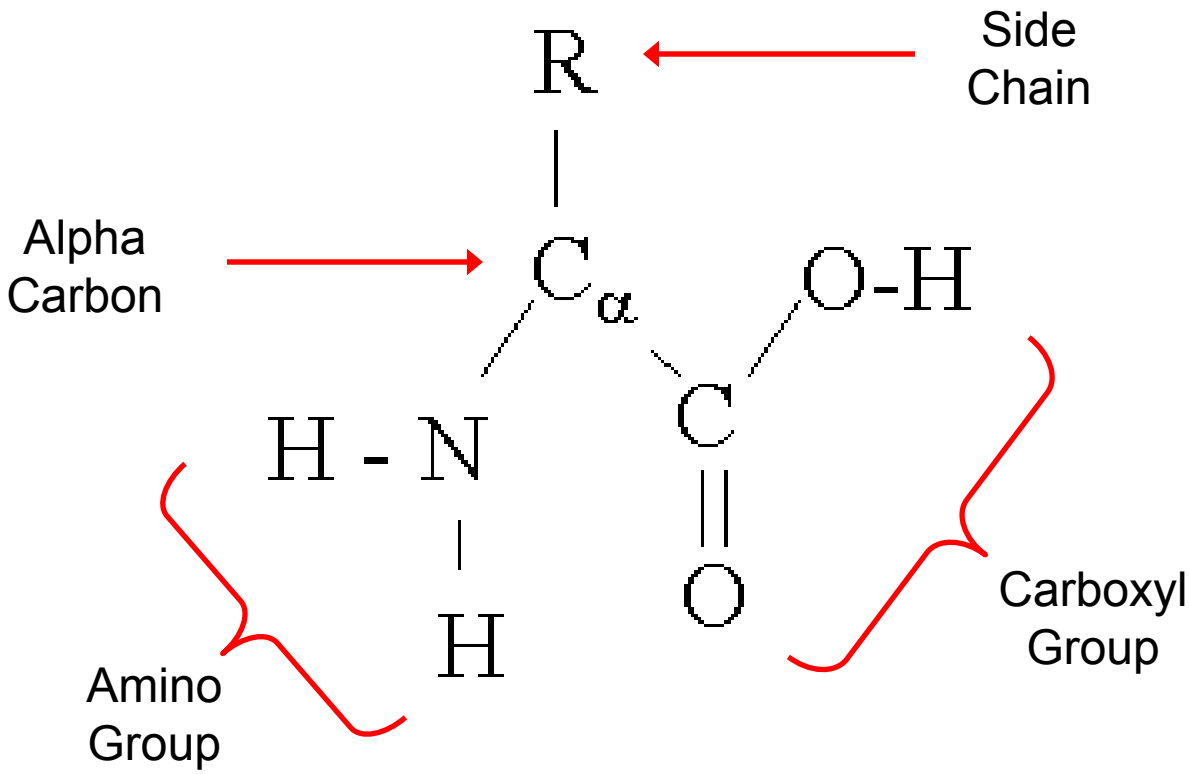
Protein Structure

Amino Acids

Amino acids are the building blocks of proteins. All AA's have the same basic structure:

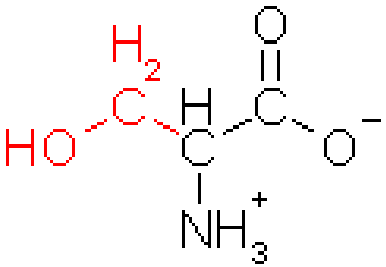


Amino Acid Properties

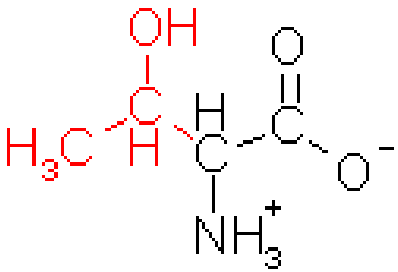
- There are 20 different, naturally occurring amino acids
- The properties of each amino acid are determined by its specific side chain
- Amino acid names are often abbreviated as either three letters or single letters (worth knowing)

SEE HANDOUT ON AMINO ACIDS

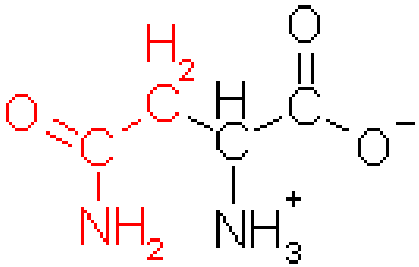
Polar Residues



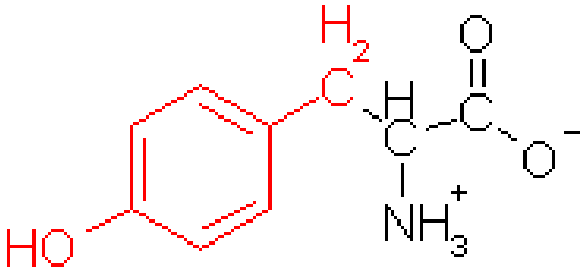
Serine (Ser)



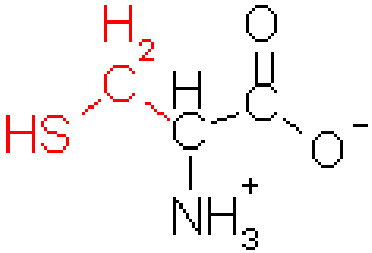
Threonine (Thr)



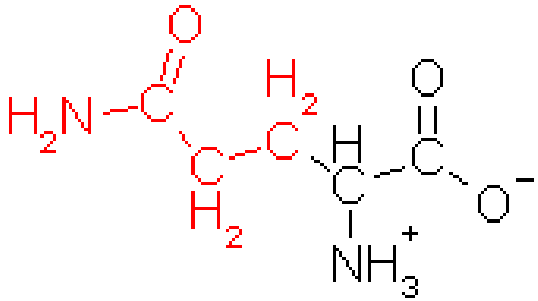
Asparagine (Asn)



Tyrosine (Tyr)

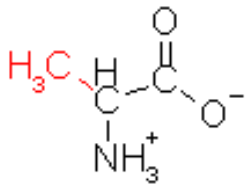


Cysteine (Cys)

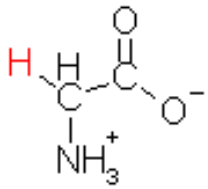


Glutamine (Gln)

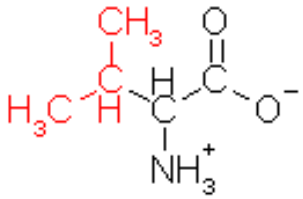
Nonpolar Residues



Alanine (Ala)

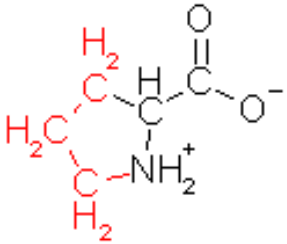


Glycine (Gly)

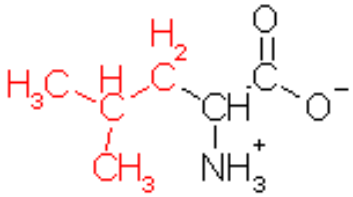


Valine (Val)

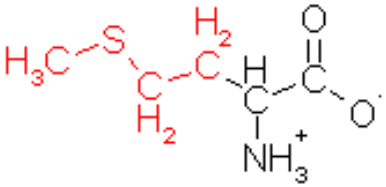
Isoleucine (Ile)



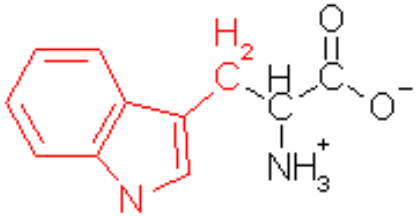
Proline (Pro)



Leucine (Leu)



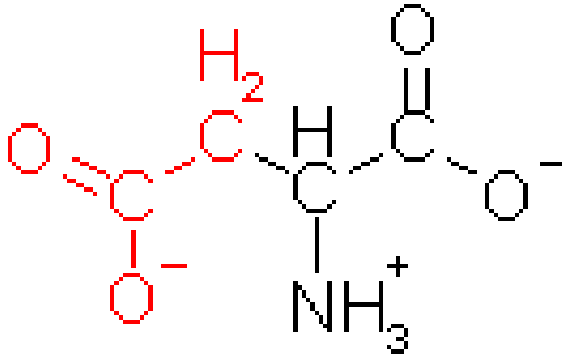
Methionine (Met)



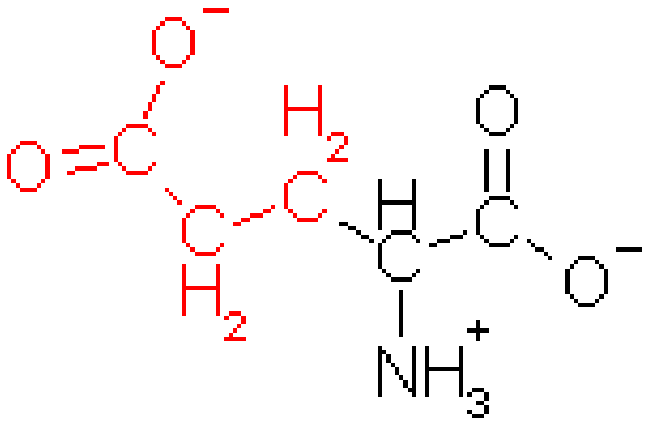
Tryptophan (Trp)

Phenylalanine (Phe)

Acidic Residues

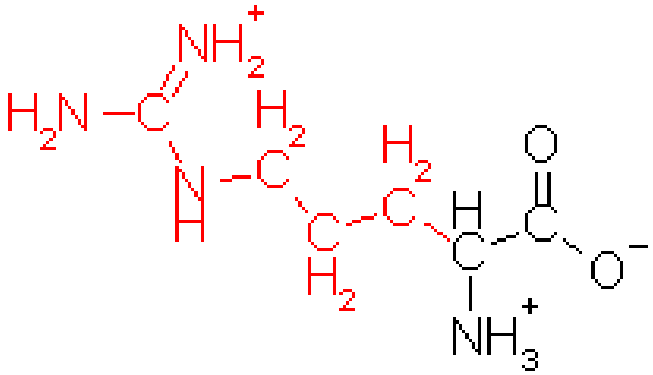


Aspartic Acid (Asp)

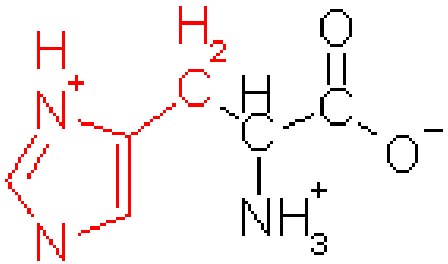


Glutamic Acid (Glu)

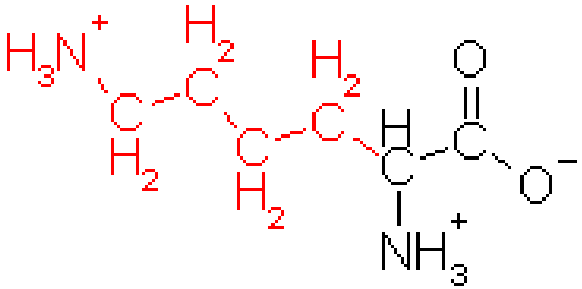
Basic Residues



Arginine (Arg)



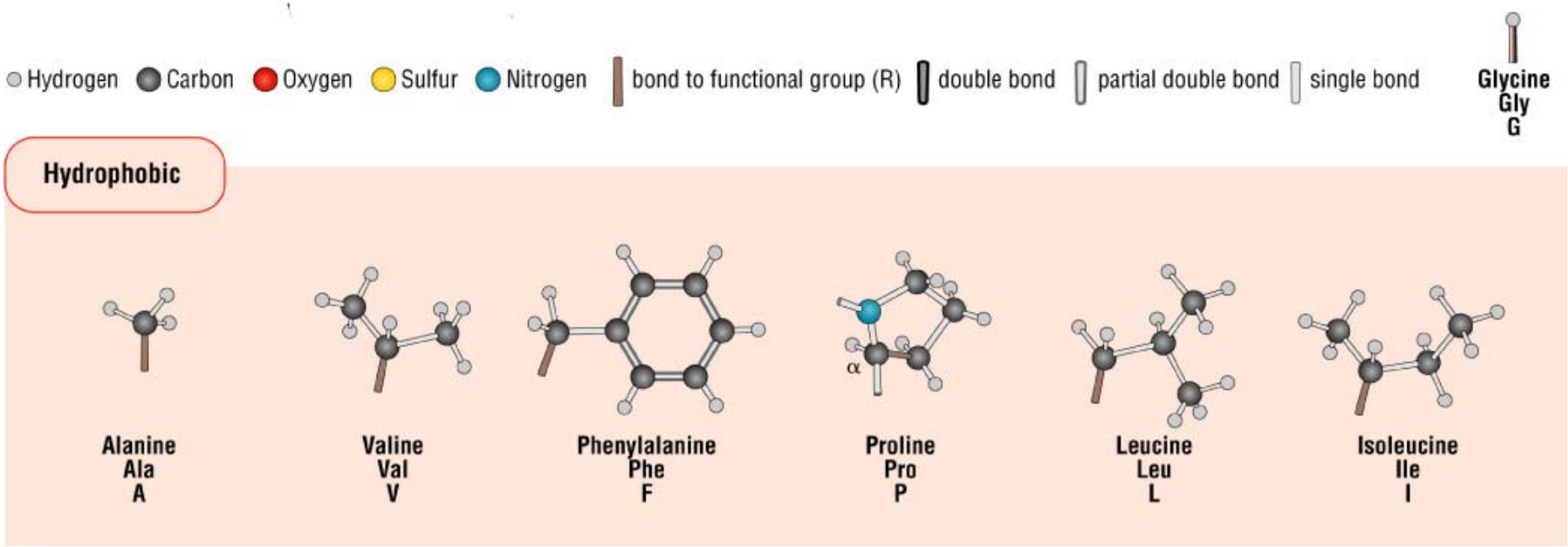
Histidine (His)

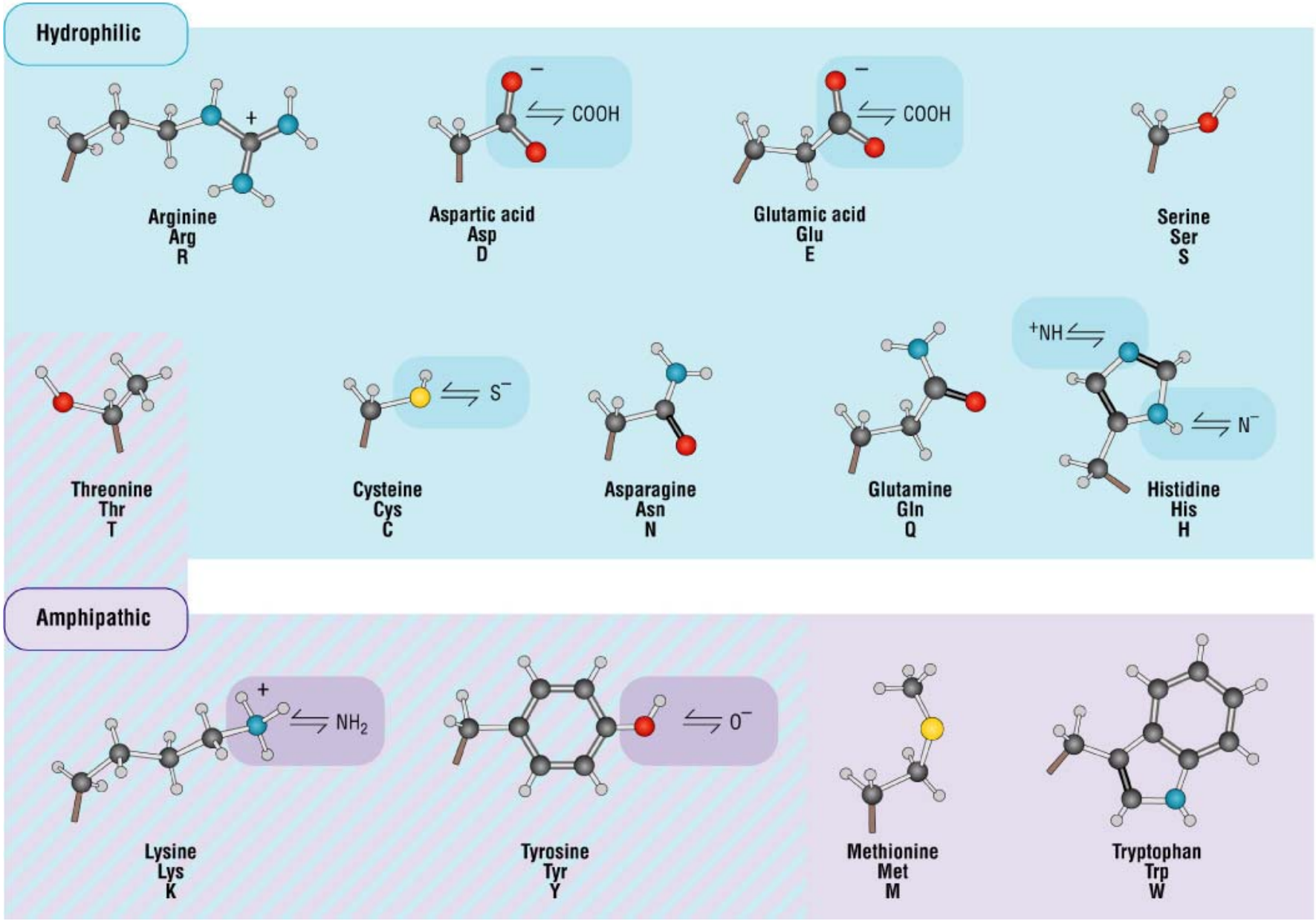


Lysine (Lys)

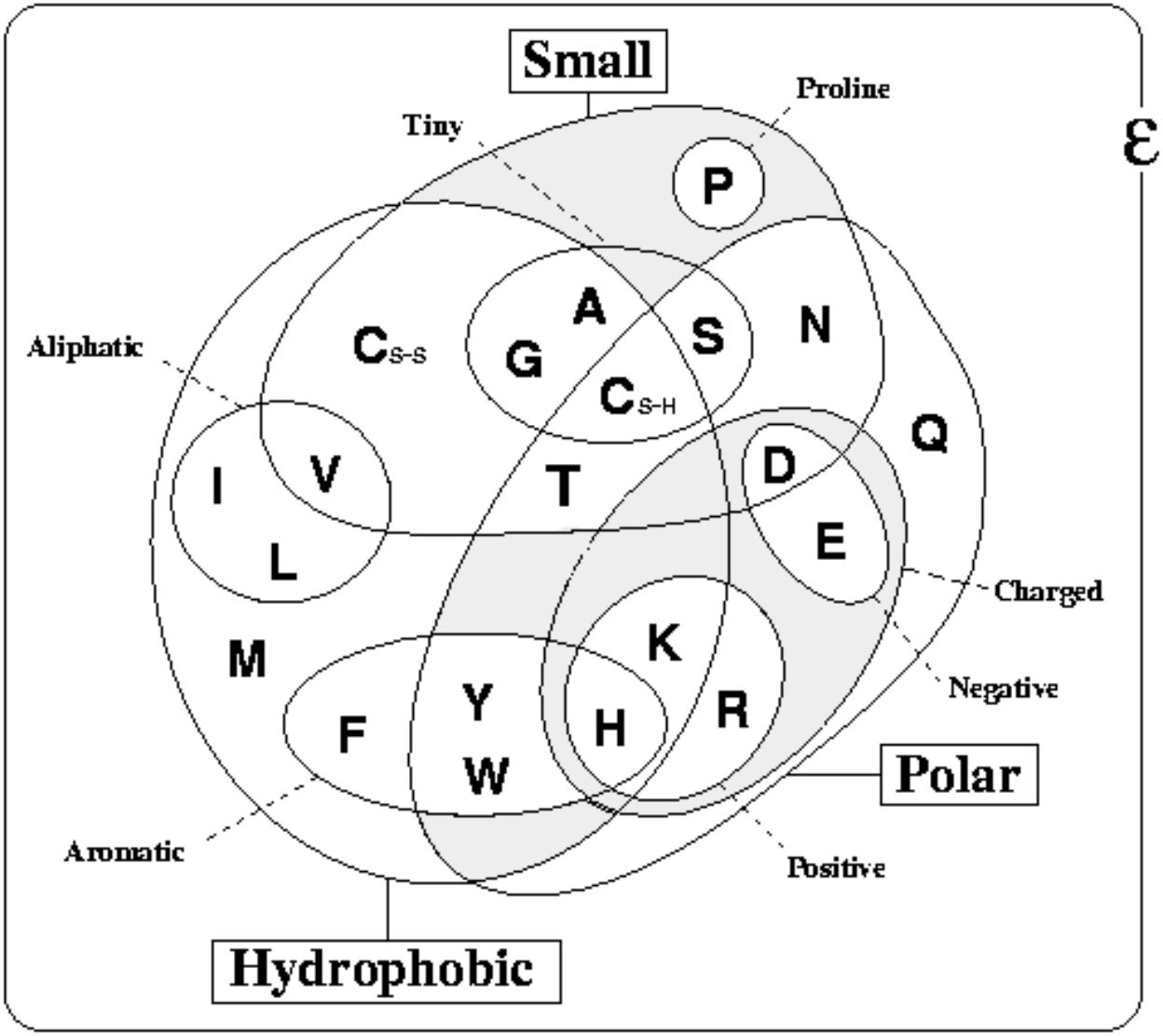
Amino Acids

More depictions from Petsko and Ringe




Hydrophilic
Arginine
Arg
R
Aspartic acid
Asp
D
COOH
Glutamic acid
Glu
E
COOH
Serine
Ser
S
Threonine
Thr
T
Cysteine
Cys
C
S⁻
Asparagine
Asn
N
Glutamine
Gln
Q
⁺NH
N⁻
Histidine
His
H
Amphipathic
Lysine
Lys
K
NH2
Tyrosine
Tyr
Y
O⁻
Methionine
Met
M
Tryptophan
Trp
W

Amino Acid Properties



pK_a

- You know that the pH is defined in terms of the proton concentration

$$pH = -\log[H^+]$$

and that this is based on an equilibrium that is reached between an acid (or base) and it constituent parts

$$HA \leftrightarrow H^+ + A^-$$

pK_a

- The equilibrium constant for this reaction is given by

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

- If we solve this equation for $[H^+]$ we get

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

pK_a

- Taking the log of each side (with a minus sign) gives us

$$-\log[H^+] = -\log K_a + \log\frac{[A^-]}{[HA]}$$

$$pH = pK_a + \log\left[\frac{base}{acid}\right]$$

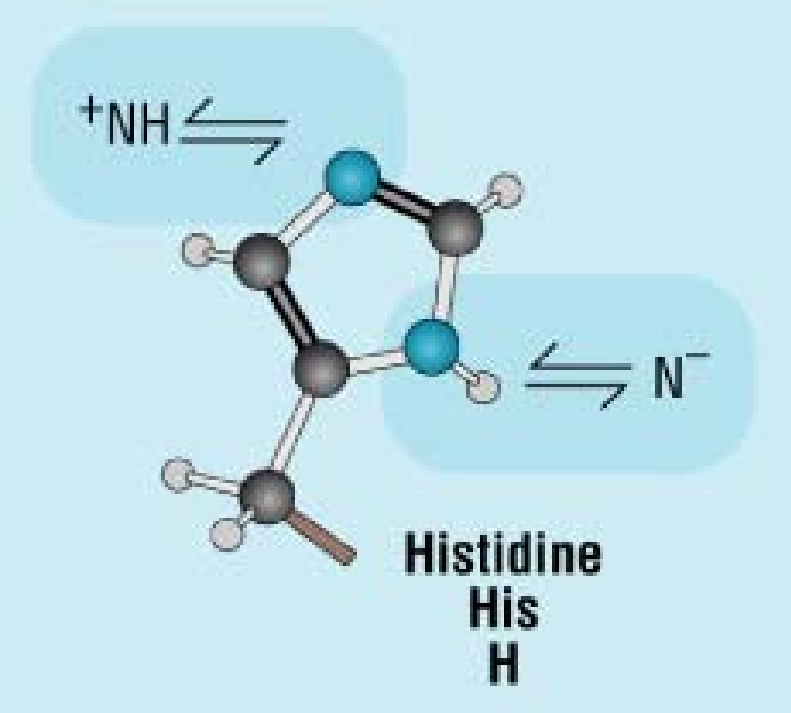
- This is the Henderson-Hasselbach equation

pK_a of amino acids

AA	COOH	NH_3^+	R	AA	COOH	NH_3^+	R
Ala	2.35	9.87	-	Leu	2.33	9.74	-
Arg	1.82	8.99	12.48	Lys	2.16	9.18	10.79
Asn	2.10	8.84	-	Met	2.13	9.28	-
Asp	1.99	9.90	3.90	Phe	2.16	9.18	-
Cys	1.92	10.78	8.33	Pro	1.95	10.65	-
Glu	2.10	9.47	4.07	Ser	2.19	9.21	~13
Gln	2.17	9.13	-	Thr	2.09	9.10	~13
Gly	2.35	9.78	-	Trp	2.43	9.44	-
His	1.80	9.33	6.04	Tyr	2.20	9.11	10.13
Ile	2.32	9.76	-	Val	2.29	9.74	-

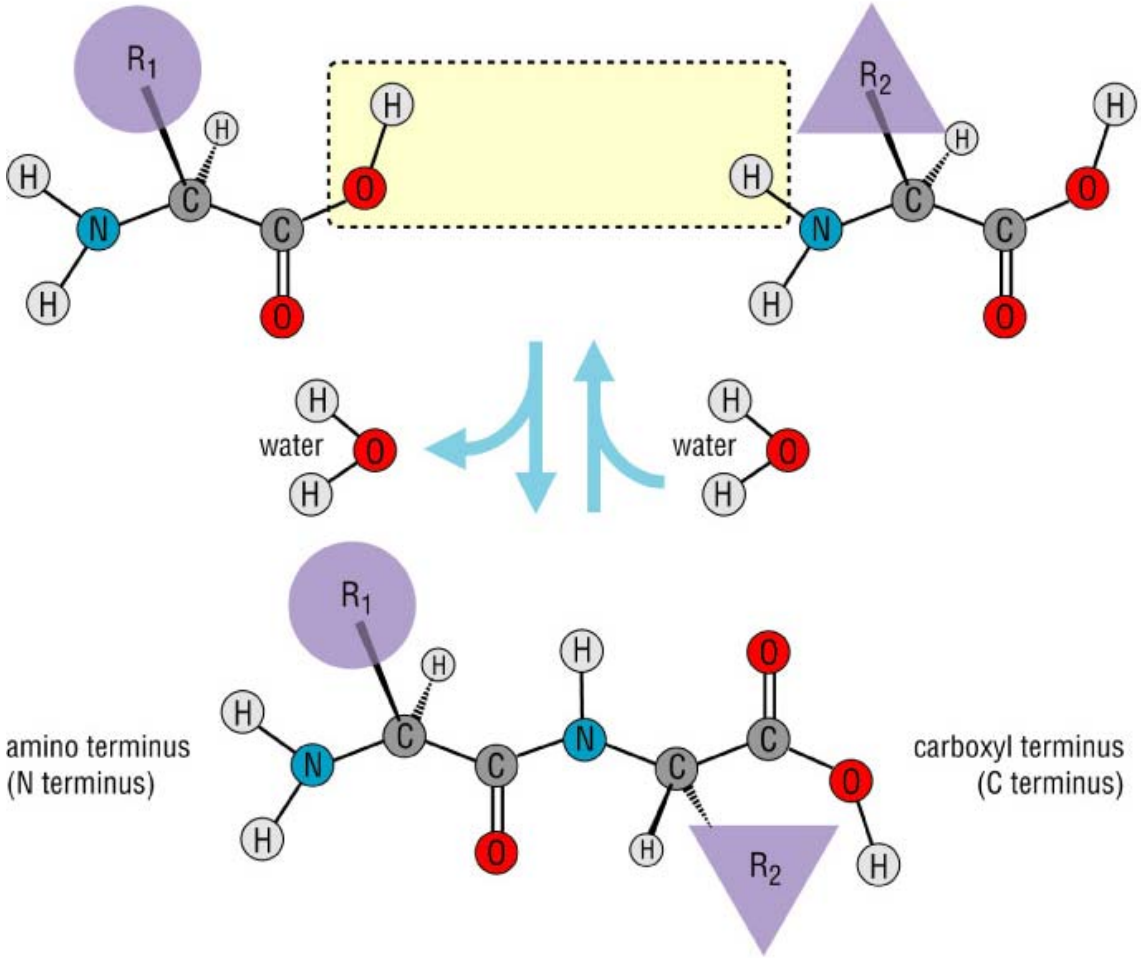
Histidine

Since the pK_a of histidine is close to neutral, its protonation state depends strongly on its local environment. This feature is often exploited and histidine is used as a molecular switch.



The Peptide Bond

To make a protein, these amino acids are joined together in a polypeptide chain through the formation of a peptide bond.

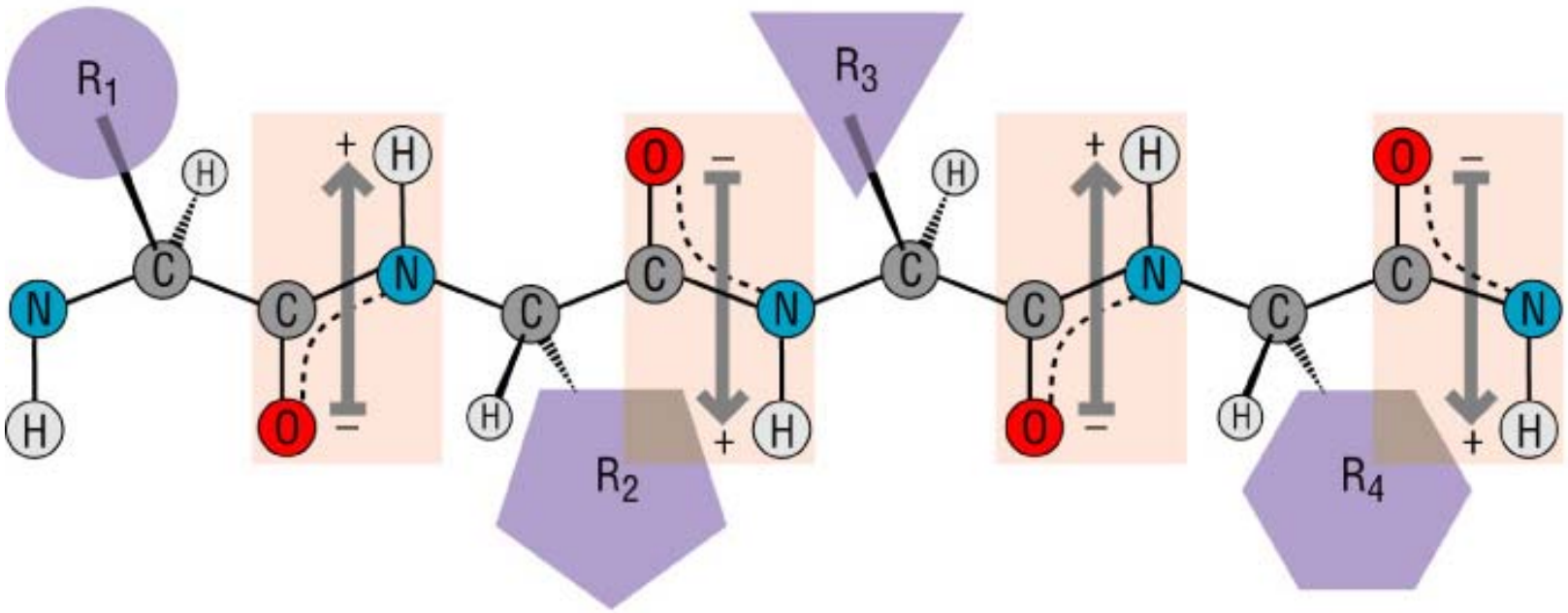


Polypeptides

- Proteins are nothing more than long polypeptide chains.
- Chains that are less than 40-50 amino acids or residues are often referred to as polypeptide chains since they are too smal to form a functional domain.
- Larger than this size, they are called proteins

- The structure, function and general properties of a protein are all determined by the sequence of amino acids that make up its primary sequence.

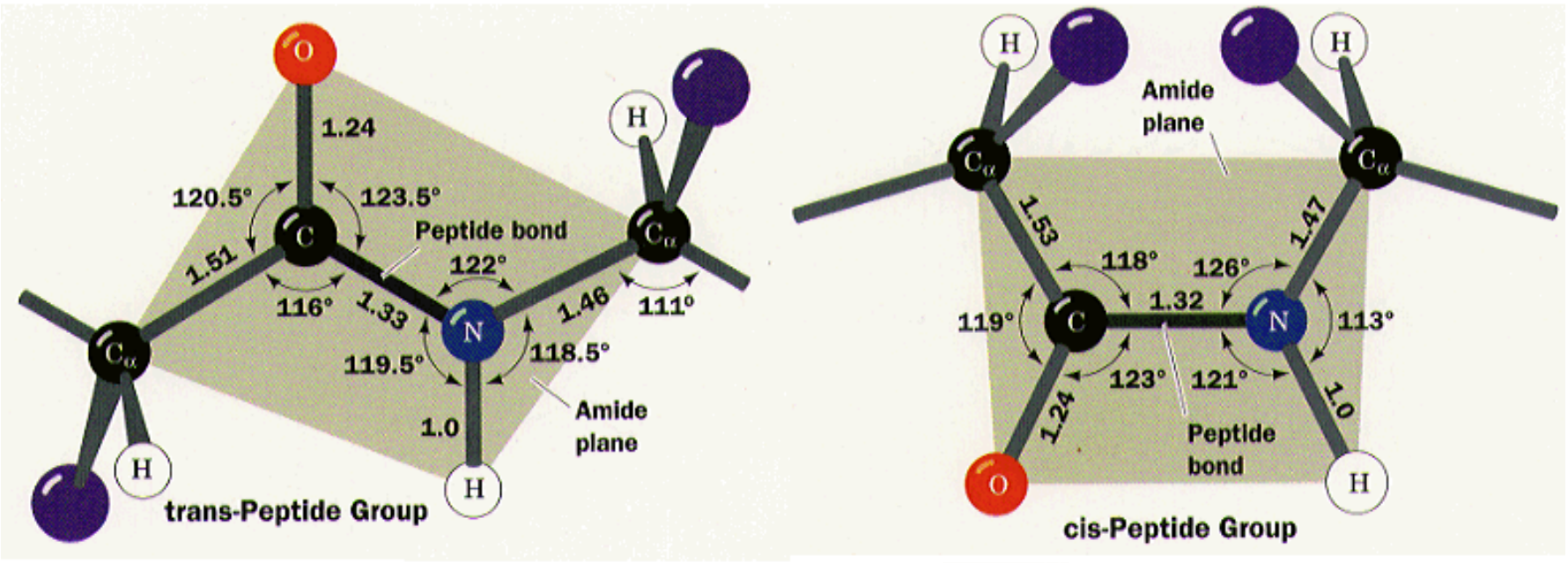
Primary Structure



…-ASP-ALA-VAL-ILE-ASP-SER-GLU-PRO-THR-…

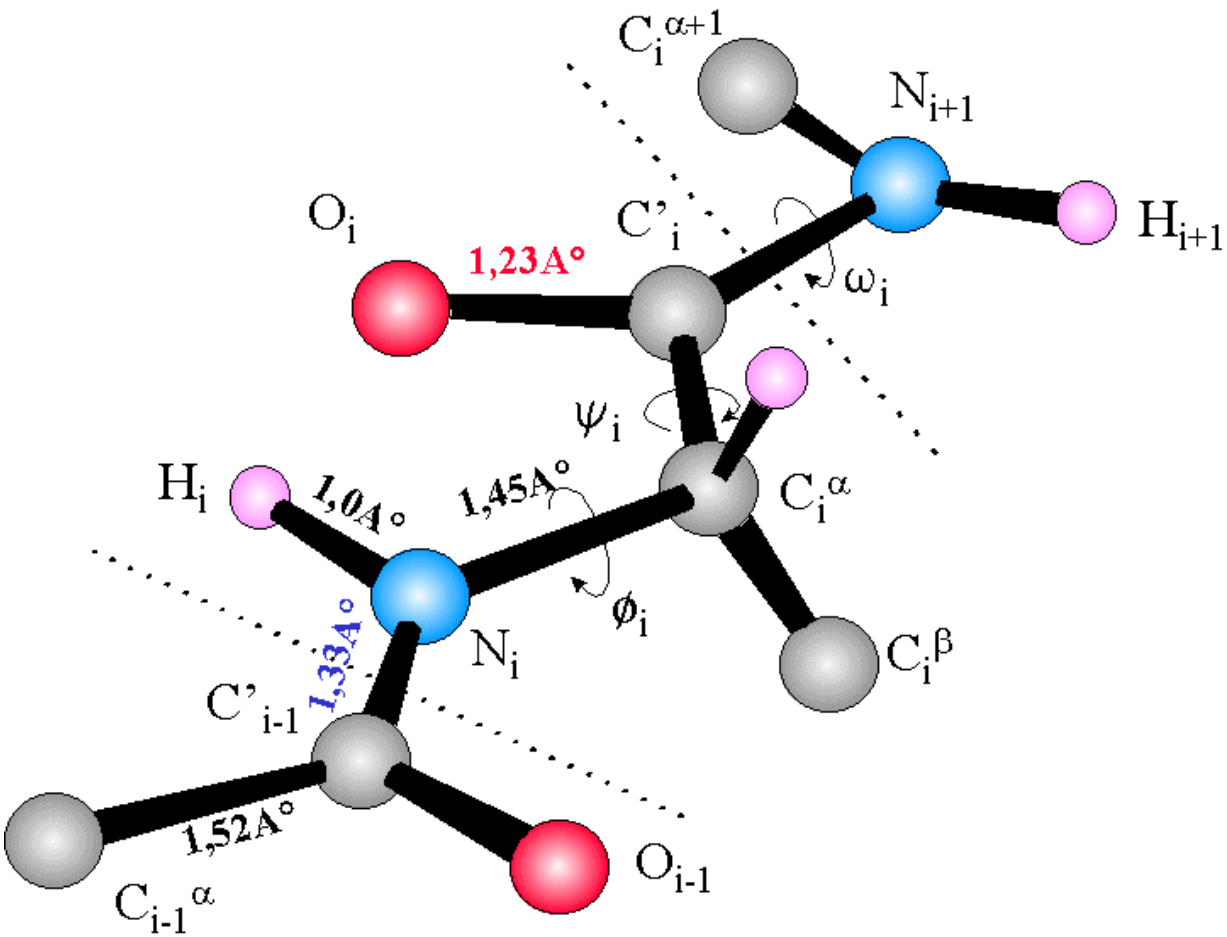
…DAVIDSEPT…

Angles and Bonds



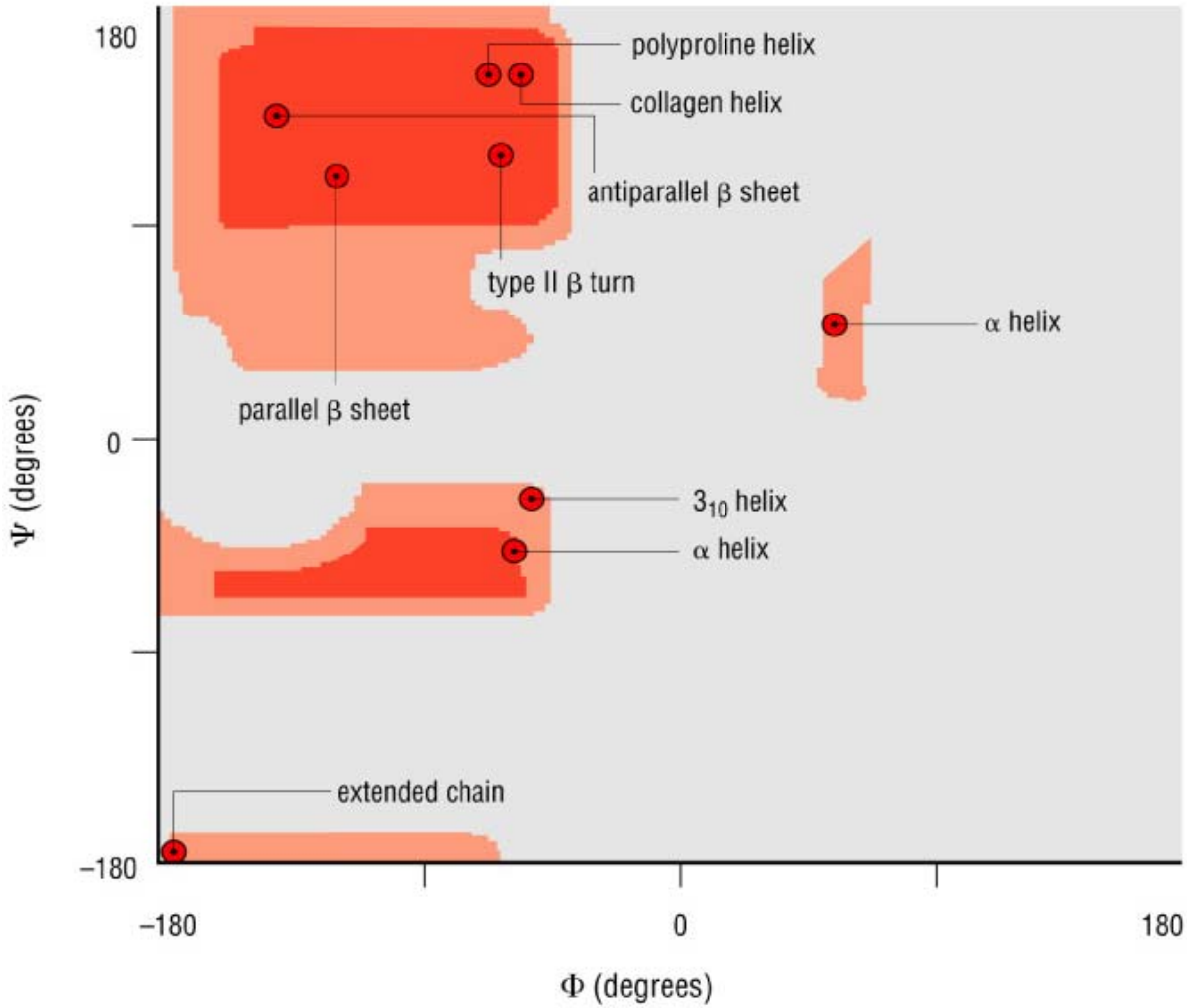
trans (ω=180) is strongly favored over cis (ω=0)

(ϕ,ψ) Angles


$C_i^{\alpha+1}$
N_{i+1}
H_{i+1}
O_i
C'_i
1,23A°
ω_i
ψ_i
H_i
1,04°
1,45A°
C_i^{α}
ϕ_i
N_i
1,33A°
C_i^{β}
C'_{i-1}
1,52A°
O_{i-1}
C_{i-1}^{α}

The Ramachandran Plot

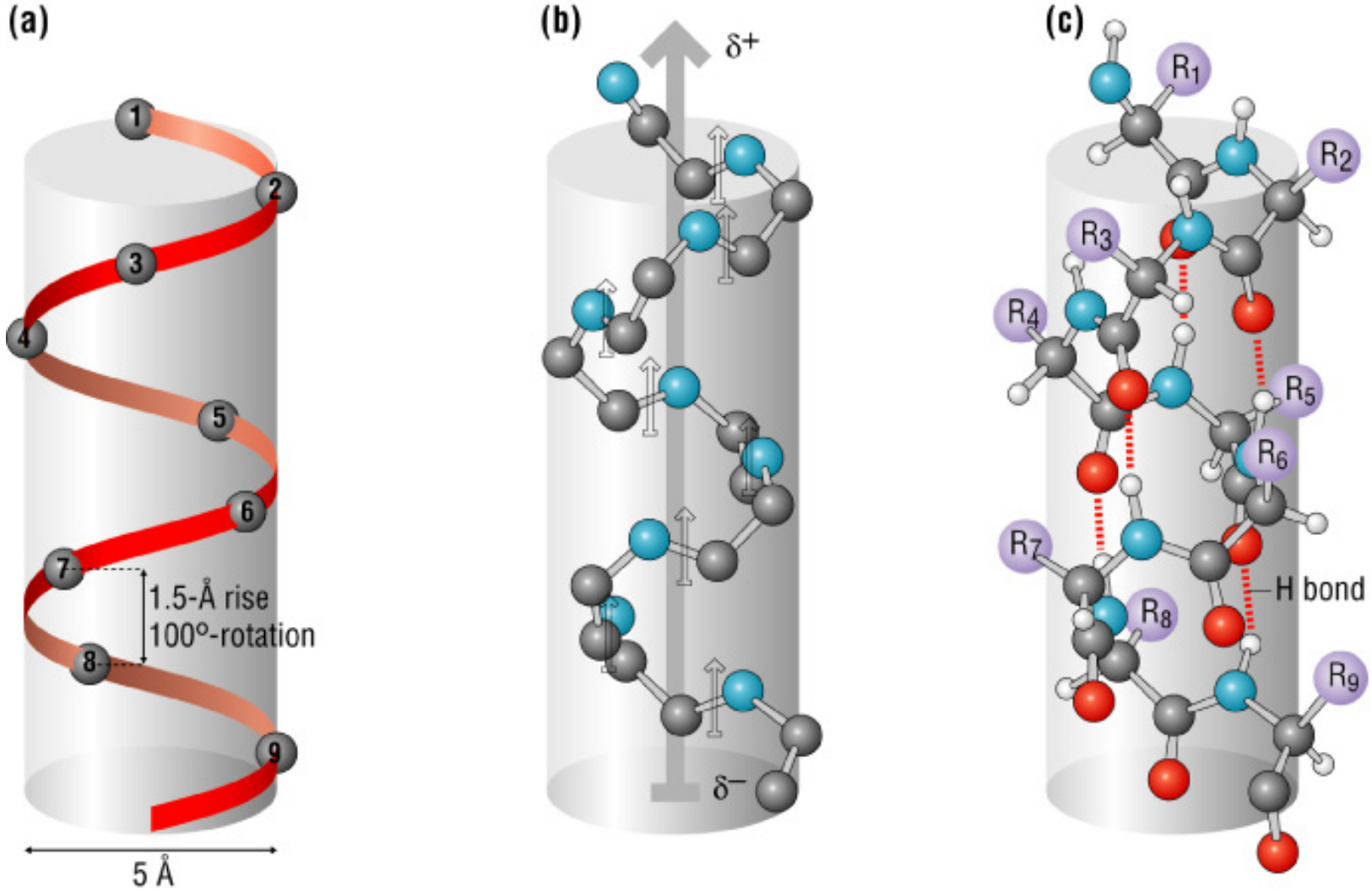
The (ϕ,ψ) angles of amino acids in a polypeptide chain or protein are restricted, largely because of steric interactions (glycine is an exception).



Secondary Structure

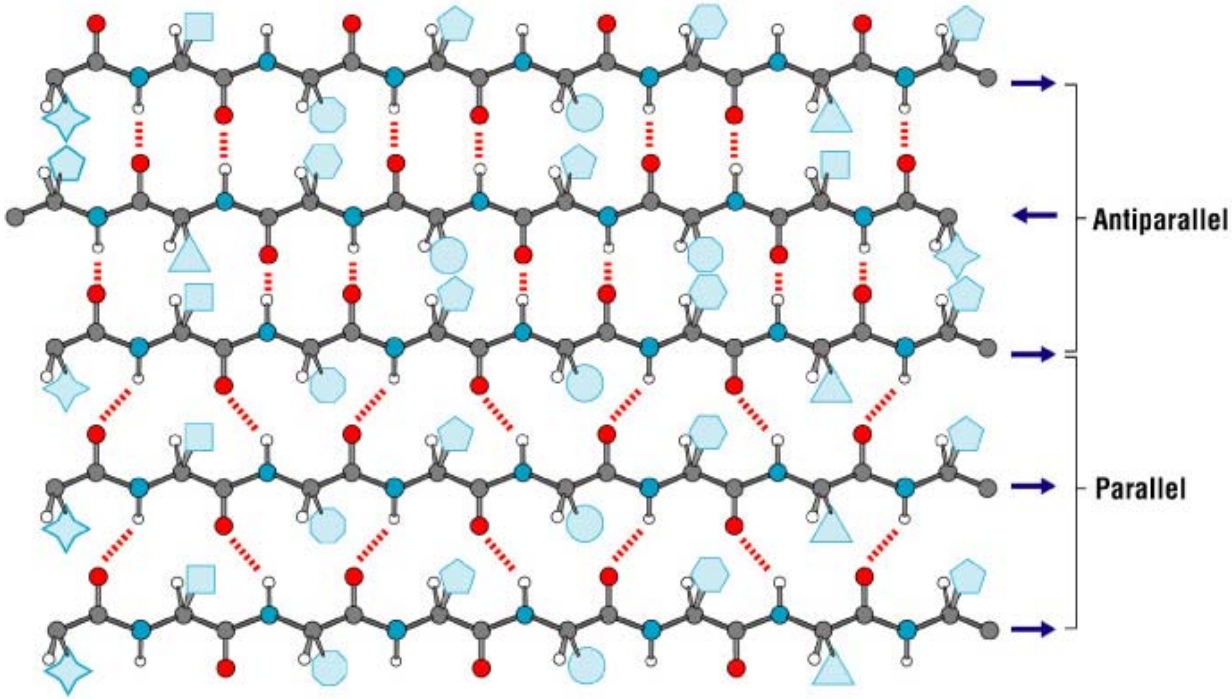
- The primary sequence or main chain of the protein must organize itself to form a compact structure. This is done in an elegant fashion by forming secondary structure elements
- The two most common secondary structure elements are alpha helices and beta sheets, formed by repeating amino acids with the same (ϕ,ψ) angles
- There are other secondary structure elements such as turns, coils, 3_{10} helices, etc.

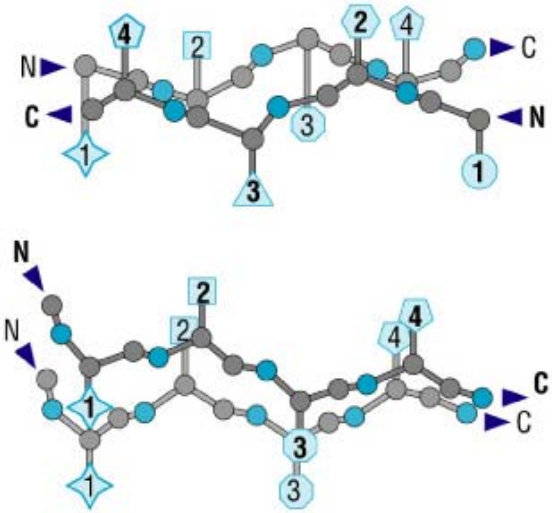
The Alpha Helix



Beta Sheets

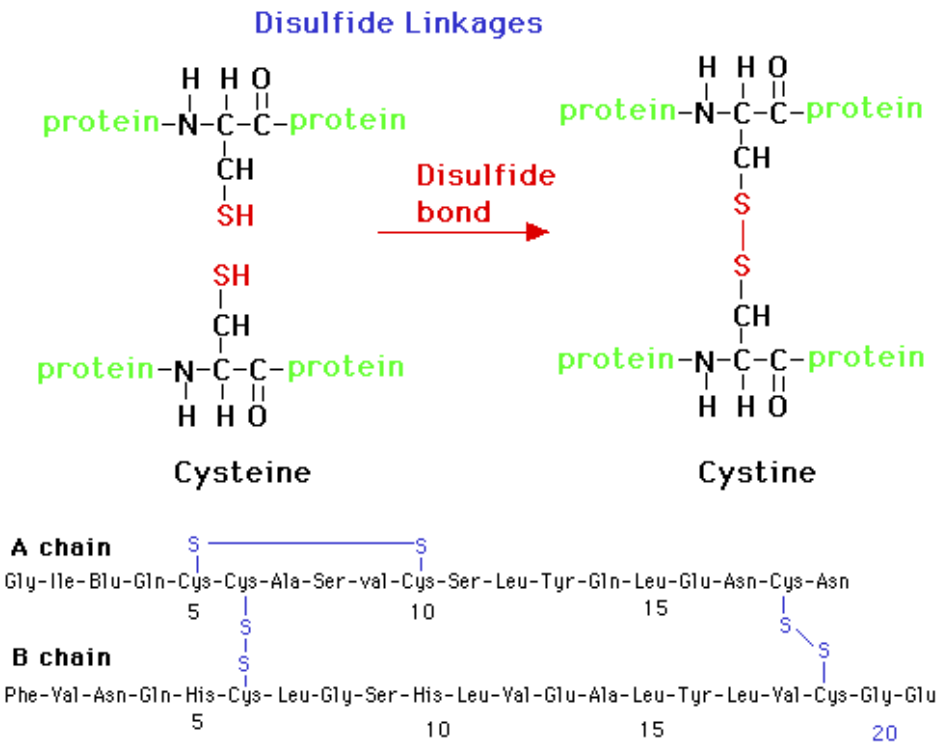
Beta sheets can be parallel or antiparallel





Disulfide Bonds

- Pairs of cysteines can form disulfide bonds between different parts of the main chain
- This adds stability and is common in extracellular proteins



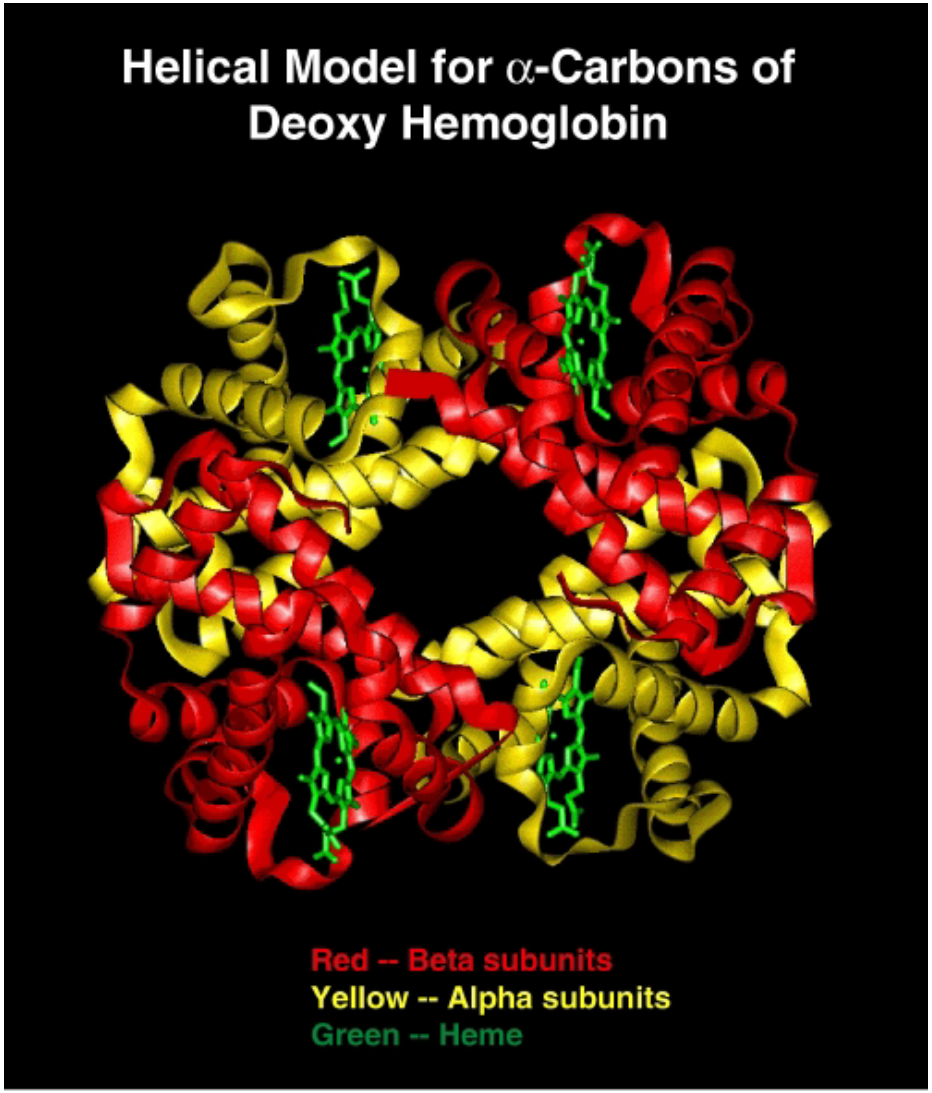
Tertiary Structure

- To make the protein look like a protein, the secondary structure elements come together to form the tertiary structure
- Most often, the secondary structure elements form motifs
  - Greek key
  - EF hand
  - Beta hairpin
  - …

Quaternary Structure

- Folded proteins then bind together to form dimer, trimers, or higher order structures

- The functional form of hemoglobin is a tetramer

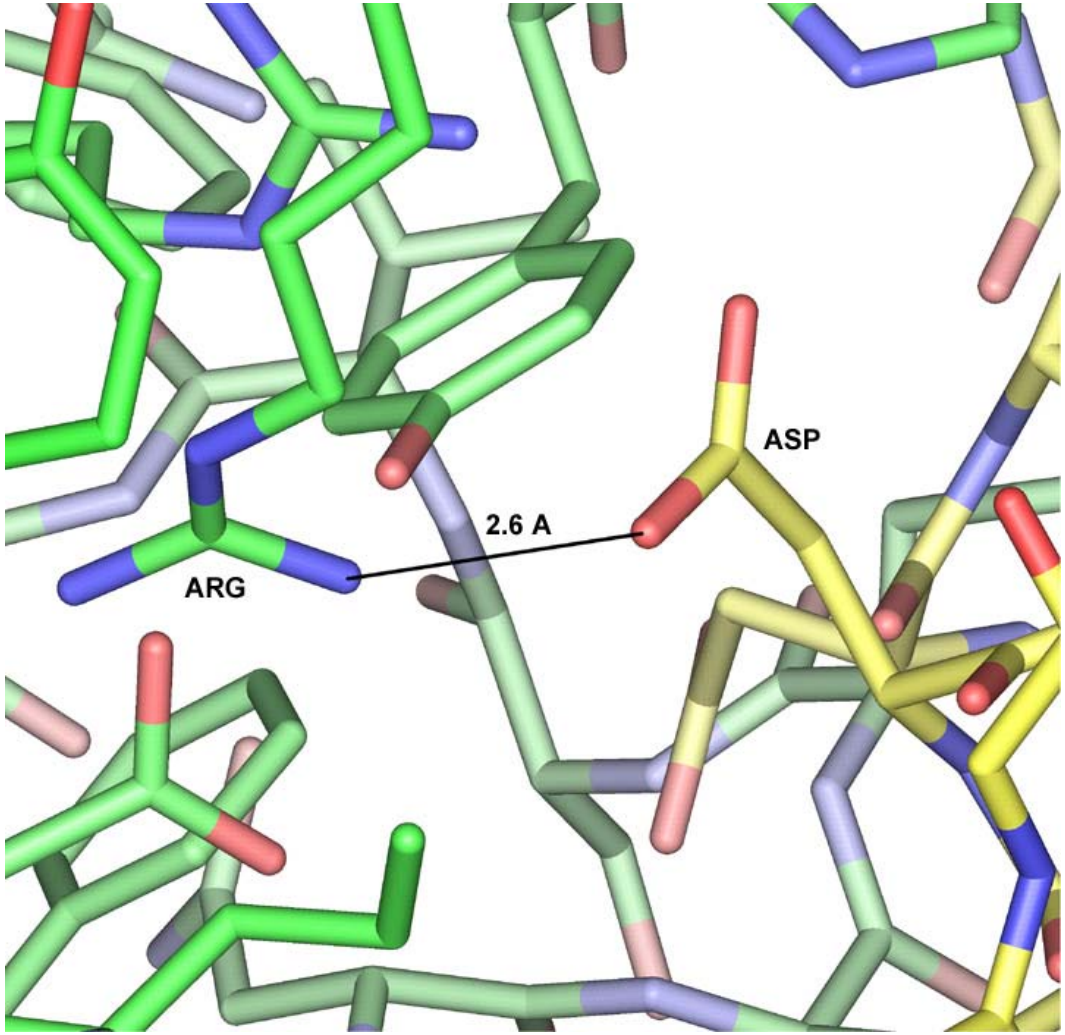


Non-Covalent Bonds

- The backbone and side chain bonds are all covalent bonds (as are disulfide bonds), but non-covalent bonds are required to maintain secondary, tertiary and quaternary structure
- These include
  - Hydrogen bonds (H-bonds)
  - Electrostatic / Salt Bridges
  - van der Waals

Salt Bridges

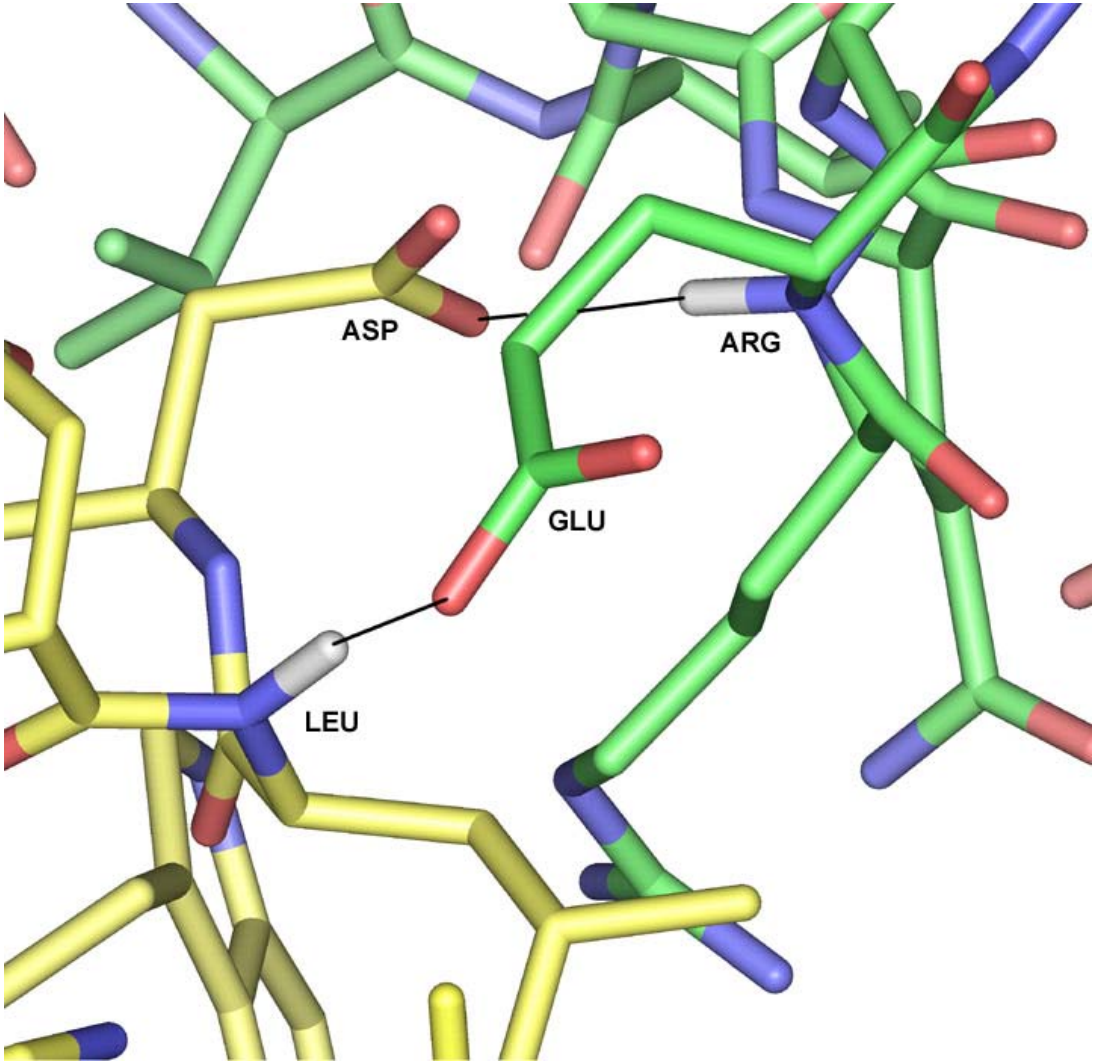
- Salt bridges are electrostatic bonds between oppositely charged groups
- The strength is usually 4-7 kcal/mol



Hydrogen Bonds

- Hydrogen bonds are formed by the sharing of a proton between donor and acceptor groups

- The strength is around 2-5 kcal/mol and the ideal distance is 2.8-3 Å

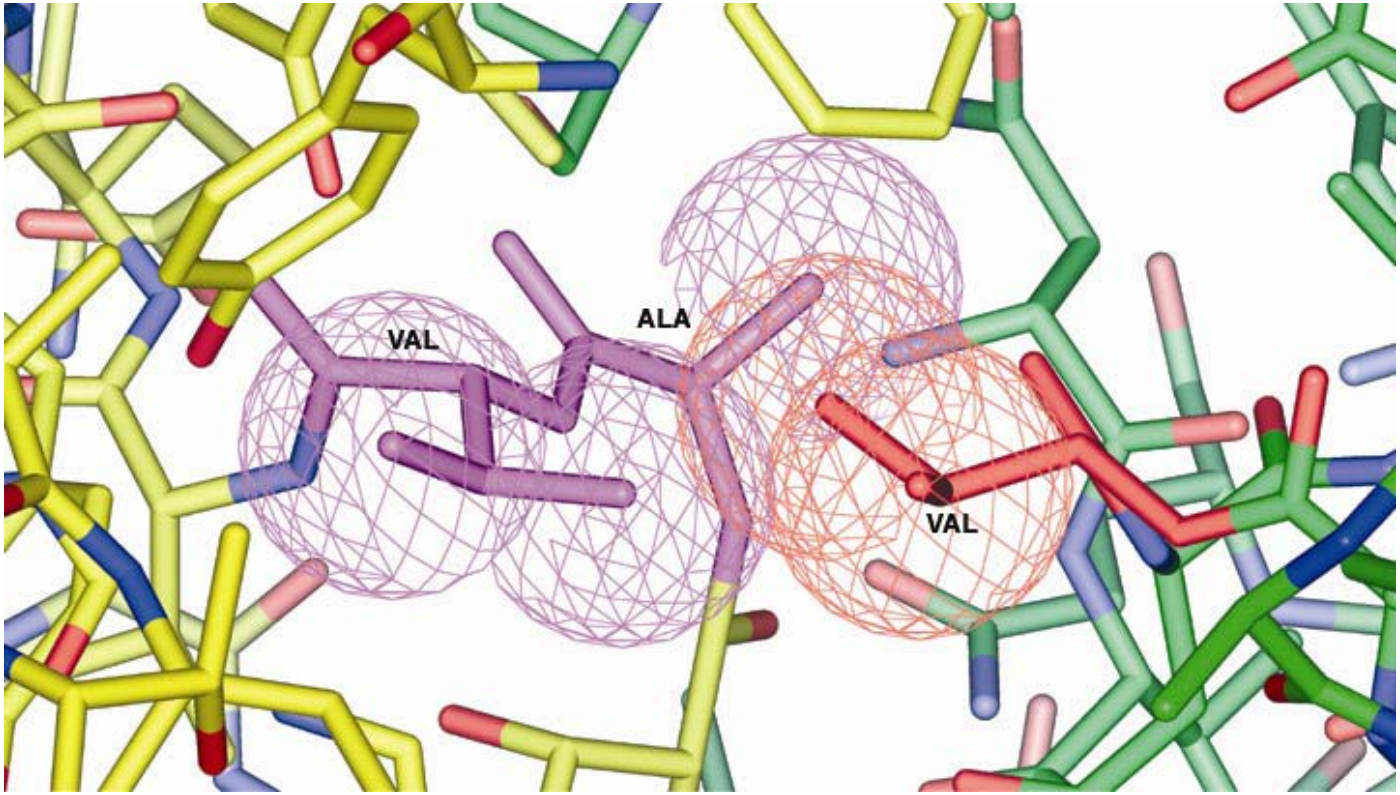


van der Waals Interactions

- Electrostatic interactions cannot account for all the non-covalent interactions observed between molecules (especially uncharged ones)
- Atoms with dipoles (and higher order multipoles) induce and interact with dipoles in other atoms via dispersion forces ($1/r^6$)

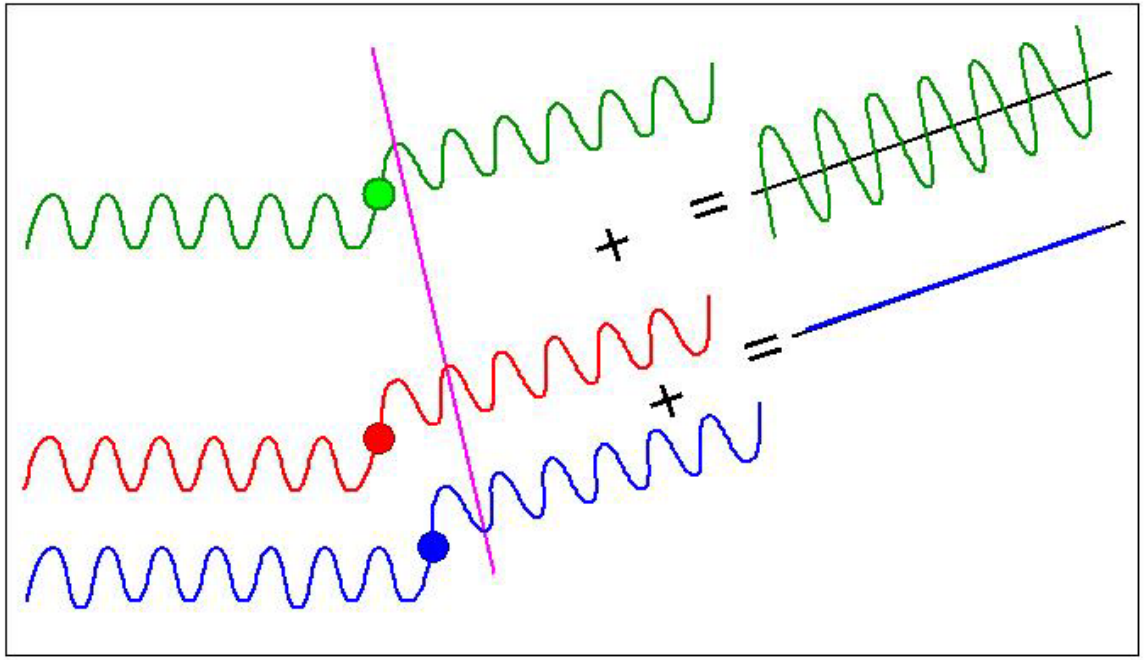
Hydrophobic Interactions

- Hydrophobic interactions are not attractive interactions, but results from the inability of water to form hydrogen bonds with certain side chains



Determining Protein Structures

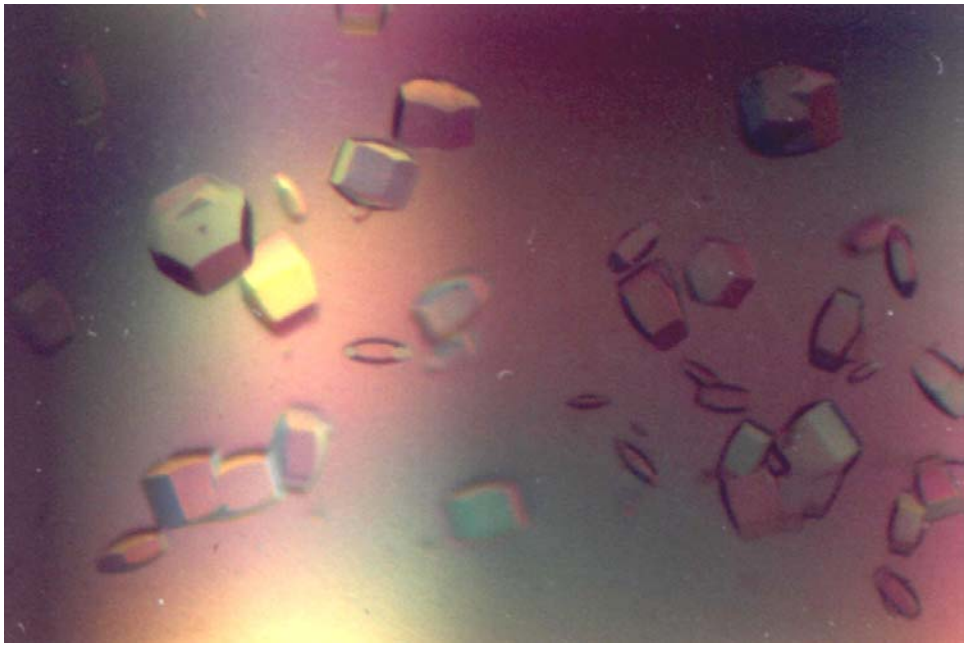
- X-ray crystallography is one of the primary means of getting high-resolution protein structures. It is based on Bragg scattering of x-rays (λ = 0.2 – 2 Å) from electron density surrounding the atoms in a protein. Higher electron density leads to more scattering.



Crystallization

- In order to have coherent scattering, the protein must first be crystallized. In many cases this is the most difficult part since proteins do not naturally form crystals

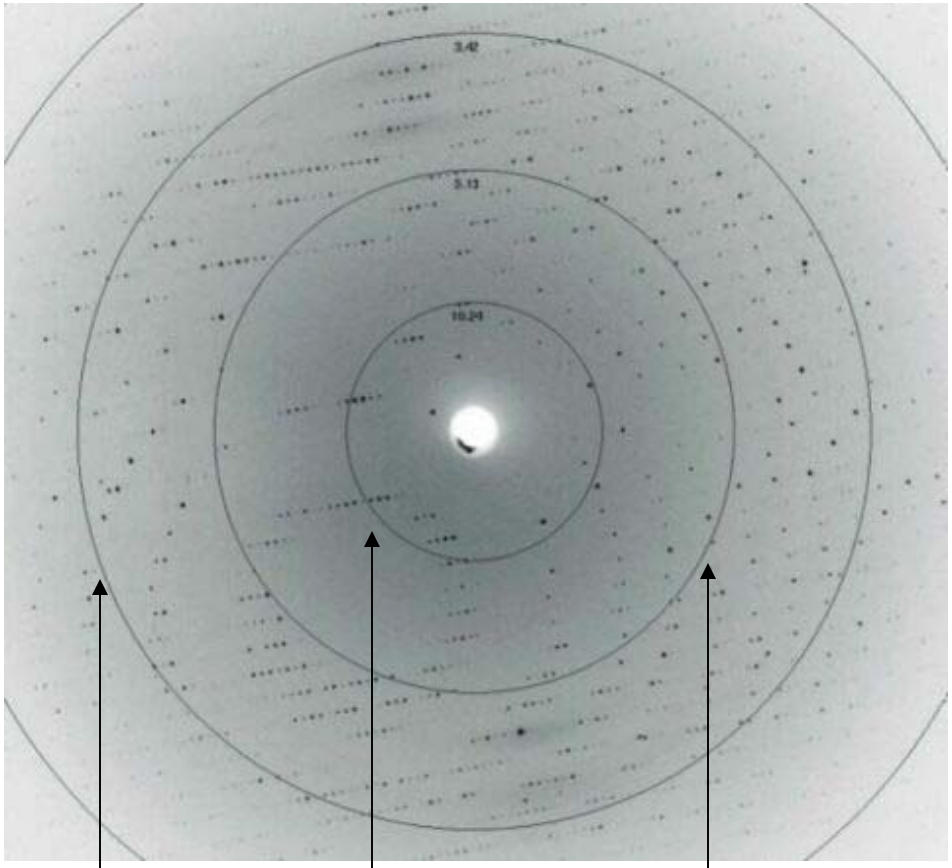
To induce crystallization, scientists must often remove flexible parts of the protein, try to crystallize a complex of the protein, etc.



Diffraction

- If successful in forming a regular crystal, the hope is that they now diffract to a high enough resolution

- This scattering density is then transformed in real space coordinates

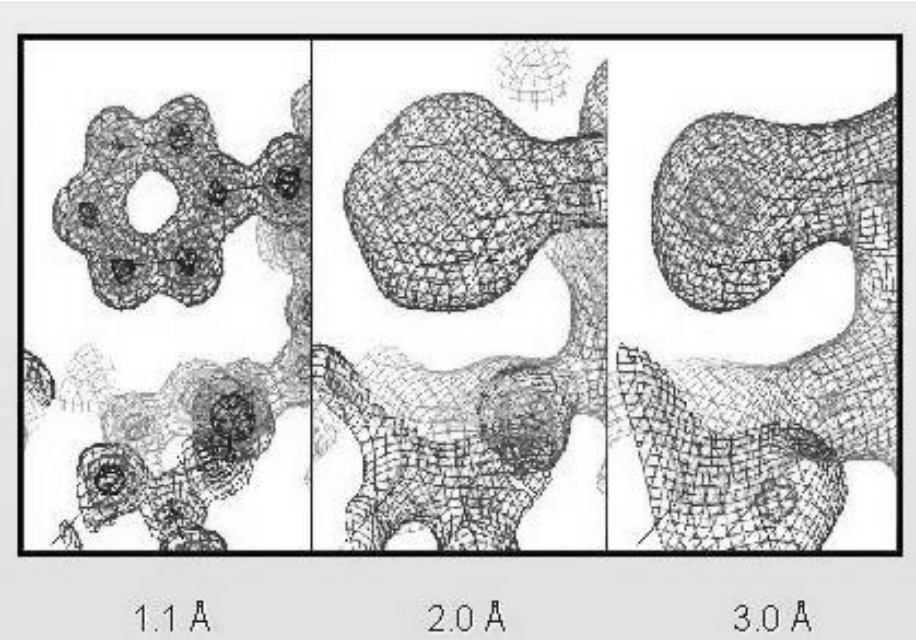


3 Å 10 Å 5 Å

Resolution of Structures

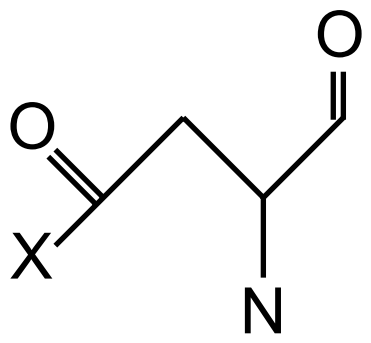
Hydrogens are typically too small to be resolved except in the highest resolution structures (< 1 Å)

- 5 Å structures resolve some secondary structure and can be useful
- 2.5 – 3 Å is more typical – constrains (ϕ,ψ) angles
- less that 1.5 Å is a very good structure - (ϕ,ψ) angles are well defined



Sequence to Structure

- Even with the highest resolution structures it is difficult or impossible to tell the difference between a N, O or C. Thus you need to know the sequence to *thread* the structure and judge the atom based on the local environment



ASP: X=O (charged)

ASN: X=N (polar)

Could it be LEU ??

Other Considerations

- Some parts of the protein may be variable or highly flexible. This means
  - They may not be resolved
  - There may be multiple orientations
  - They might have a large temperature factor (also called a β factor) – 20=good, 80=poor

- Crystallization or formation of a protein complex *may* distort the structure

PDB Files

	atom no.	atom type	residue name	residue no.	x	y	z	occ	β/TF
ATOM	1	N	ARG	1	31.758	13.358	-13.673	1.00	18.79
ATOM	2	CA	ARG	1	31.718	13.292	-12.188	1.00	14.26
ATOM	3	C	ARG	1	33.154	13.224	-11.664	1.00	18.25
ATOM	4	O	ARG	1	33.996	12.441	-12.225	1.00	20.10
ATOM	5	CB	ARG	1	30.886	12.103	-11.724	1.00	16.74
ATOM	6	CG	ARG	1	29.594	11.968	-12.534	1.00	15.96
ATOM	7	CD	ARG	1	28.700	13.182	-12.299	1.00	15.45
ATOM	8	NE	ARG	1	27.267	12.895	-12.546	1.00	12.82
ATOM	9	CZ	ARG	1	26.661	13.087	-13.727	1.00	17.38
ATOM	10	NH1	ARG	1	27.370	13.558	-14.735	1.00	18.38
ATOM	11	NH2	ARG	1	25.367	12.797	-13.838	1.00	25.73
ATOM	12	N	PRO	2	33.800	13.936	-10.586	1.00	17.07
ATOM	13	CA	PRO	2	34.976	13.367	-9.840	1.00	14.99

...

Structure and PDB Examples

- Go to the pdb

  http://www.rcsb.org/pdb/

  and search for the BPTI structure (1BPI) or some other structure of interest.
- More examples in Rasmol tutorial on class webpage