Protein Structure, Function and Disease

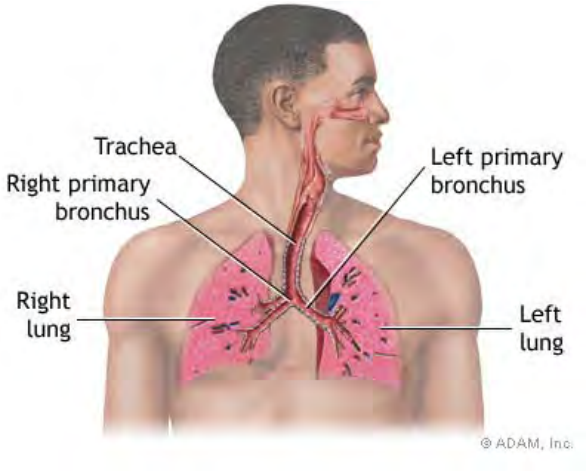
Amino acids: the building blocks of proteins

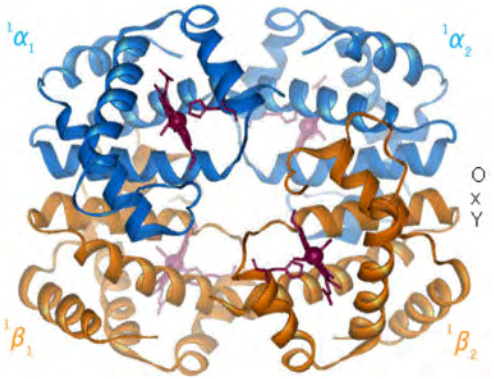
(Partially adopted from Prof. John Baenziger's former lectures)

Jyh-Yeuan (Eric) Lee, Assistant Professor, BMI

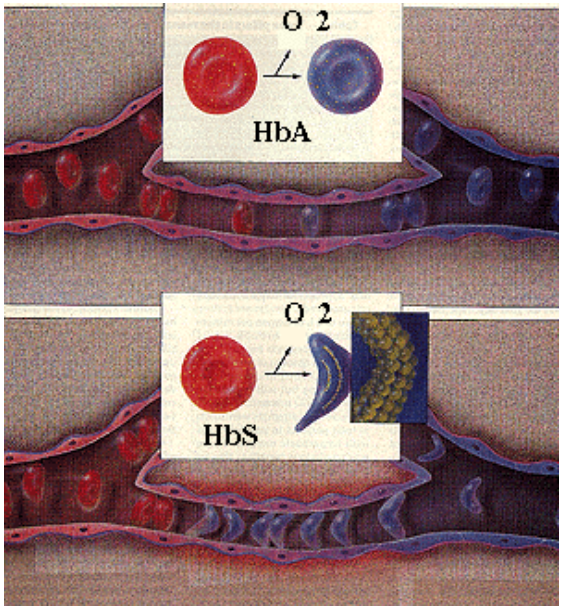
Much of our understanding of the basic structural features of proteins is due to the pioneering work of Linus Pauling (left) and Robert Corey (right).

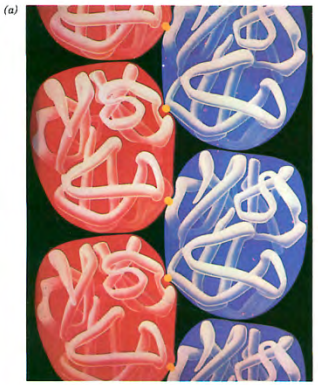
From clinical diagnosis to protein structure





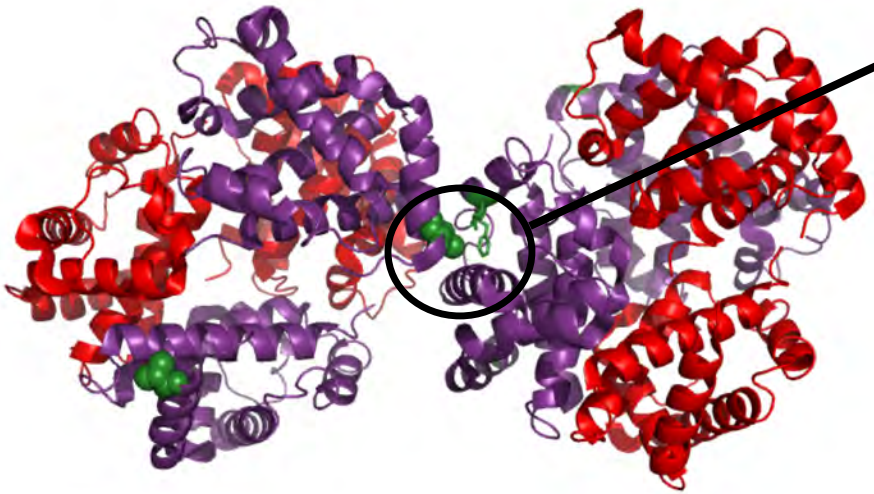
βGlu6 to βVal6 mutation

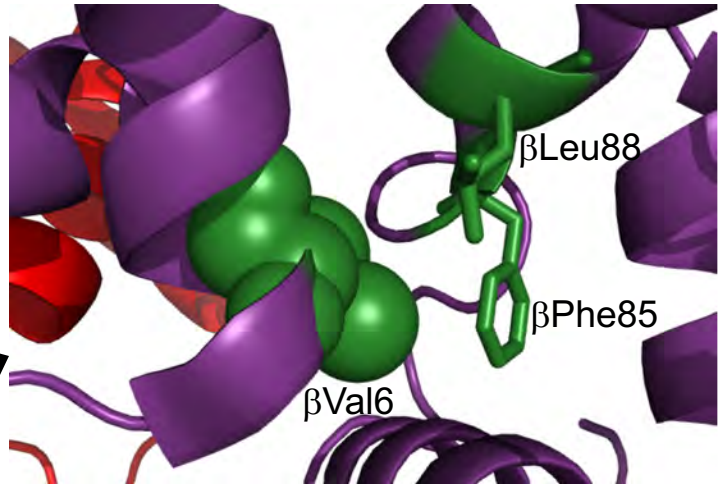




The 3D arrangement of amino acids dictates function!

Dimer of HbS molecules





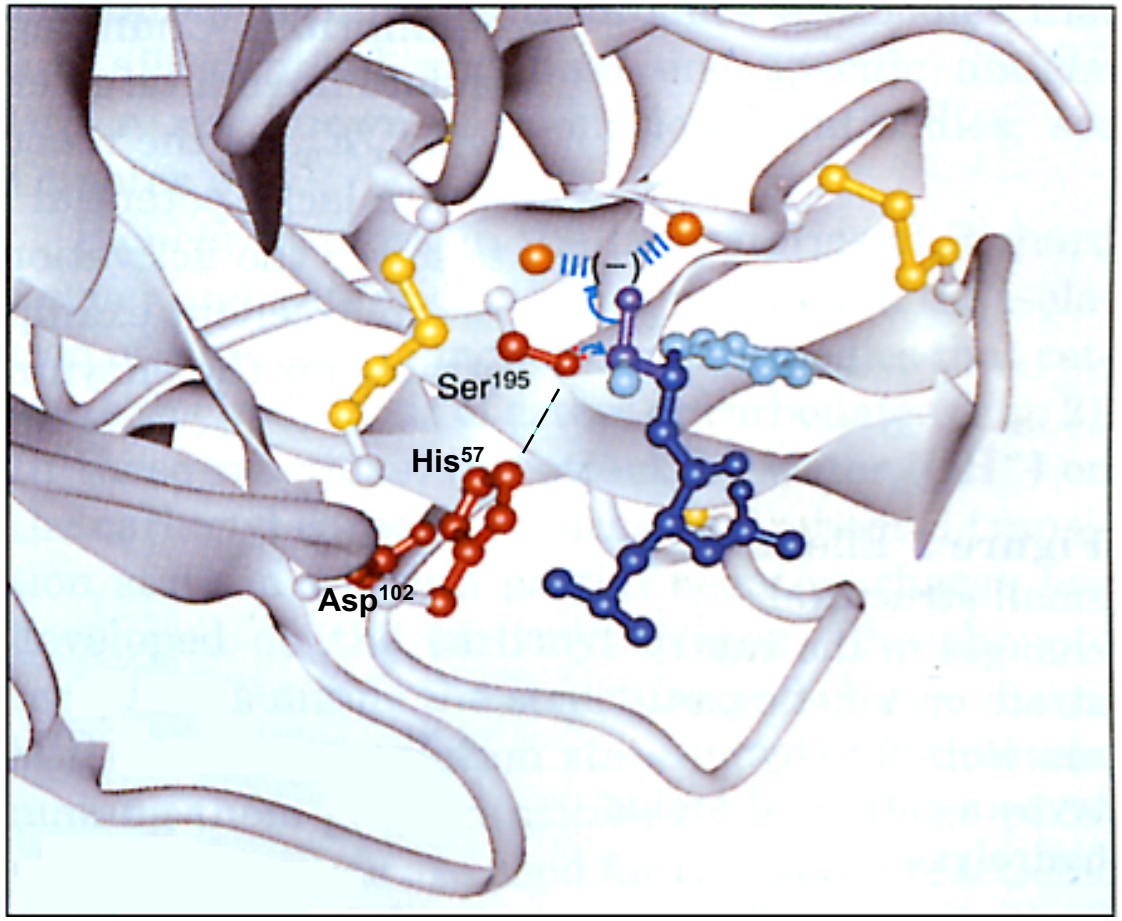
Close up view of the hydrophobic residues that interact to drive dimerization

The βE6V mutation in HbS creates a hydrophobic knob that fits into a hydrophobic pocket on an adjacent HbS tetramer leading to polymerization. ***The 3D structure dictates shape complementarity, but side chain chemistry drives the binding!***

The 3D arrangement of amino acids dictates function!

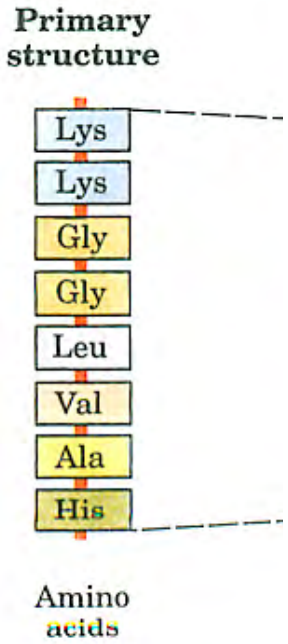
This diagram shows the active site of an enzyme, *trypsin*. The nucleophilic OH of Ser195 is positioned to attack the carbonyl group of a substrate (blue).

Asp102 and His57 hydrogen bond with Ser195, increasing its nucleophilicity.



The 3D structure orients the side chains of Ser195, Asp102, and His57 perfectly in 3D to create molecular scissors – ***i.e. side chain chemistry in 3D dictates function!***

Proteins are linear sequences of amino acids in 3-D

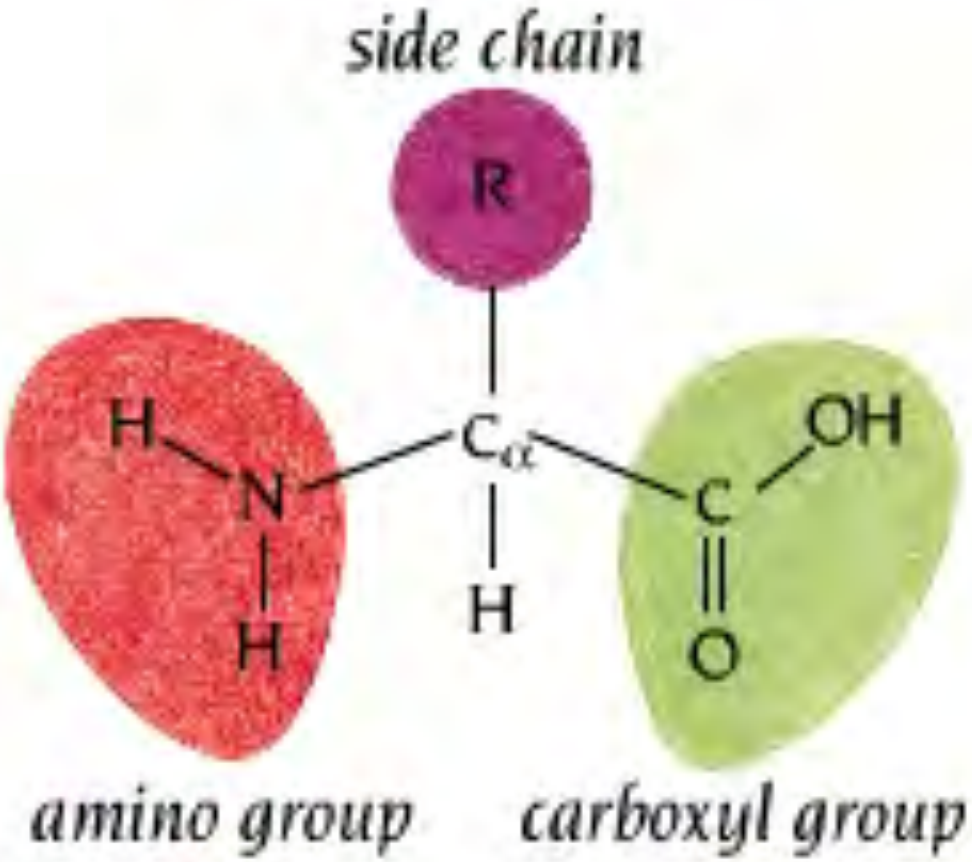


Amino acids are the building blocks of proteins. To understand structure and function, we must understand the chemistry of these amino acid building blocks:

1) Side chain chemistry

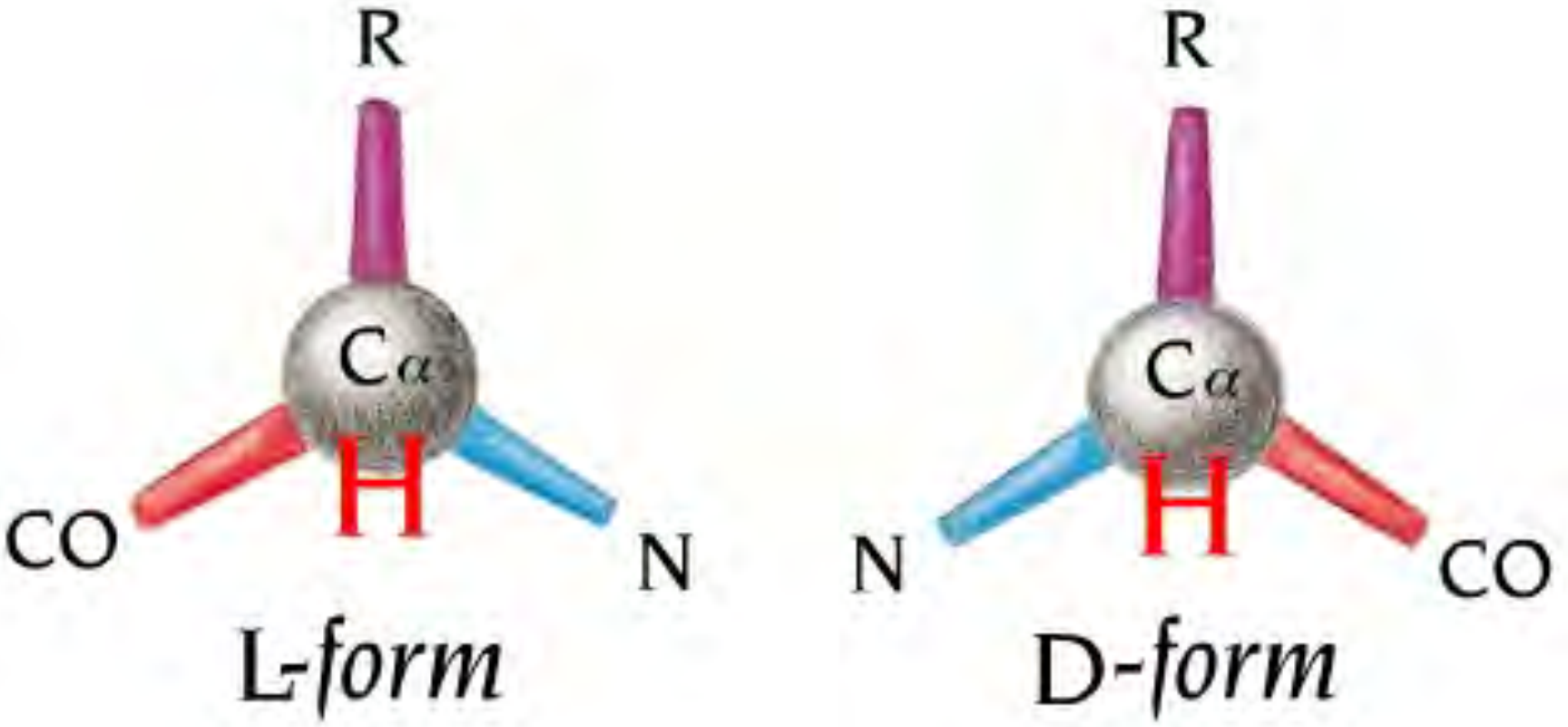
2) The peptide bond

α-Amino acids



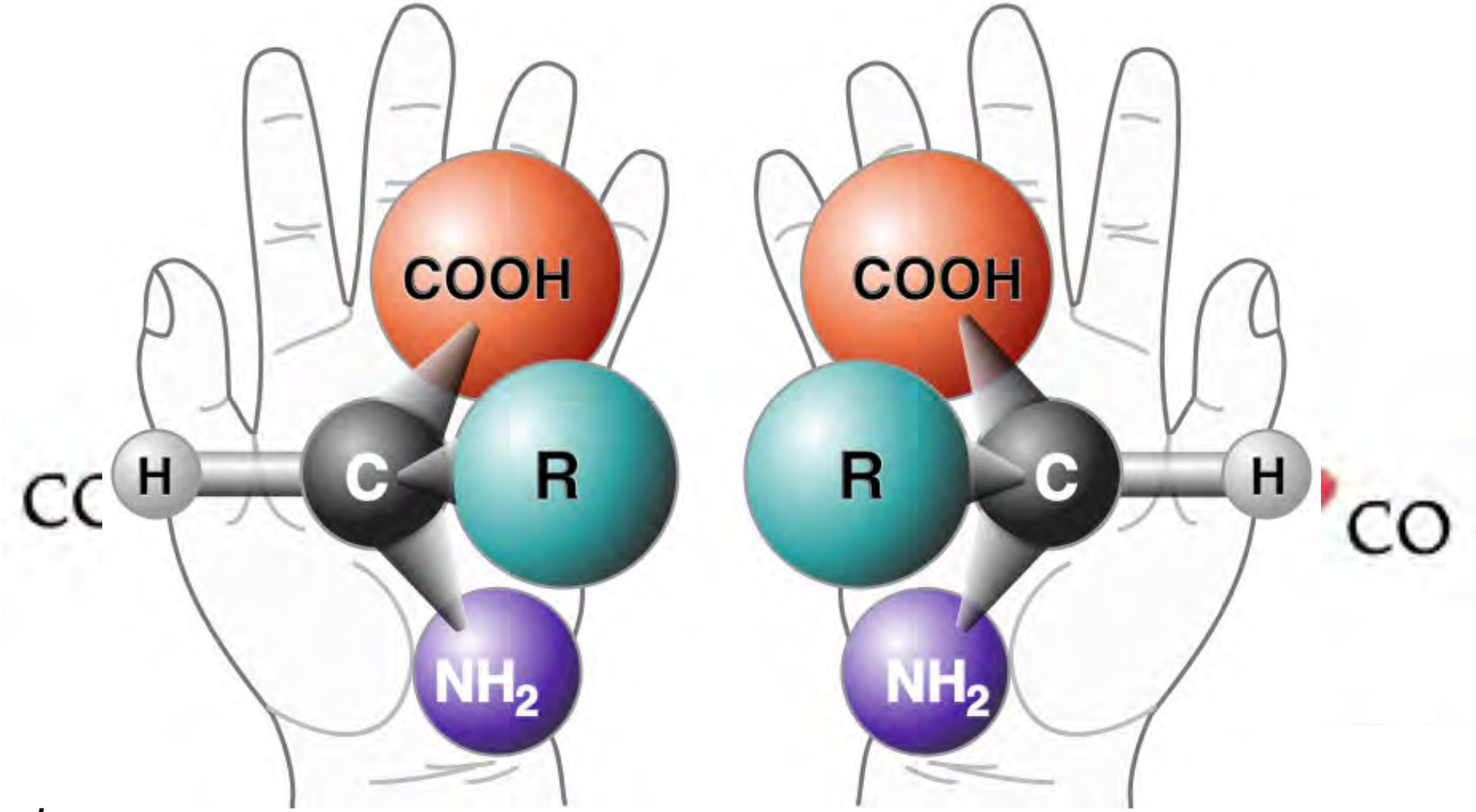
Proteins are linear polymers formed from the 20 naturally occurring α-amino acids

Amino acids adopt either L or D configuration (chirality)



L = levorotary, rotates plane polarize light counterclockwise
D = dextrorotary, rotates plane polarize light clockwise

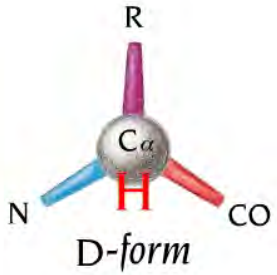
Amino acids adopt either L or D configuration (chirality)



L = levorotary, rotates plane polarize light counterclockwise

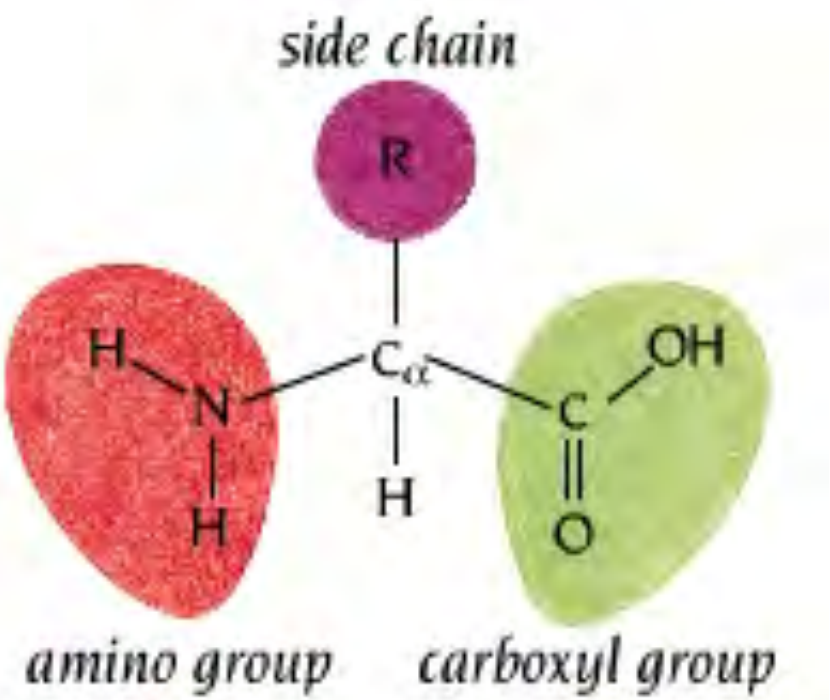
D = dextrorotary, rotates plane polarize light clockwise

L amino acids dominate on earth



All amino acids in proteins are *L. Bacteria enzymatically (i.e. non-ribosomally) synthesize peptides with D-amino acids for their cell walls and as antibiotics.*

Genetic code specifies 20 unique side chains

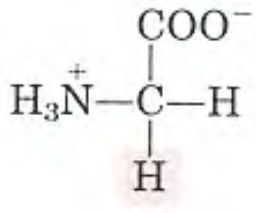


The chemistry of the R groups (R= Glycine, Alanine, Valine, Leucine, Isoleucine, …) dictate protein function. There are four types of side chains:

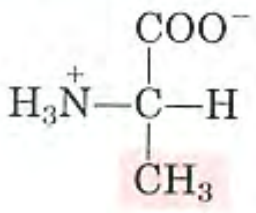
1) Aliphatic
2) Aromatic
3) Neutral with H bonding
4) Charged and “highly” polar

Aliphatic amino acids

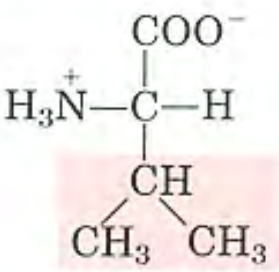
The R groups of the aliphatic amino acids are generally considered to be NON-POLAR as none has hydrogen bond forming capabilities.



Glycine

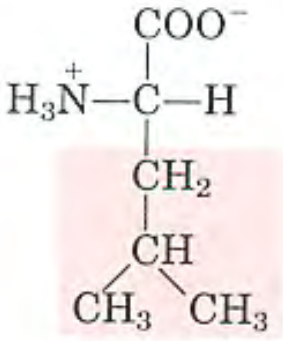


Alanine

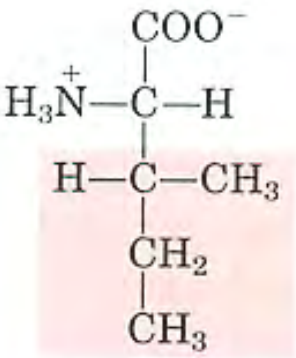


Valine

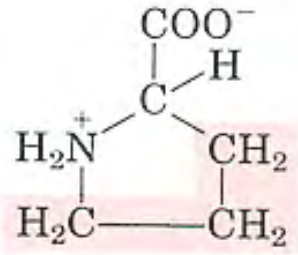
Glycine is generally considered to be a polar side chain because its physico-chemical properties are dominated by the peptide NH and C=O. The small size of Gly allows greater flexibility of the polypeptide backbone at this location



Leucine

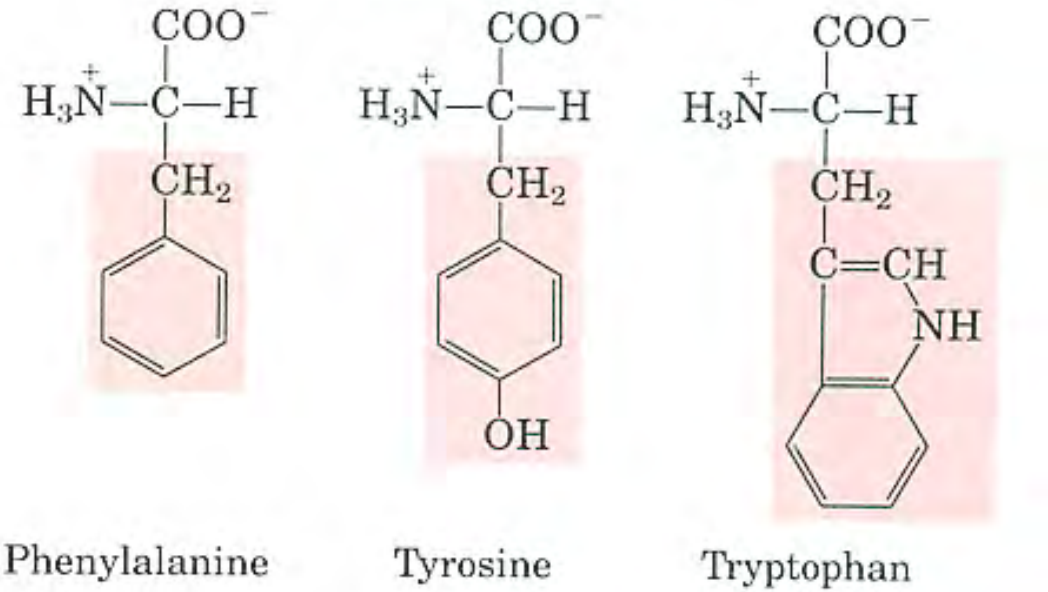


Isoleucine



Proline

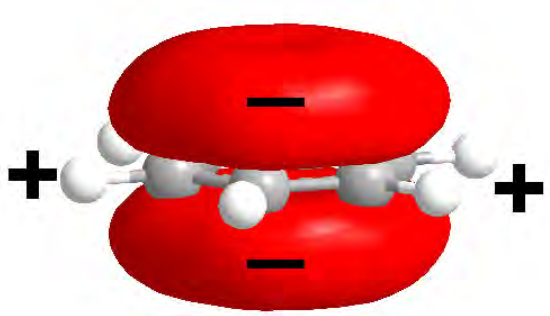
Aromatic amino acids



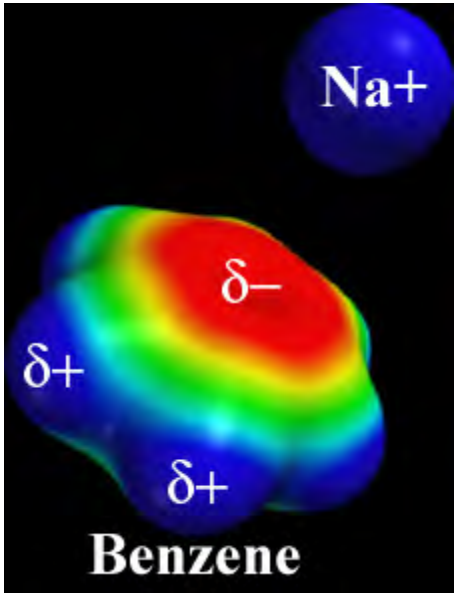
The R groups of Phe and Trp are generally considered non-polar due to their aromatic moieties. The Trp N-H can form a H bond – so it is a borderline member.

The R group of Tyr is moderately polar due to the H bonding capabilities of the OH group. If the Tyr deprotonates to form a tyronsinate, the residue becomes polar.

Aromatic amino acids



The quadrupole moment of benzene



The π electrons of benzene interact favorably with a sodium ion (right)

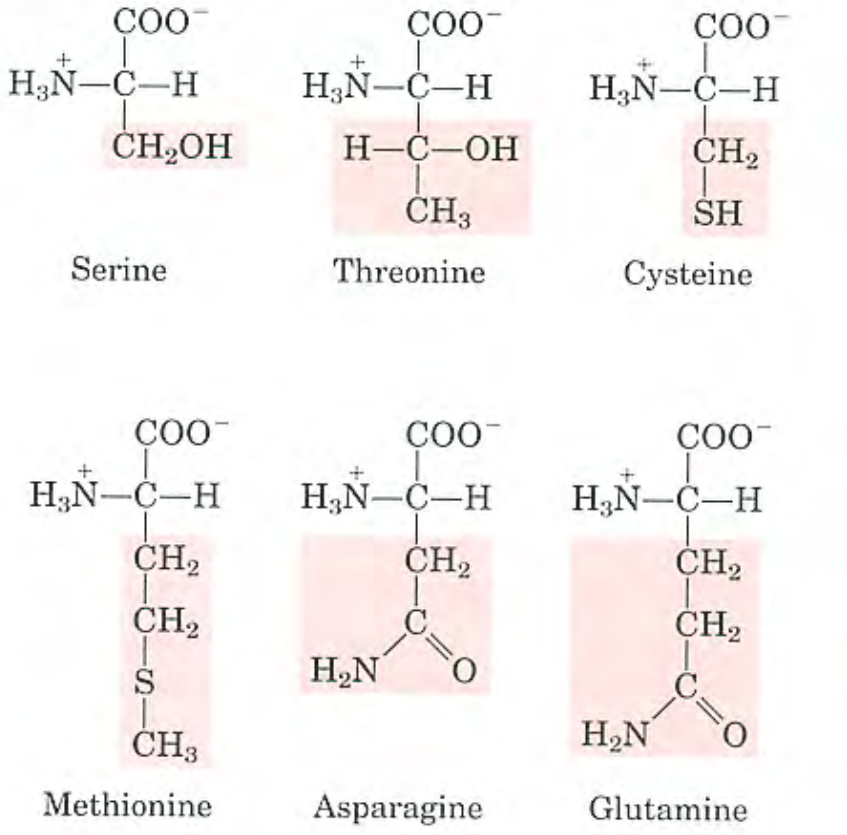
Although “non-polar”, the distribution of electrons in each aromatic leads to a quadrupole moment, which allows aromatic residues to form electrostatic interactions with other residues or even ions.

Images from Wikipedia and provided by Denis Dougherty

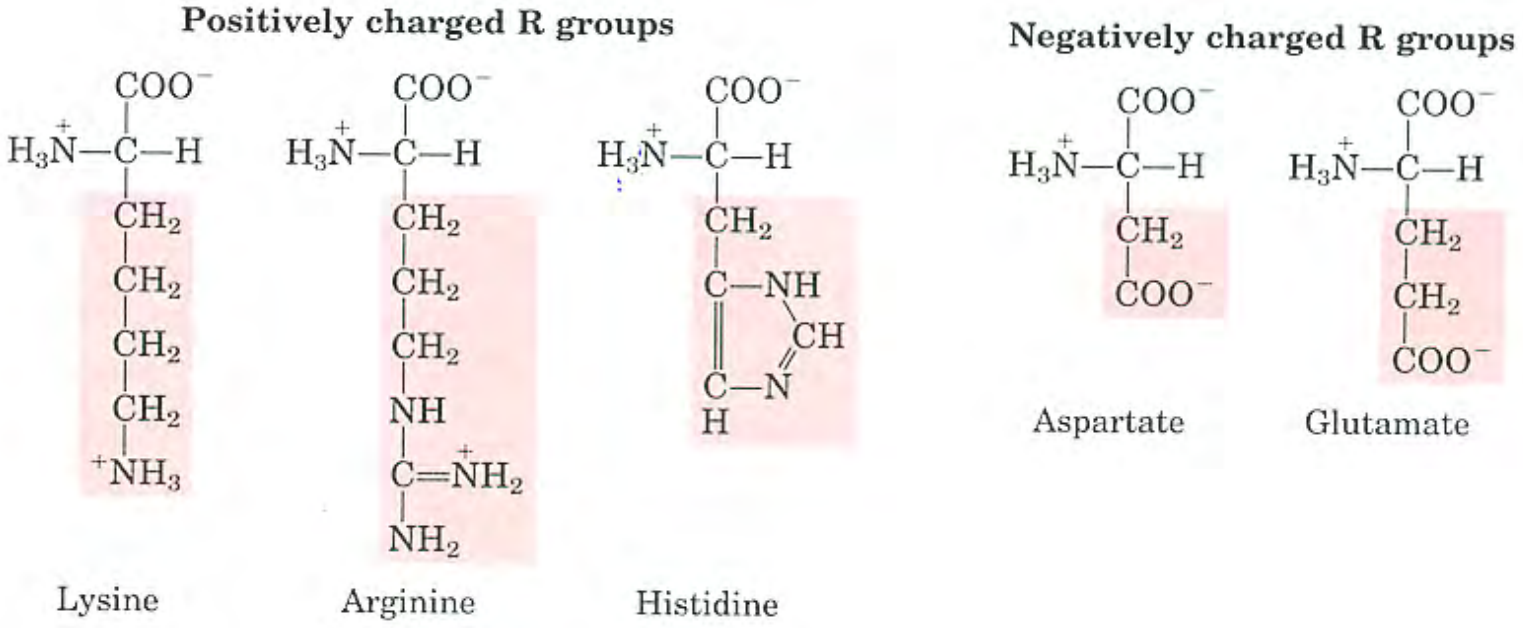
Neutral amino acids with hydrogen bonding

Met is the only member of this group that cannot readily form hydrogen bonds (it can be a weak H-bond acceptor). It is considered non-polar.

The remaining amino acid side chains have strong hydrogen bonding capabilities. Serine, threonine, and cysteine can deprotonate at high pH. In oxidizing environments (outside the cell), Cys forms disulphide bonds.



Charged amino acids



The charged R groups are considered to be the most polar of all side chains. The pKa of His is close to 7, so it can be found in either a neutral or a charged state in proteins. Changes in protonation state leading to the movement of protons from one residue to another within a protein are often essential to the mechanisms of protein function.

Properties of the 20 natural amino acids

Amino acid	Abbreviated names[1]		M_r[3]	pKa (R group) [1,2]	Polarity[1]	% Occurance in proteins
Nonpolar, aliphatic						
Glycine	Gly	G	75		Moderate	7.5
Alanine	Ala	A	89		Non-polar	9.0
Valine	Val	V	117		Non-polar	6.9
Leucine	Leu	L	131		Non-polar	7.5
Isoleucine	Ile	I	131		Non-polar	4.6
Proline	Pro	P	115		Moderate	4.6
Aromatic						
Phenylalanine	Phe	F	165		Non-polar	3.5
Tyrosine	Tyr	Y	181	10.5	Moderate	3.5
Tryptophan	Trp	W	204		Non-polar	1.1
Neutral, H-bonding						
Serine	Ser	S	105	13.6	Moderate	7.1
Threonine	Thr	T	119	13.6	Moderate	6.0
Cysteine	Cys	C	121	8.4	Non-polar	2.8
Methionine	Met	M	149		Non-polar	1.7
Asparagine	Asn	N	132		Polar	4.4
Glutamine	Gln	Q	146		Polar	3.9
Charged						
Aspartate	Asp	D	133	3.9	Very Polar	5.5
Glutamate	Glu	E	147	4.1	Very Polar	6.2
Lysine	Lys	K	146	10.5	Very Polar	7.0
Arginine	Arg	R	174	12.5	Very Polar	4.7
Histidine	His	H	155	6.0	Polar	2.1

[1]must know information in dark green including [2]pKa of N-termus ($-NH_3^+$) ≅ 9 and pKa of C-terminus (-COOH) ≅ 2; [3]the average molecular weight for an amino acid is 110 g/mol.

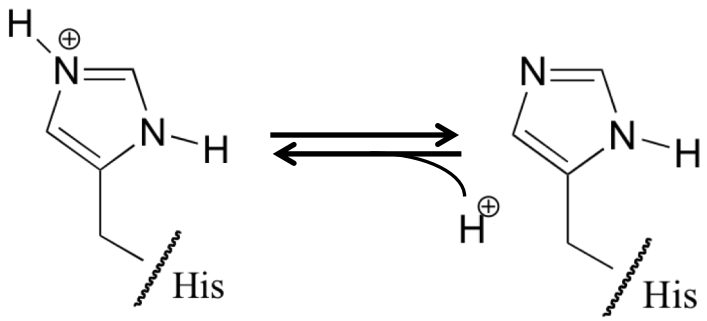
For amino acid quiz, you must know:

- one and three letter codes
- chemical structures
- side chain pK_a (**not in the quiz**)
- relative polarity (non-polar, moderately polar, polar, very polar)
  (not in the quiz)

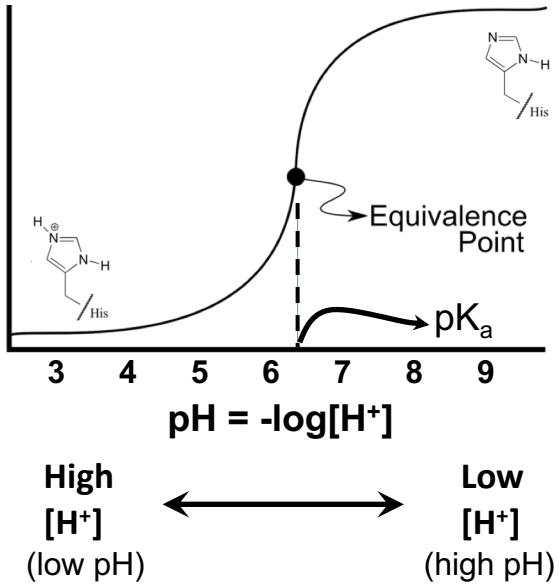
Note that pKa values tell us about side chain "chemistry" and are important parameters that often dictate function. Side chain polarity can also be used as a predictive tool to help predict protein structure, identify transmembrane segments, etc.

Amino acid side chain pK_a

Consider Histidine:



low pH high pH

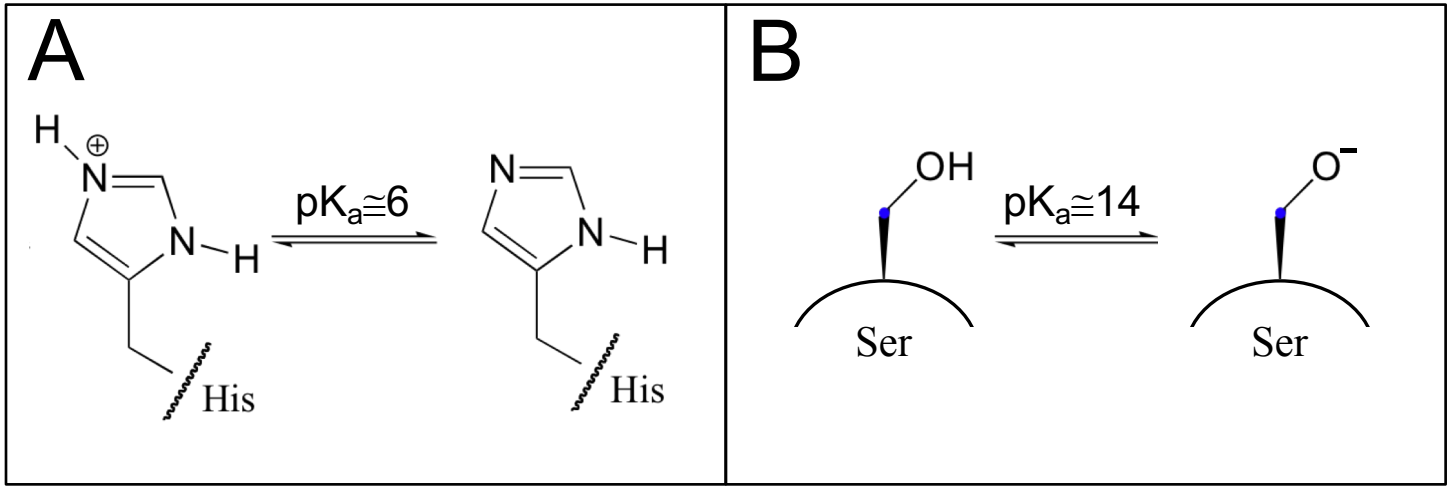


Histidine exists in both a protonated positively charged and a deprotonated neutral form. At pH values << 6, there is a high $[H^+]$ that favors the protonated state. If we titrate by adding OH^-, we raise the pH and lower the $[H^+]$ so that we now favor the de-protonated or neutral form.

The pK_a is the pH at which the side chain is 50% protonated and 50% deprotonated and is thus a measure of how strongly the side chain holds onto its proton.

His deprotonates at higher [H⁺] than Ser

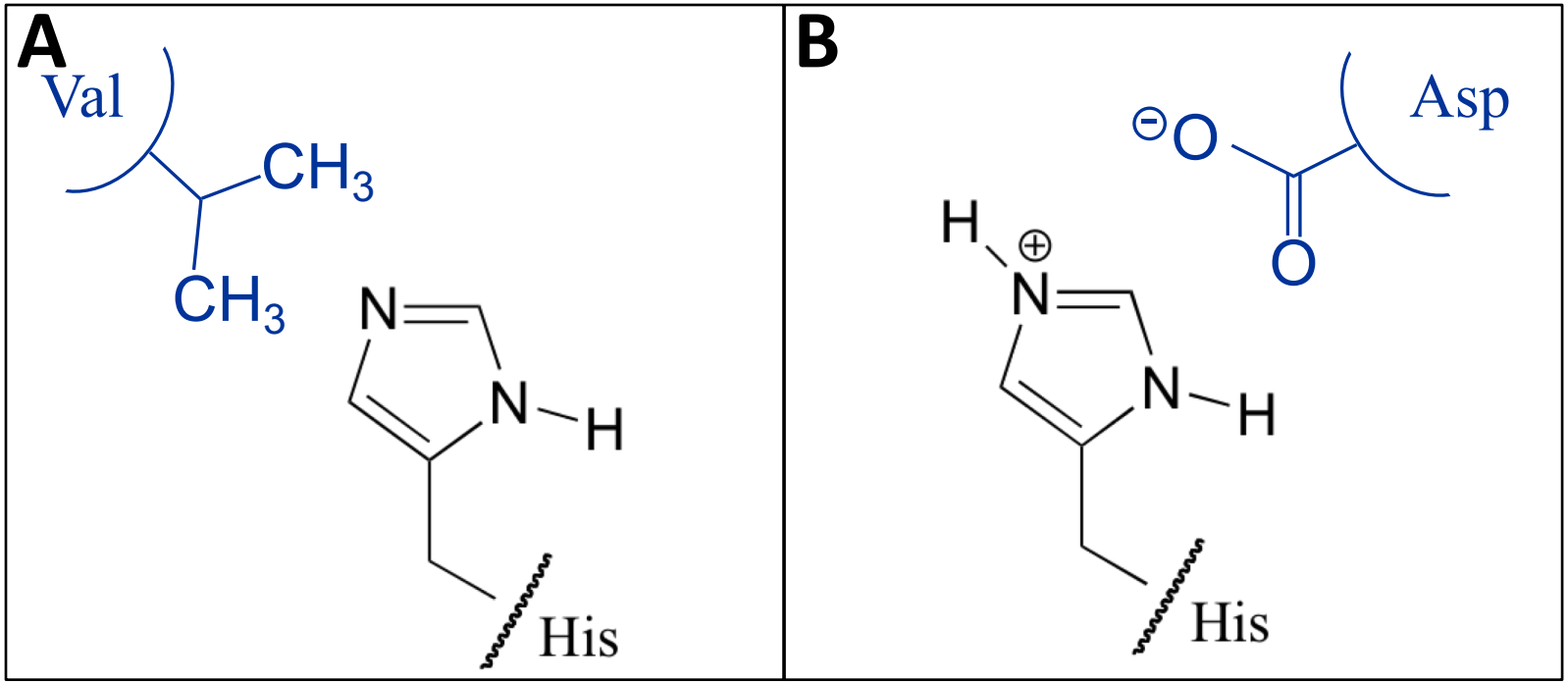
The pKa is a measure of the relative ease by which the side chain gives up its proton. Which has a greater propensity to deprotonate?



At what concentration of protons will His give up 50% of its protons? At what concentration of protons will Ser give up 50% of its protons?

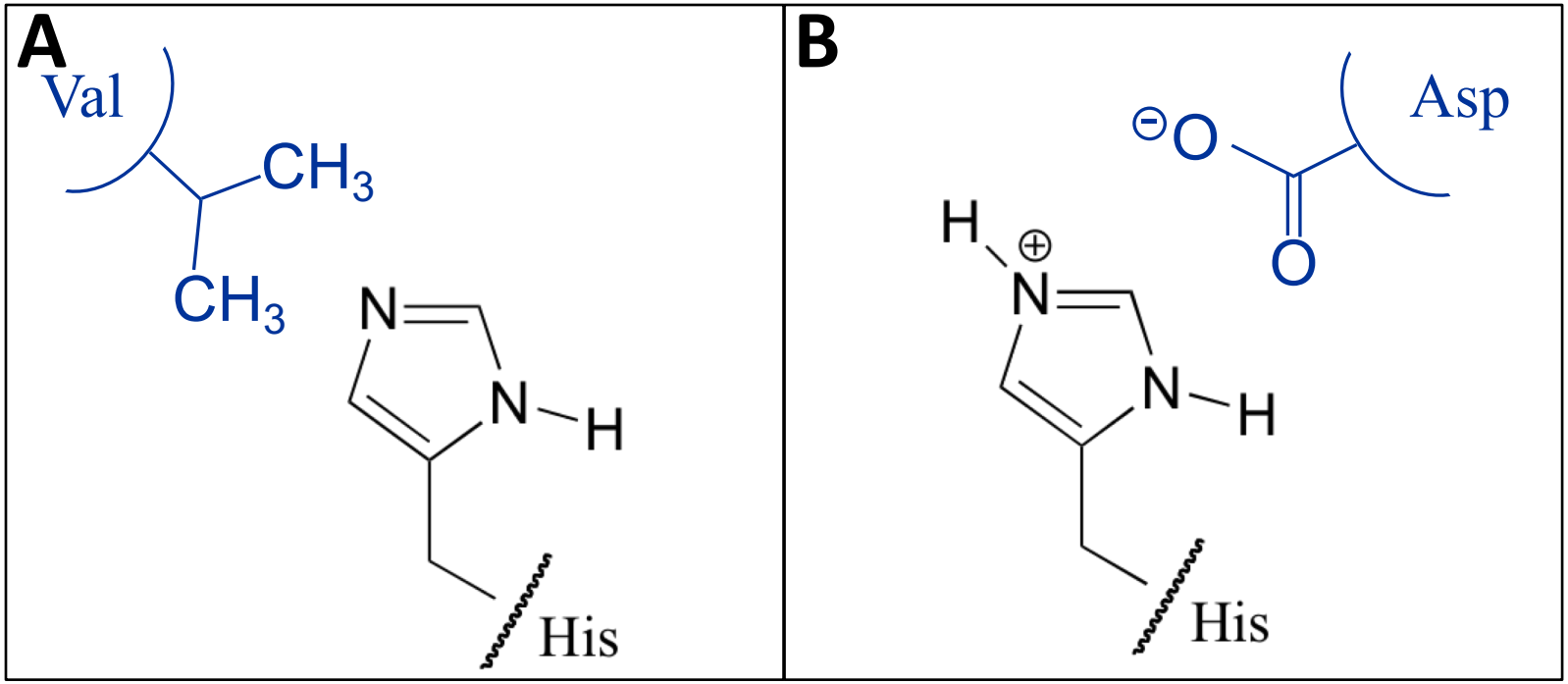
His with a moderately low pK_a ($pK_a \cong 6$)(*A*) deprotonates much more easily than Ser ($pK_a \cong 14$)(*B*) – i.e. a higher pH (↓[H^+]) is required for Ser to deprotonate.

pK_a is environment dependent



Consider the pKa of His surrounded by different residues on the interior of a protein. On the left, we have a His that is close to a hydrophobic Val. On the right, we have a His that is close to a charged Asp residue. How will these adjacent residues influence the pKa of the His residues?

pK_a is environment dependent

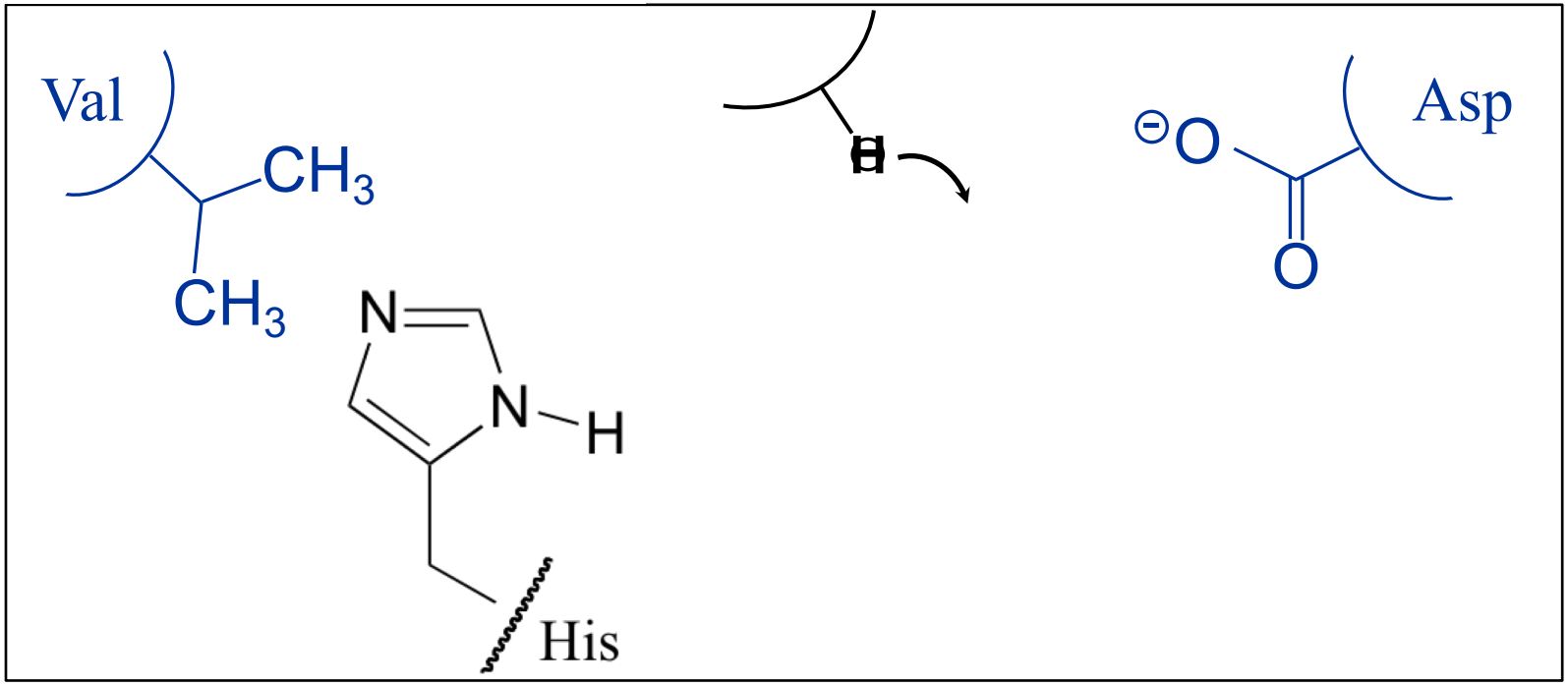


The oppositely charged Asp stabilizes the protonated form of His (B, right) and shifts the pKa to a higher pH value) – His holds on stronger to the proton.

So there is a greater chance that His will be protonated next to an Asp than next to a neutral Val.

Changes in protein structure change pKas

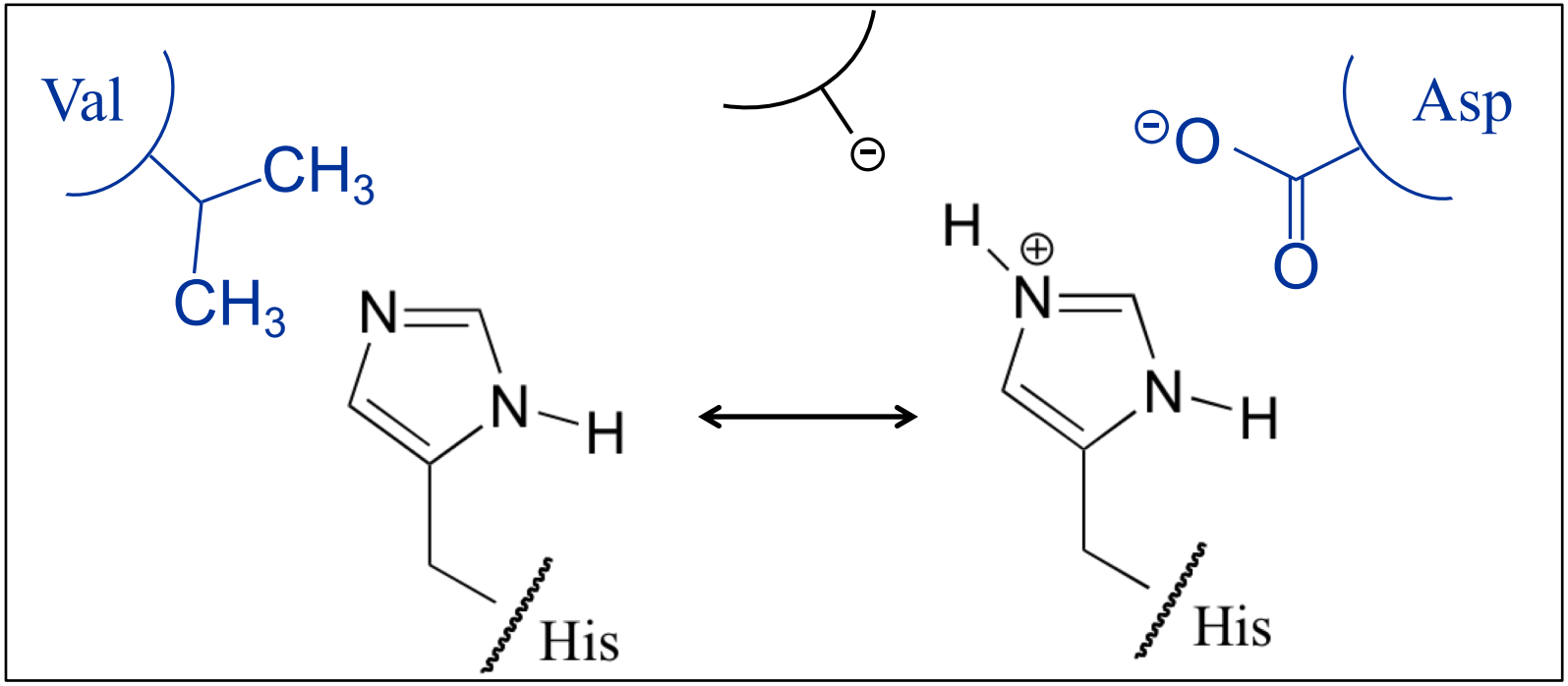
Conformation B



Consider a protein that changes conformation moving the His from a position close to a Val to a position close to an Asp:

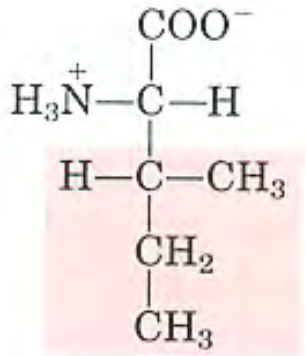
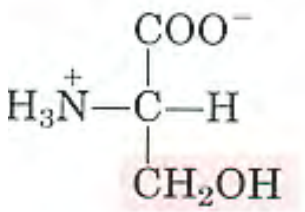
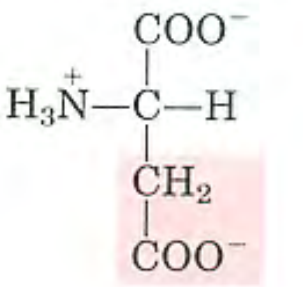
In conformation B, His is close to an Asp and will prefer to be protonated thus shifting the pKa higher – His will likely be protonated at neutral pH=7.

Changes in protein structure change pKas



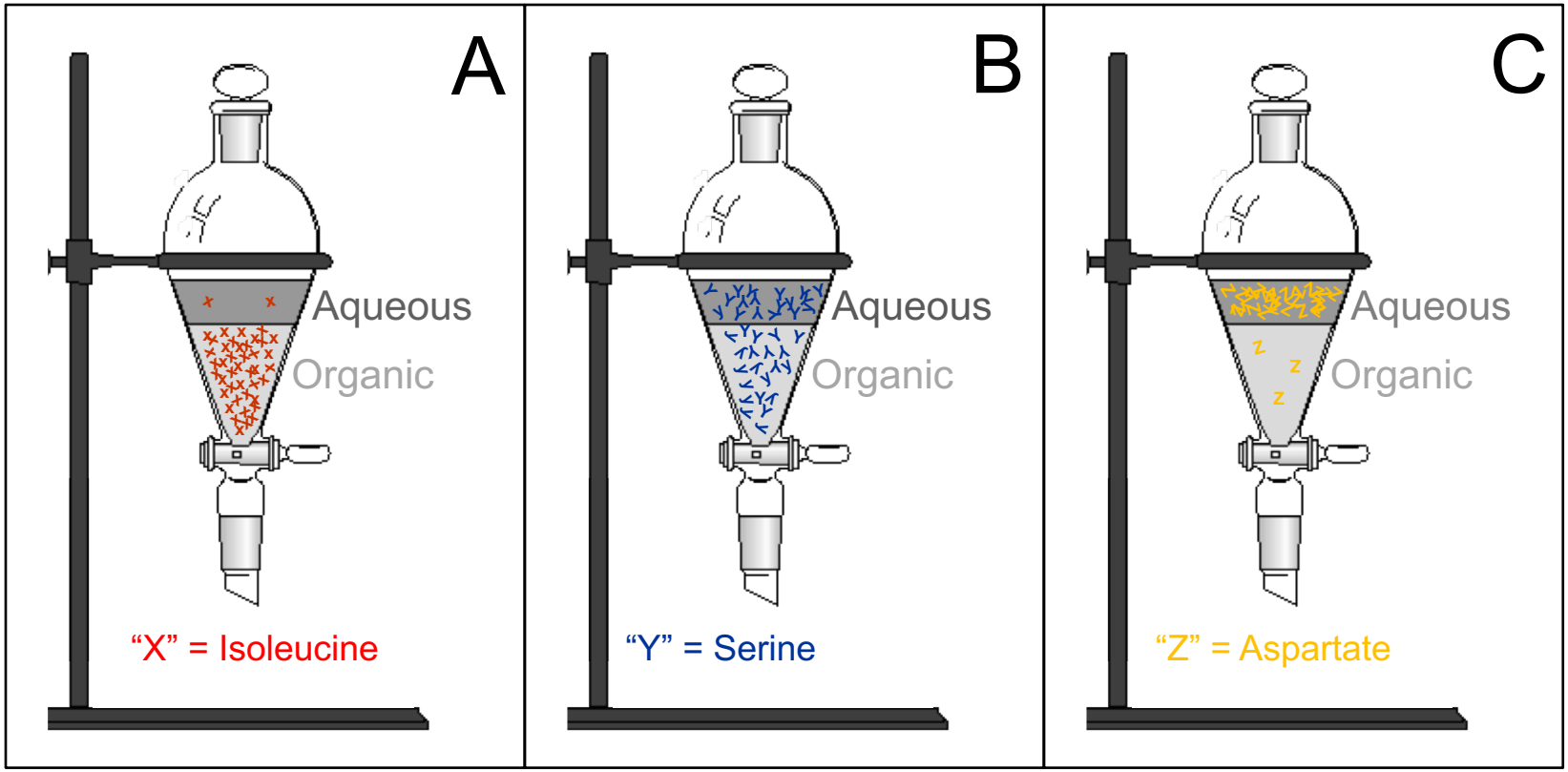
By changing the surrounding environment, the pKa of a residue can be changed so that the residue either acquires or donates a proton! Changes in structure leading to changes in pKa can cause the transfer of protons from one side chain to another. Such transfers are essential during protein function!

Amino acid side chain polarity

A	B	C
		
Isoleucine	Serine	Aspartate
Non-polar	Moderately polar	Polar

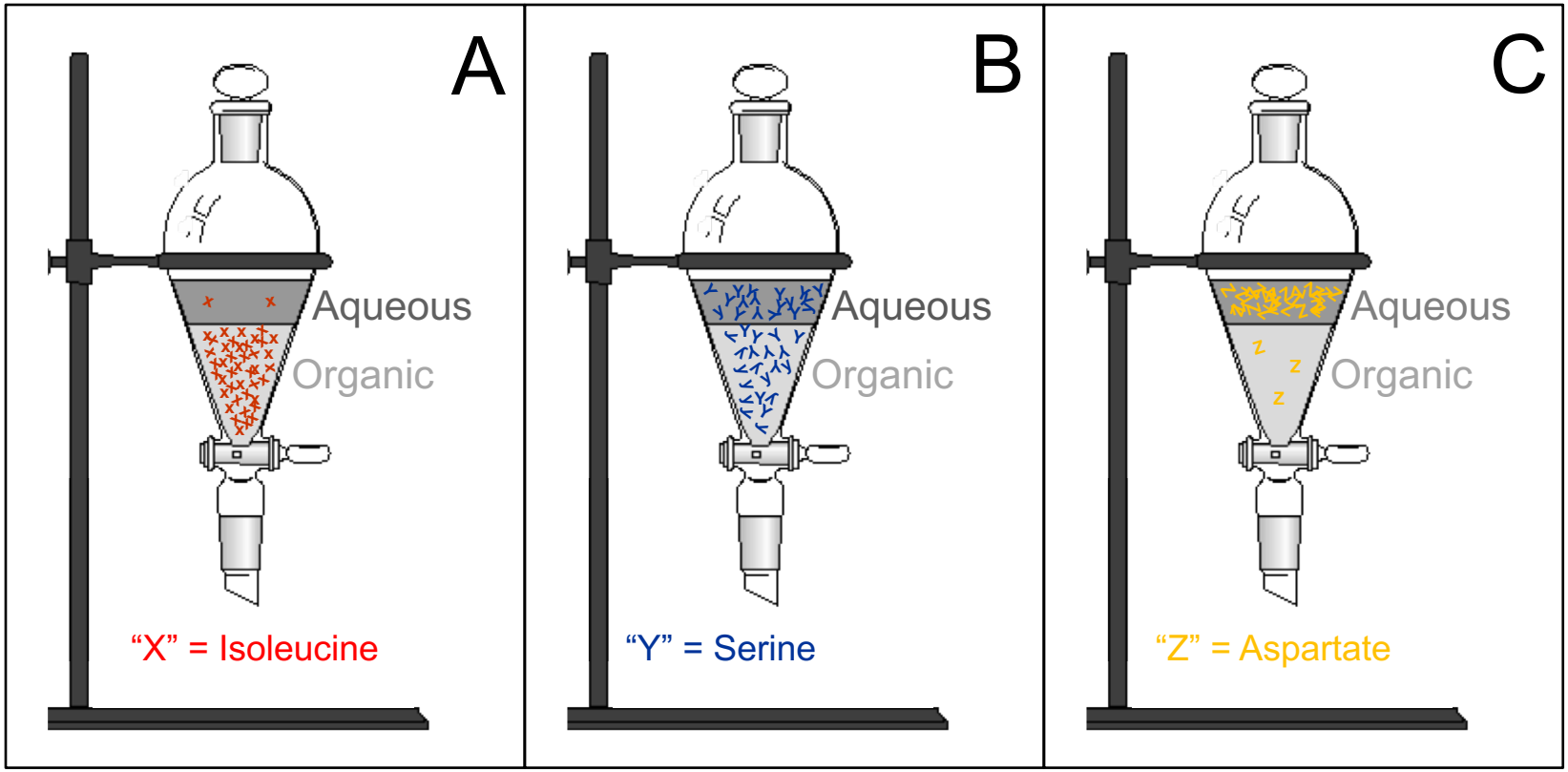
We can intuitively estimate polarity from the structures of side chains. For example, the aliphatic chain of isoleucine with four carbons is non-polar; serine with one carbon, but a polar hydrogen bonding OH group is moderately polar, while the charged aspartate is polar. *But how do we experimentally determine amino acid side chain polarity?*

Quantifying amino acid polarity



Several methods have been used to measure the relative polarities of the amino acid side chains, but in essence all are based on the relative partitioning of the amino acids between an aqueous and an organic phase. Many scales have been developed, the most useful at http://blanco.biomol.uci.edu/hydrophobicity_scales.html

Quantifying amino acid polarity



(A) Isoleucine has a greater affinity for the organic phase, serine has moderate affinity for both the organic and the aqueous phase, and aspartate has a greater affinity for the aqueous phase. By measuring the partition coefficients, we can establish a relative polarity scale

Amino acid polarity scales

Table 12.1 Hydrophobicity scales

Amino acid	Phe	Met	Ile	Leu	Val	Cys	Trp	Ala	Thr	Gly	Ser	Pro	Tyr	His	Gln	Asn	Glu	Lys	Asp	Arg
A	2.8	1.9	4.5	3.8	4.2	2.5	−0.9	1.8	−0.7	−0.4	−0.8	−1.6	−1.3	−3.2	−3.5	−3.5	−3.5	−3.9	−3.5	−4.5
B	3.7	3.4	3.1	2.8	2.6	2.0	1.9	1.6	1.2	1.0	0.6	−0.2	−0.7	−3.0	−4.1	−4.8	−8.2	−8.8	−9.2	−12.3

Row A is from J.Kyte and R.F. Doolittle; row B, from D.A. Engelman, T.A. Steitz, and A. Goldman.

Non-polar **Moderate Polarity** **Polar** **Very Polar**

A and B are two different hydrophobicity scales of the amino acids calculated using two different methods – they give slightly different relative polarities. For our use, you need to know whether the amino acids are non-polar, moderately polar, polar, or very polar, although these designations are somewhat arbitrary…

Amino acid polarity scales

Table 12.1 Hydrophobicity scales

Amino acid	Phe	Met	Ile	Leu	Val	Cys	Trp	Ala	Thr	Gly	Ser	Pro	Tyr	His	Gln	Asn	Glu	Lys	Asp	Arg
A	2.8	1.9	4.5	3.8	4.2	2.5	−0.9	1.8	−0.7	−0.4	−0.8	−1.6	−1.3	−3.2	−3.5	−3.5	−3.5	−3.9	−3.5	−4.5
B	3.7	3.4	3.1	2.8	2.6	2.0	1.9	1.6	1.2	1.0	0.6	−0.2	−0.7	−3.0	−4.1	−4.8	−8.2	−8.8	−9.2	−12.3

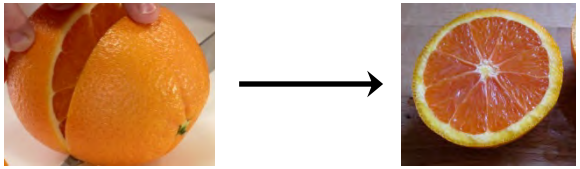
Row A is from J.Kyte and R.F. Doolittle; row B, from D.A. Engelman, T.A. Steitz, and A. Goldman.

Non-polar **Moderate Polarity** **Polar** **Very Polar**

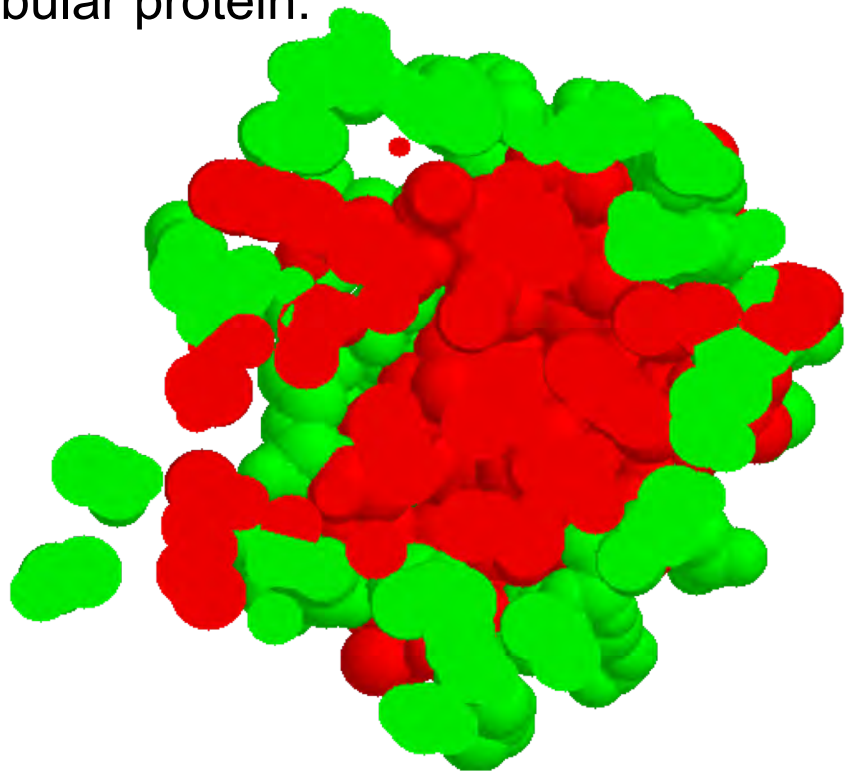
…for example, the polarities of Ala and Thr are close – yet we classify Ala as non-polar and Thr as moderately polar. There is no clear distinction between polar and non-polar. Instead, there is a continuum where some residues are more or less polar than others. You should understand why the above designations are somewhat arbitrary.

Hydrophobic residues are found inside proteins

Non-polar residues tend to be found on the interiors of globular proteins and at the interfaces between different subunits.



A slice through a space filling model of a globular protein:



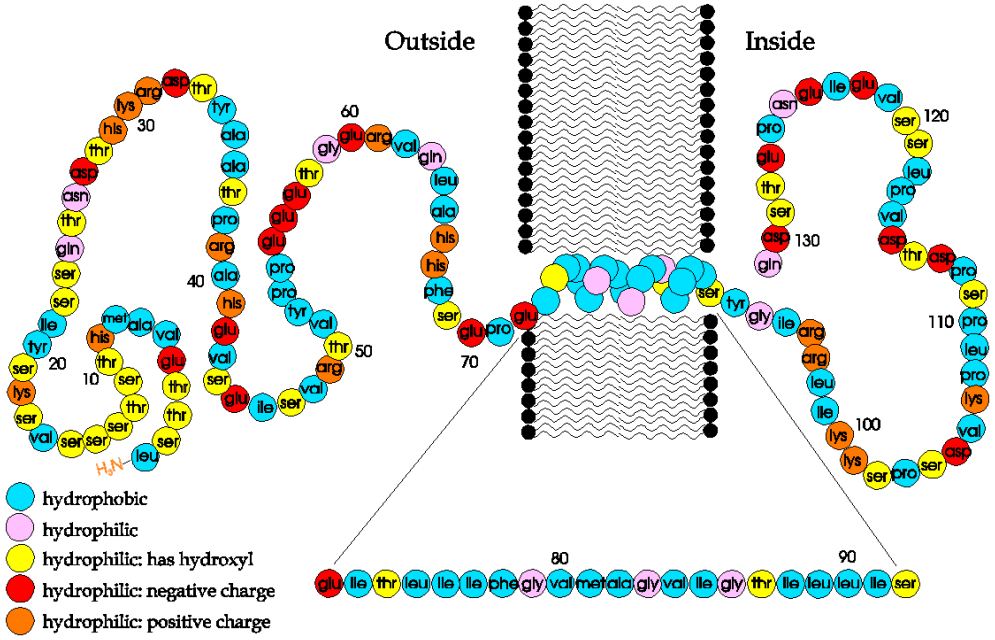
The hydrophobic residues (red) are clustered on the interior of a protein away from the aqueous environment. Polar hydrophilic residues (green) are typically found on the surface of the protein.

Hydrophobic residues are found in membranes

Non-polar residues are particularly common in α-helical transmembrane (TM) segments of membrane proteins.

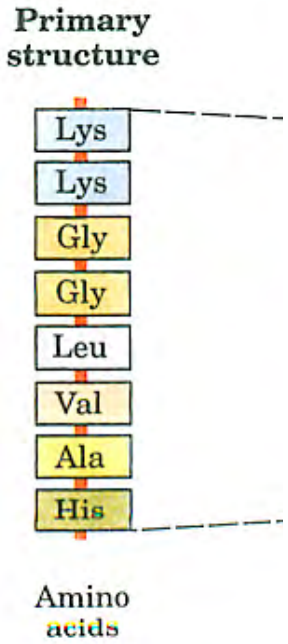
This strong correlation suggested that hydrophobicity can be used as a tool to predict the location of transmembrane α-helices in a protein sequence?

The TM segment of Glycophorin



The transmembrane (TM) portion of glycophorin (a red blood cell protein) has many non-polar residues, while the extra-membranous domains have many hydrophilic residues

Proteins are linear sequences of amino acids in 3-D



Amino acids are the building blocks of proteins. To understand structure and function, we must understand the chemistry of these amino acid building blocks:

1) Side chain chemistry

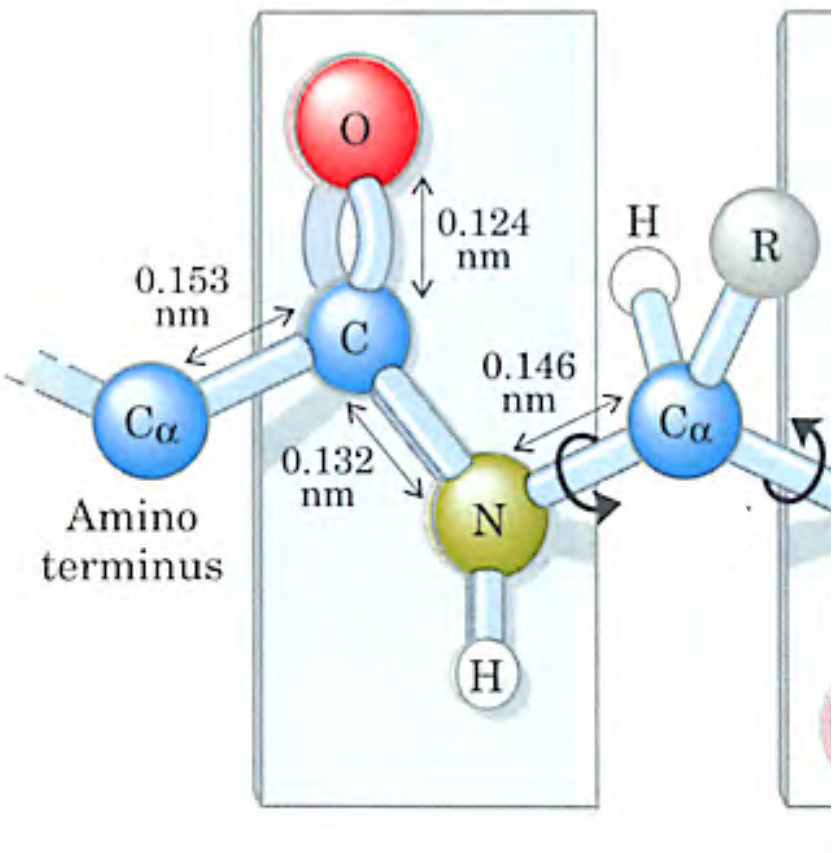
2) The peptide bond

Crystal structures (structural biology) define peptide bond geometries

In the 1930s and 1940s, Linus Pauling and Robert Corey solved the structures of many di- and tri-peptides and thus defined the peptide bond geometries and bond lengths, as shown on the right.

Note that we usually use the unit Å, when discussing bond lengths.

1Å = 0.1 nm



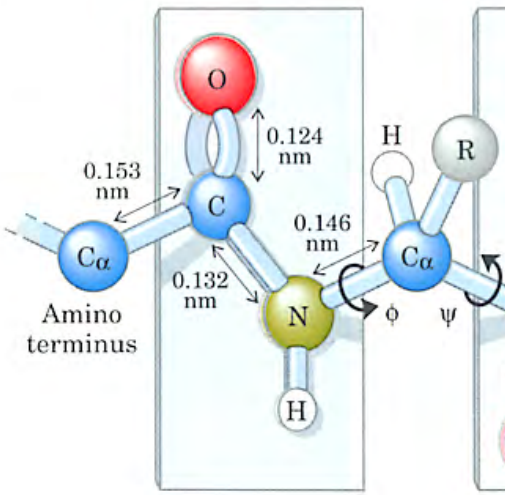
Peptide bond is shorter than single C-N bond...

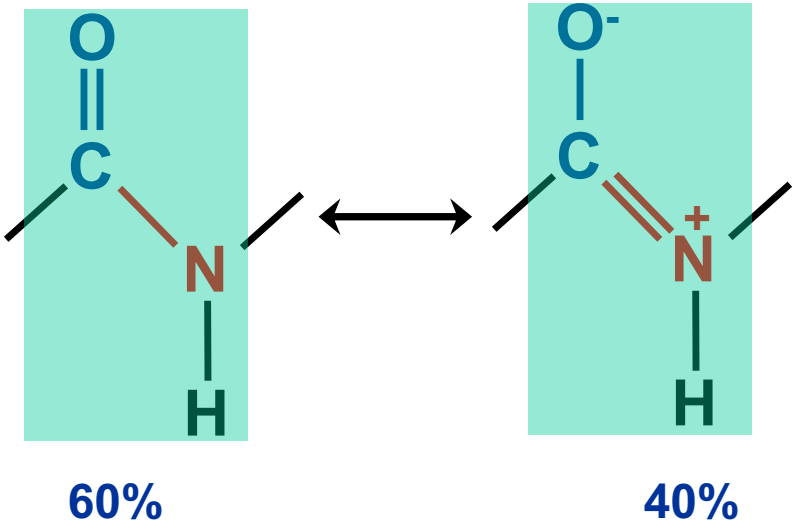
1) The **C=O** bond is 1.24 Å, 0.02 Å longer than a double C=O bond (~1.22 Å), but shorter than a single C-O bond (~1.27 Å)

2) The **CO-N** peptide bond is 1.32 Å, 0.13 Å shorter than a single C-N bond (~1.45 Å), but longer than a double C=N bond (~1.13 Å)

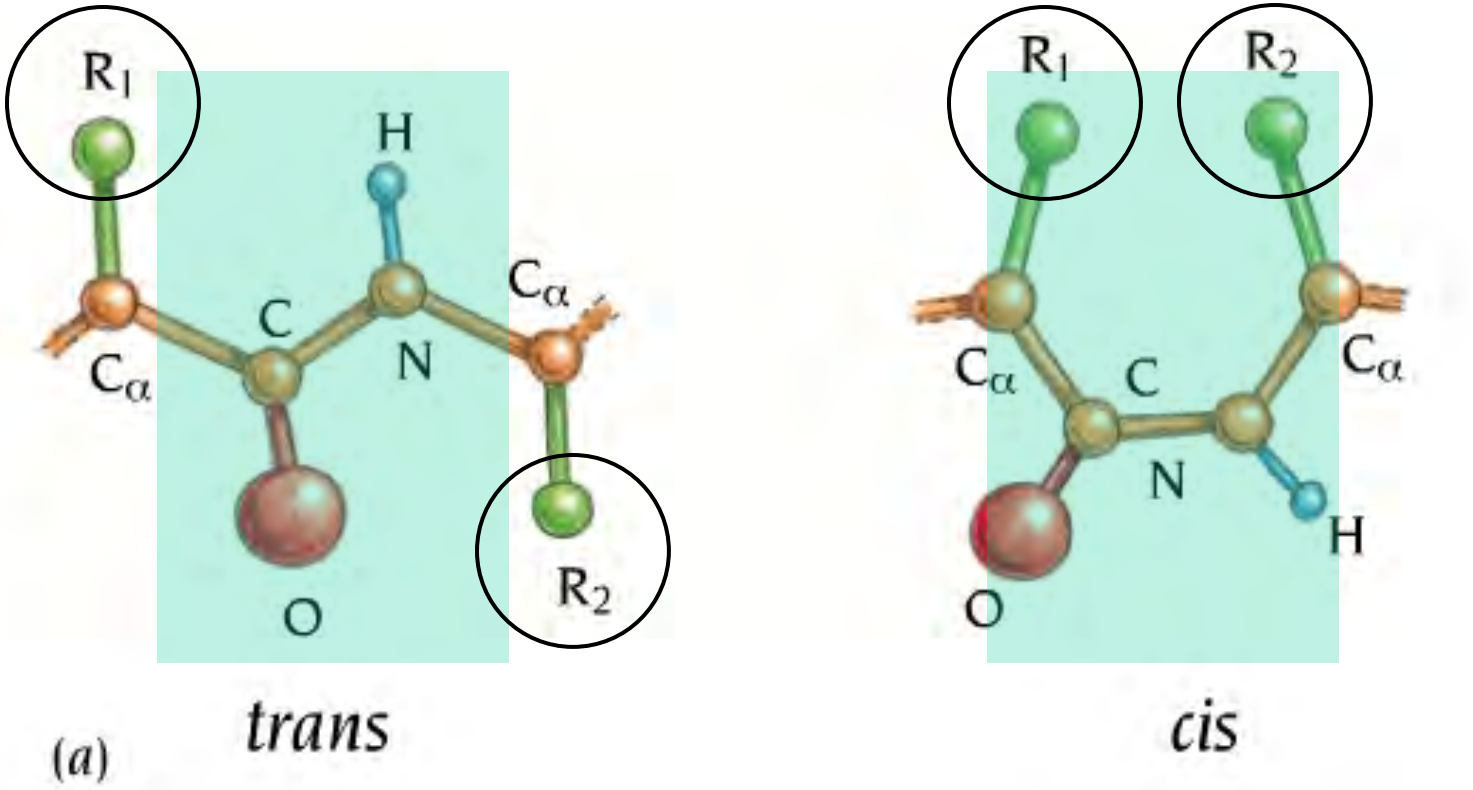
These both suggest resonance – or partial sharing of the π-electrons between the C=O and the C-N. The consequence of this resonance is that the C=N bond has a partial (~40%) double bond character.

The 40% double bond nature of the peptide bond means that there is no rotation about the (**C=O**) **–** (**N**-H) bond. The peptide bond is thus planar (green) i.e. the atoms associated with the C-N peptide bond are coplanar.



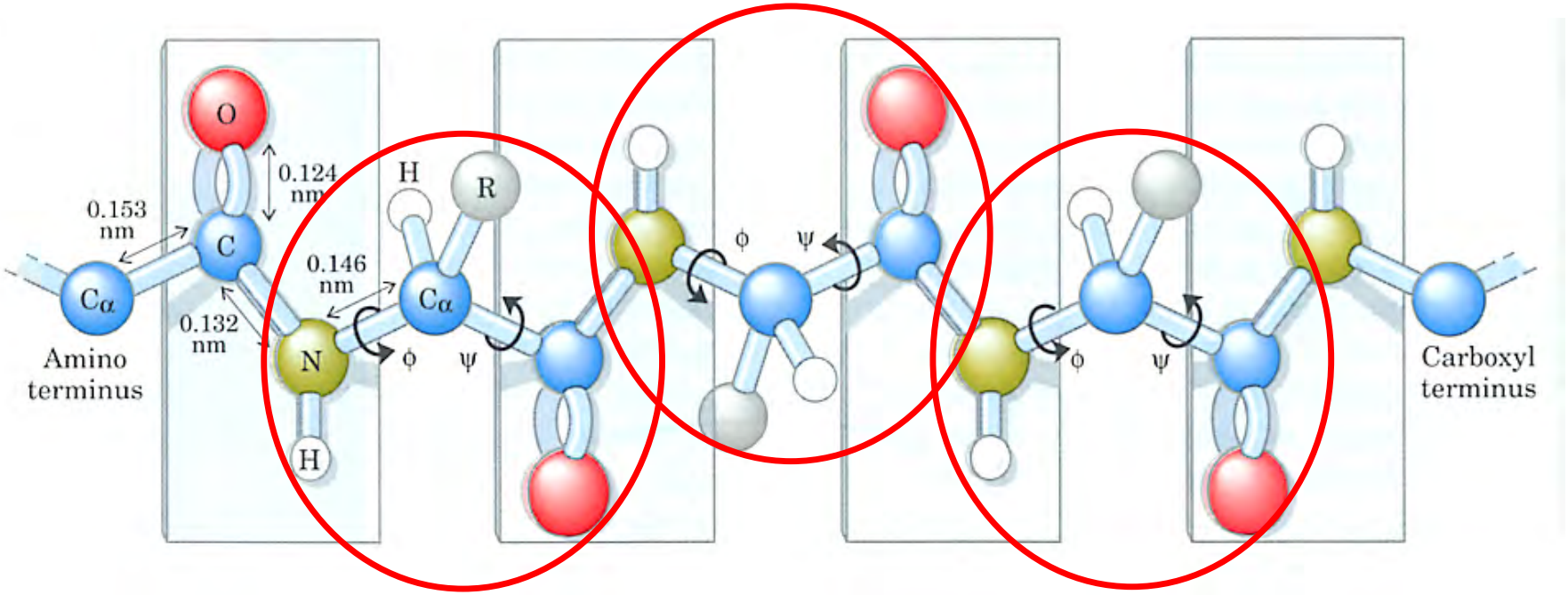


Peptide bond is usually trans



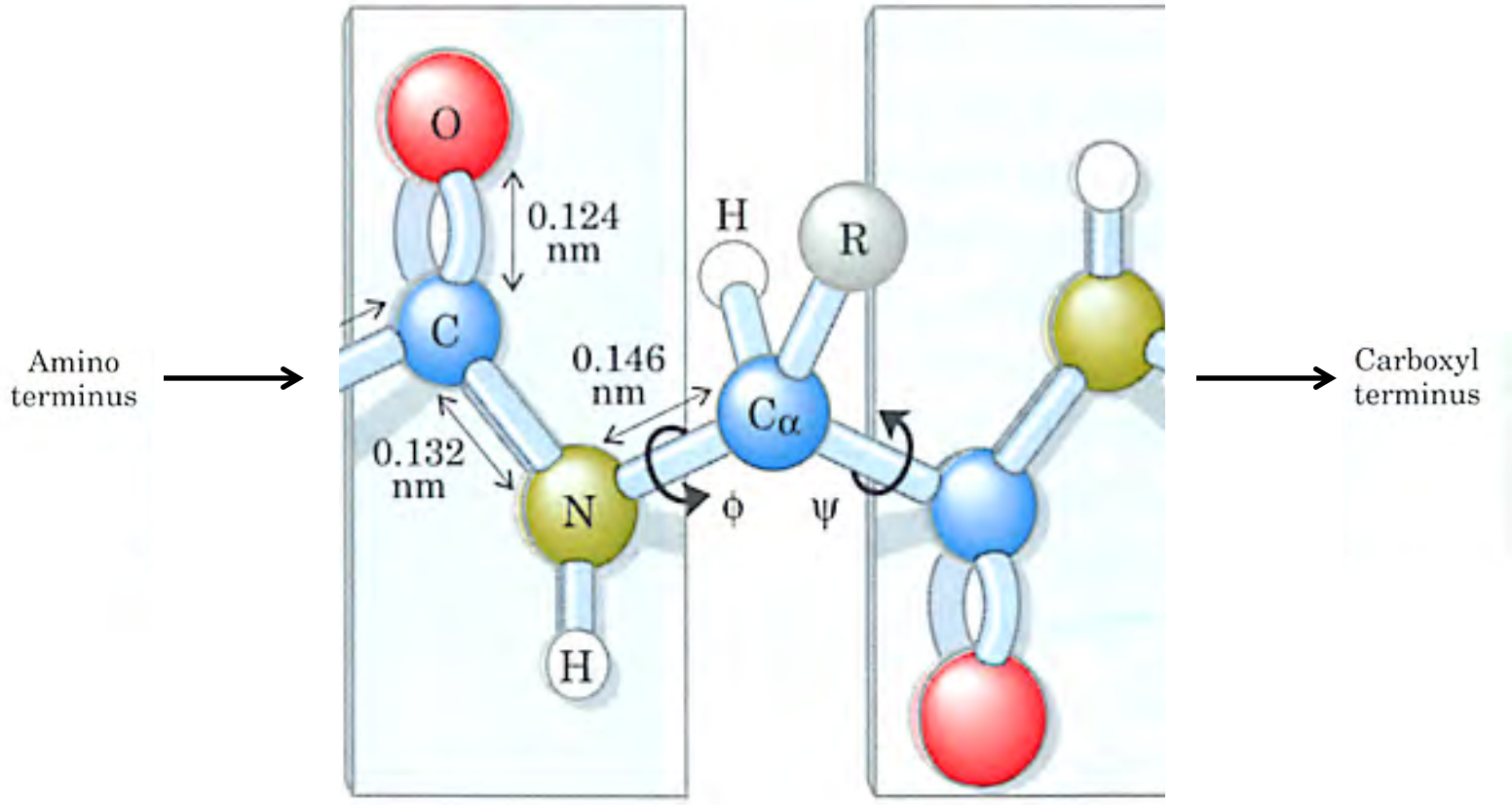
Pauling and Corey also found that the *trans* conformation (i.e the C-C_∞-R1 and the N-C_∞-R2 bonds are on opposite sides of the peptide bond) is found 99.9% of the time (except with Pro, where the *trans* form is favored “only” 95% of the time).

Polypeptide Chains are a series of planar units



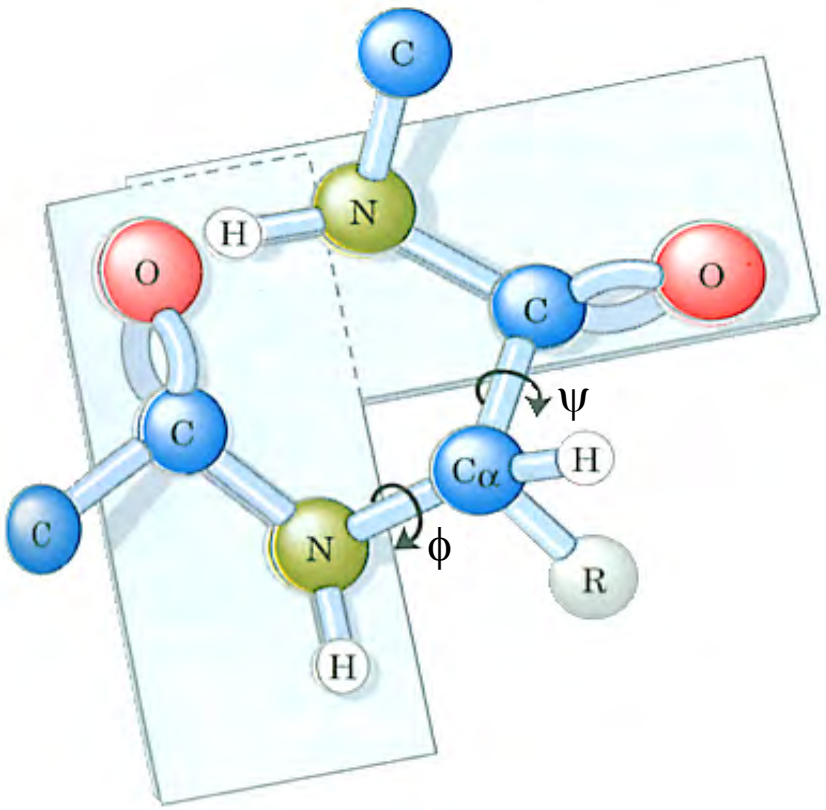
The backbone of the polypeptide chain is a series of planar structures, each plane separated by a methylene group, -C_αH(R)-. The only bond rotations allowed are rotations about the two single bonds attached to the C_α methylene group of each amino acid – called the ϕ and φ angles. These rotations, although limited, allow the polypeptide to fold into a globular or other structure. **There is one set of ϕ and φ for each amino acid!**

Flexibility to fold is by rotation of ϕ and ψ angles



Rotation of the peptide backbone is only permitted about the N - Cα and the Cα - C=O single bonds. By convention, these rotations are labeled ϕ (phi) and ψ (psi), respectively. **Positive rotations are clockwise rotations looking in the C_α-to-N or C_α-to-CO directions. So the arrows in the diagram shows "negative" rotations.**

ϕ and ψ are 0° when adjacent peptides are coplanar



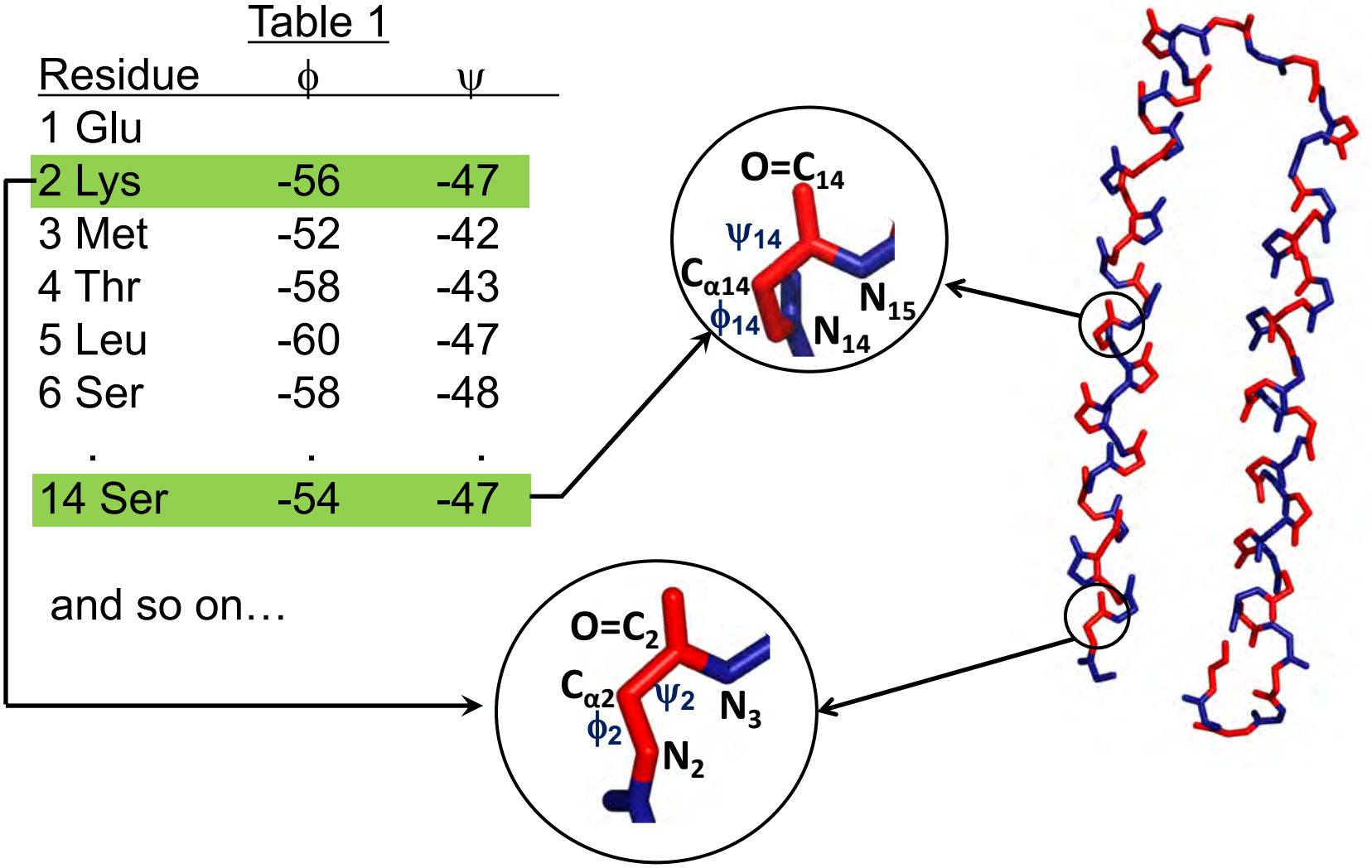
By convention, we assign the angles ϕ and ψ to 0° when the two peptide bonds adjacent to a single Cα carbon are in the same plane. Both angles are 180° when the polypeptide is in a planar fully extended configuration

ϕ and ψ angles define 3D backbone fold

Table 1

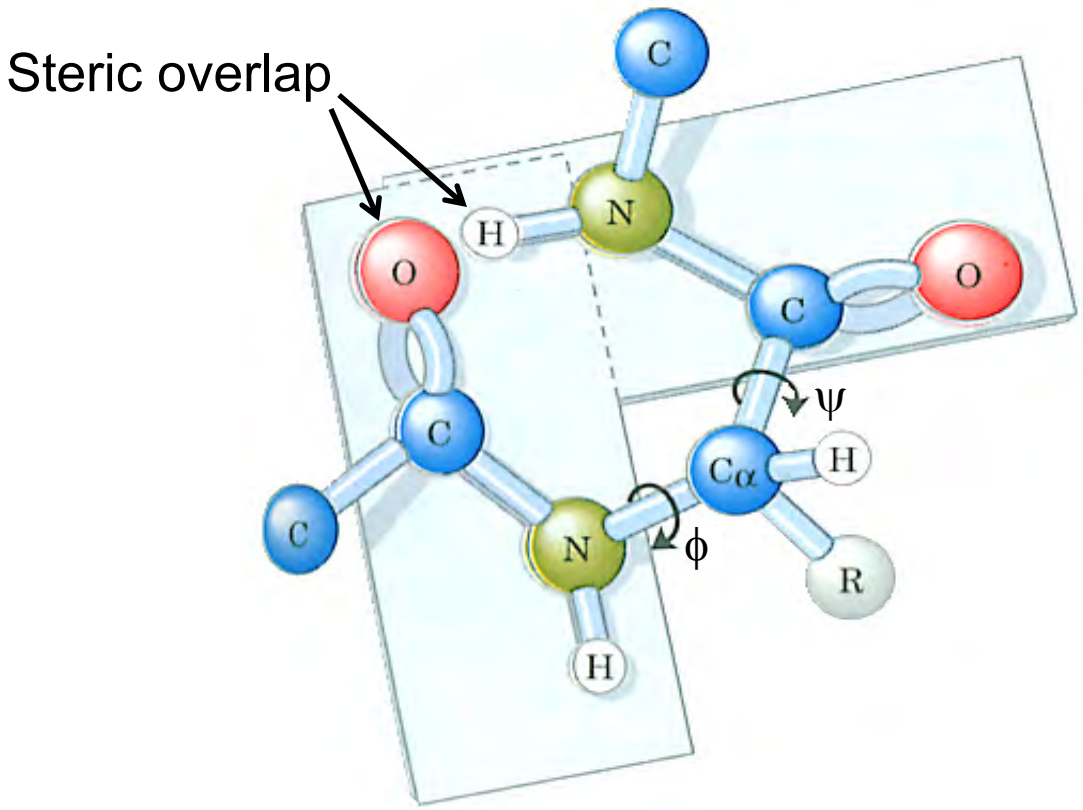
Residue	ϕ	ψ
1 Glu		
2 Lys	-56	-47
3 Met	-52	-42
4 Thr	-58	-43
5 Leu	-60	-47
6 Ser	-58	-48
.	.	.
14 Ser	-54	-47

and so on…



A table of ϕ and ψ angles for each amino acid in a protein defines the 3D globular fold.

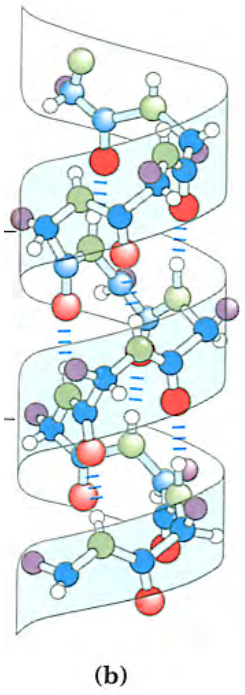
Not all ϕ and ψ angles are sterically possible



When ϕ and ψ are both 0°, there is steric overlap between the carbonyl oxygen of one amino acid and the adjacent peptide N-H – *so not all ϕ and ψ angles are allowed.*

Ramachandran calculated which

In the 1960s, G.N. Ramachandran calculated the "allowed" ϕ (phi) and ψ (psi) angles *where there is no steric overlap between adjacent atoms.* The ϕ and ψ angles that give rise to every type of secondary structure were also calculated:


(b)

α-helix: ϕ ~ -57° and ψ ~ -47°

β-sheet: ϕ ~ -139° and ψ ~ +135°

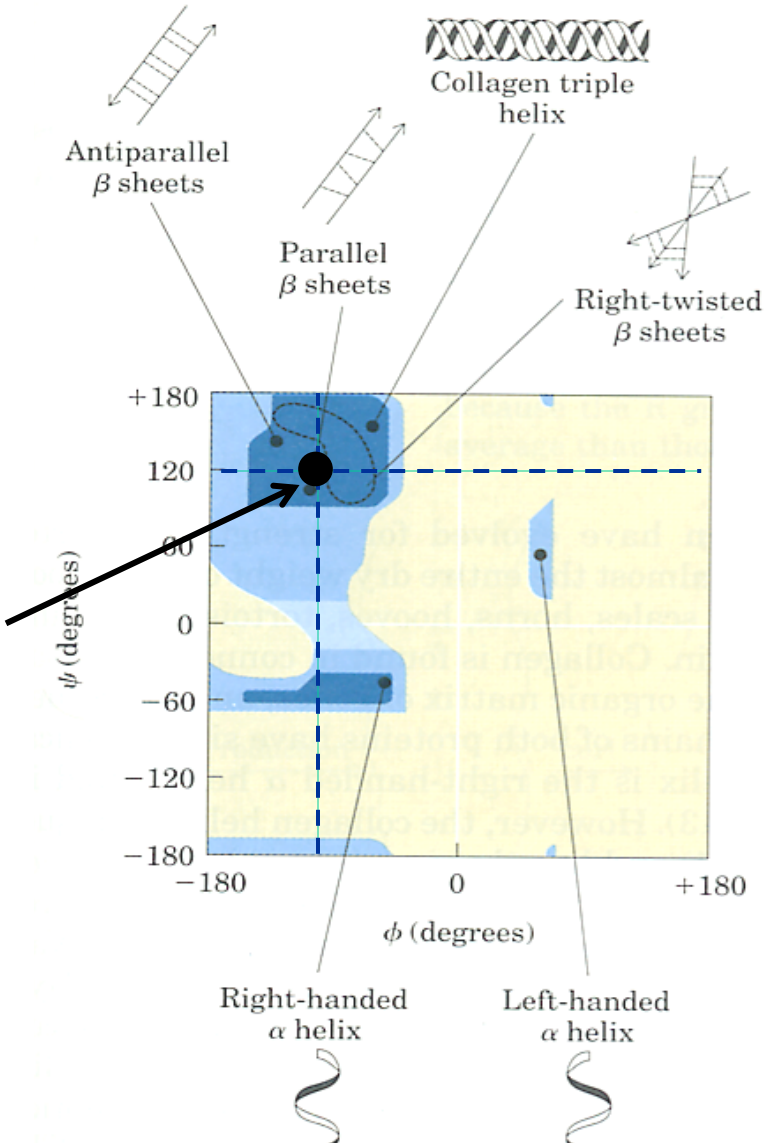
In order to plot all these data in a simple visual format, he developed what he called a Ramachandran plot:

Ramachandran Plot shows allowed ϕ and ψ angles

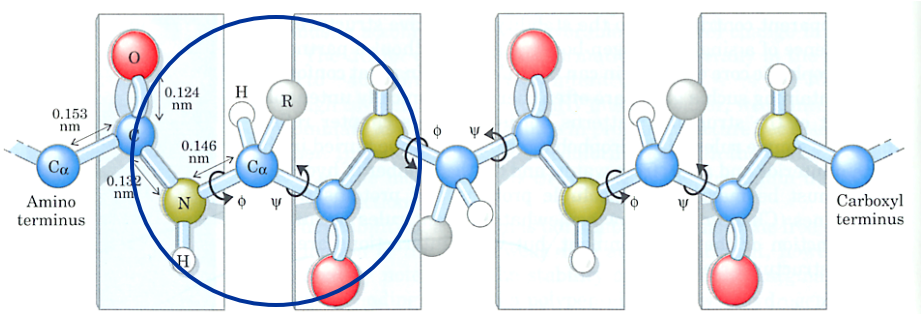
A Ramachandran plot is a 2D plot with one axes corresponding to the ϕ angle and the other corresponding to the ψ angle.

A typical plot shows one spot defining the ϕ and ψ angels for each residue in a protein (so there are usually many spots). For example, a residue with a ϕ angle of -120 and ψ angle of +120, would give rise to the spot located in the diagram on the right.

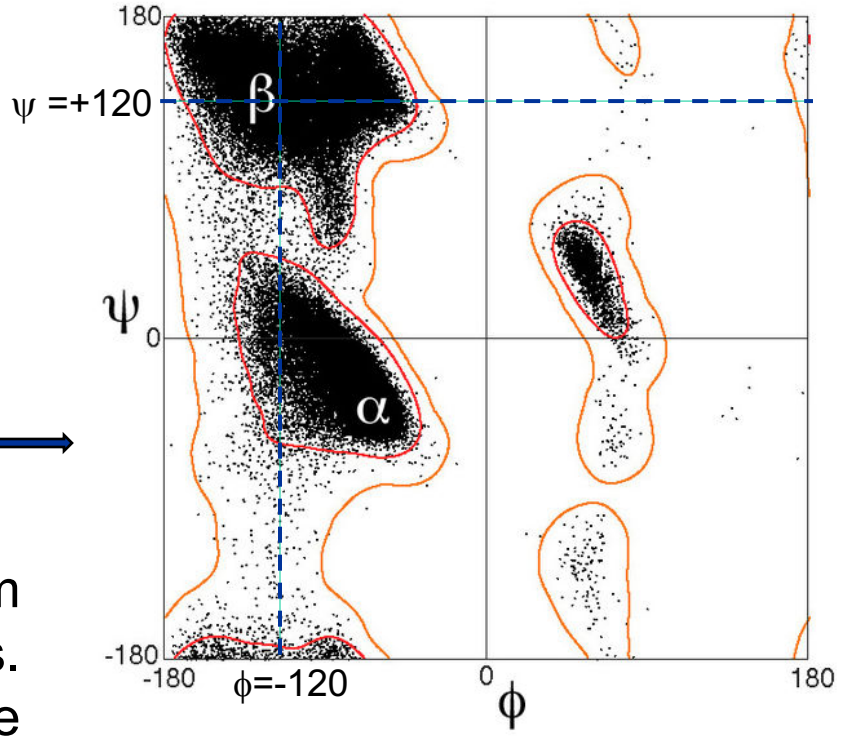
Dark blue show ϕ and ψ angles where there is no steric overlap. Lighter blue shows angles where mild overlap can occur, but can often be accommodated through subtle distortions in structure.



A typical Ramachandran plot has many spots



Residue X
ϕ=-120 and ψ =+120



This plot shows the ϕ and ψ angles from every residue in 500 different proteins. We use the plot to test whether the angles fall within the theoretically allowed ϕ and ψ angles that correspond to low or no steric overlap.

The shown plot containing data from 500 proteins reveals that the vast majority of experimentally observed ϕ and ψ angles match the predictions of Ramachandran.