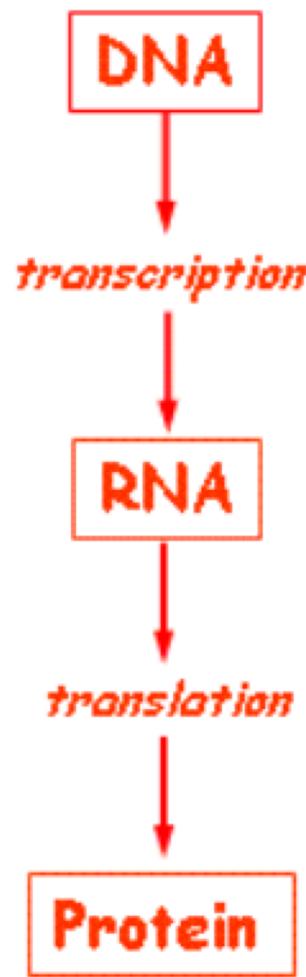


Machine Learning Approaches to Analyze & Predict Ramachandran Plot Distributions

Lecture 03

Central Dogma of Molecular Biology

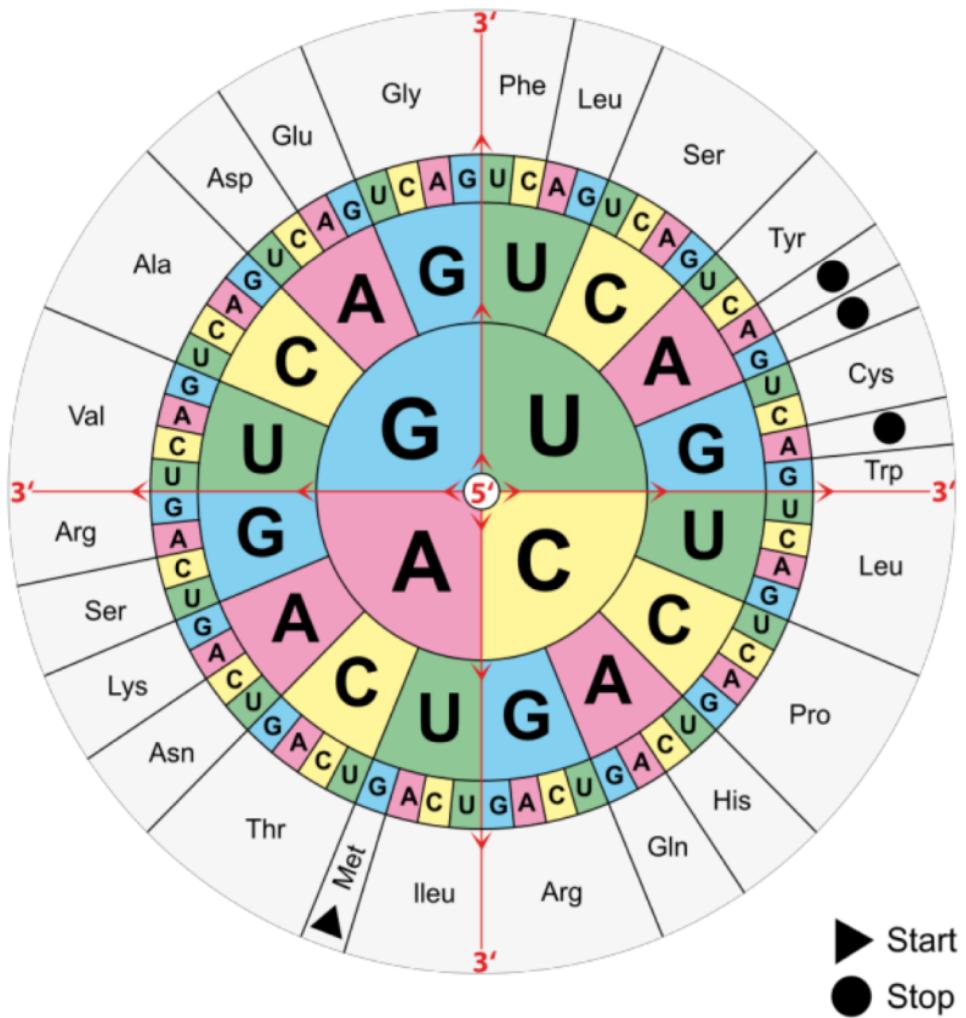


General transfers of biological sequential information

General	Special	Unknown
DNA → DNA (DNA replication)	RNA → DNA (reverse transcription)	protein → DNA
DNA → RNA (transcription)	RNA → RNA (RNA replication)	protein → RNA
RNA → protein (translation)	DNA → protein (direct translation)	protein → protein

Codon Wheel

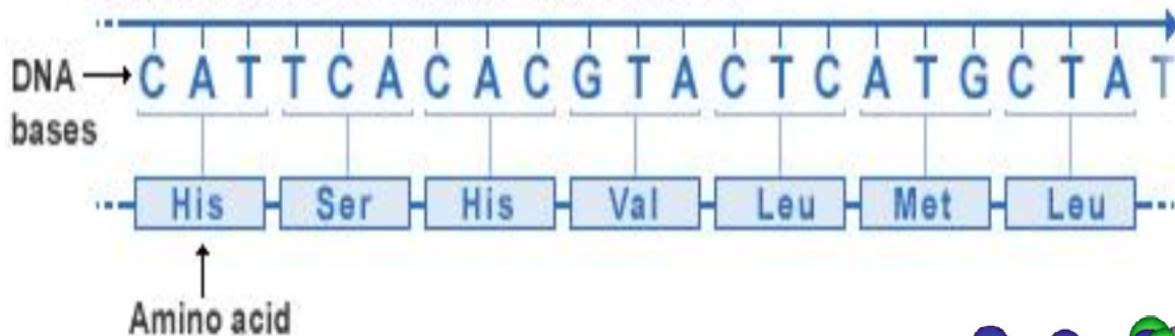
Amino Acids



Amino Acid	3-Letter Code	1-Letter Code
Alanine	Ala	A
Cysteine	Cys	C
Aspartic acid or aspartate	Asp	D
Glutamic acid or glutamate	Glu	E
Phenylalanine	Phe	F
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Leucine	Leu	L
Methionine	Met	M
Asparagine	Asn	N
Proline	Pro	P
Glutamine	Gln	Q
Arginine	Arg	R
Serine	Ser	S
Threonine	Thr	T
Valine	Val	V
Tryptophan	Trp	W
Tyrosine	Tyr	Y

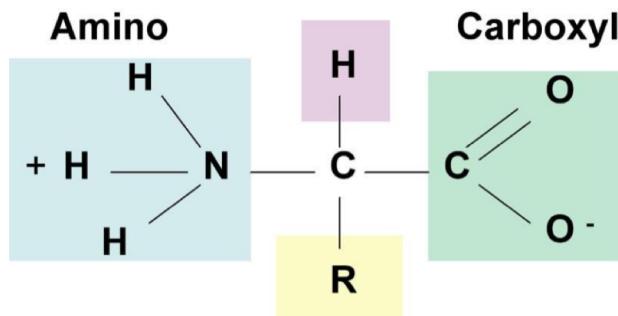
Protein Sequencing

Original DNA code for an amino acid sequence.

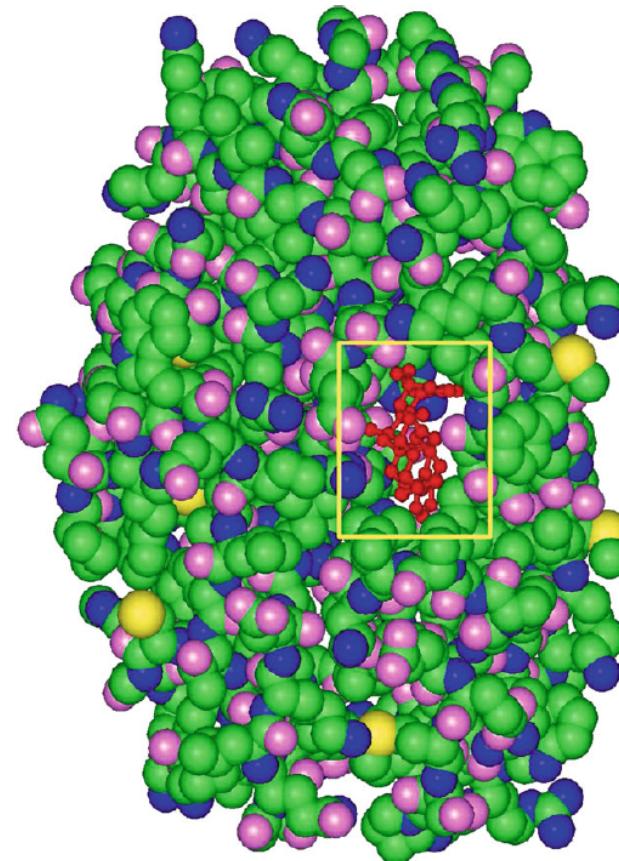


Amino Acid Structure

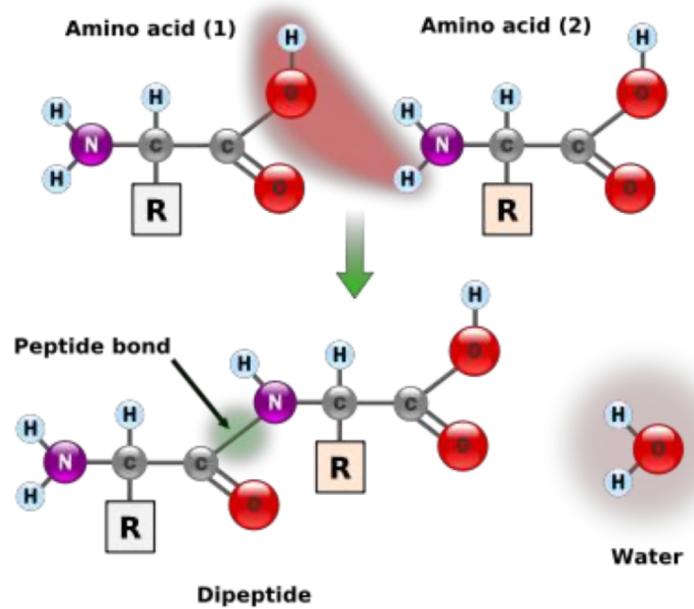
Hydrogen



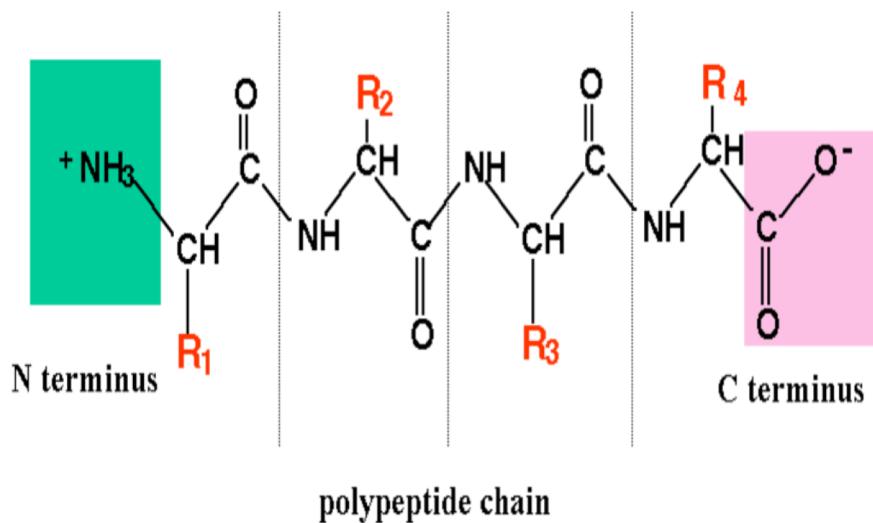
R-group
(variant)



Peptide Formation



Peptide Bonds



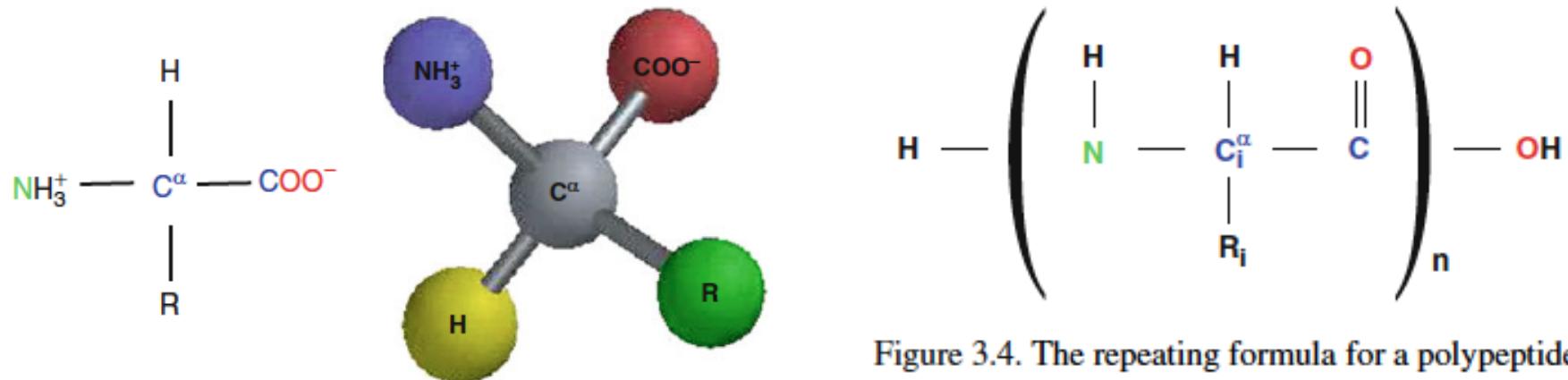


Figure 3.4. The repeating formula for a polypeptide.

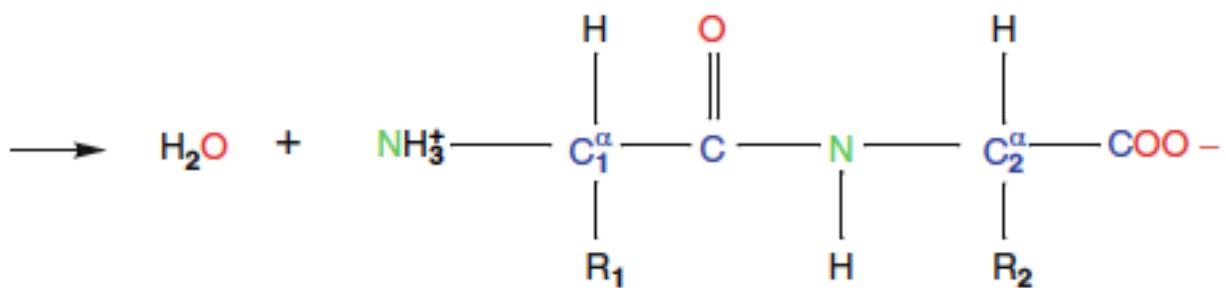
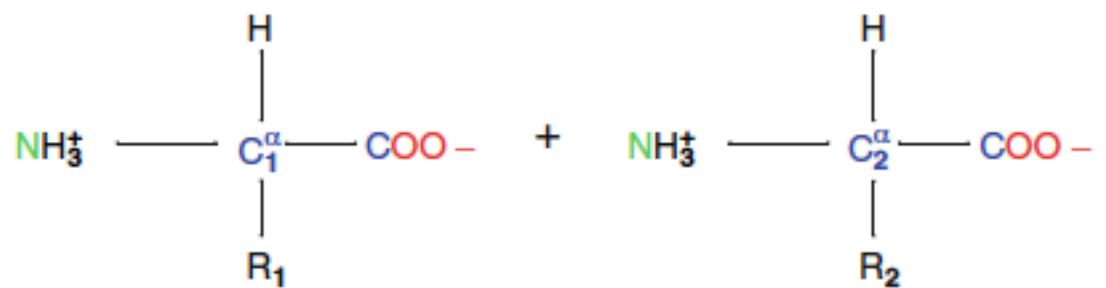
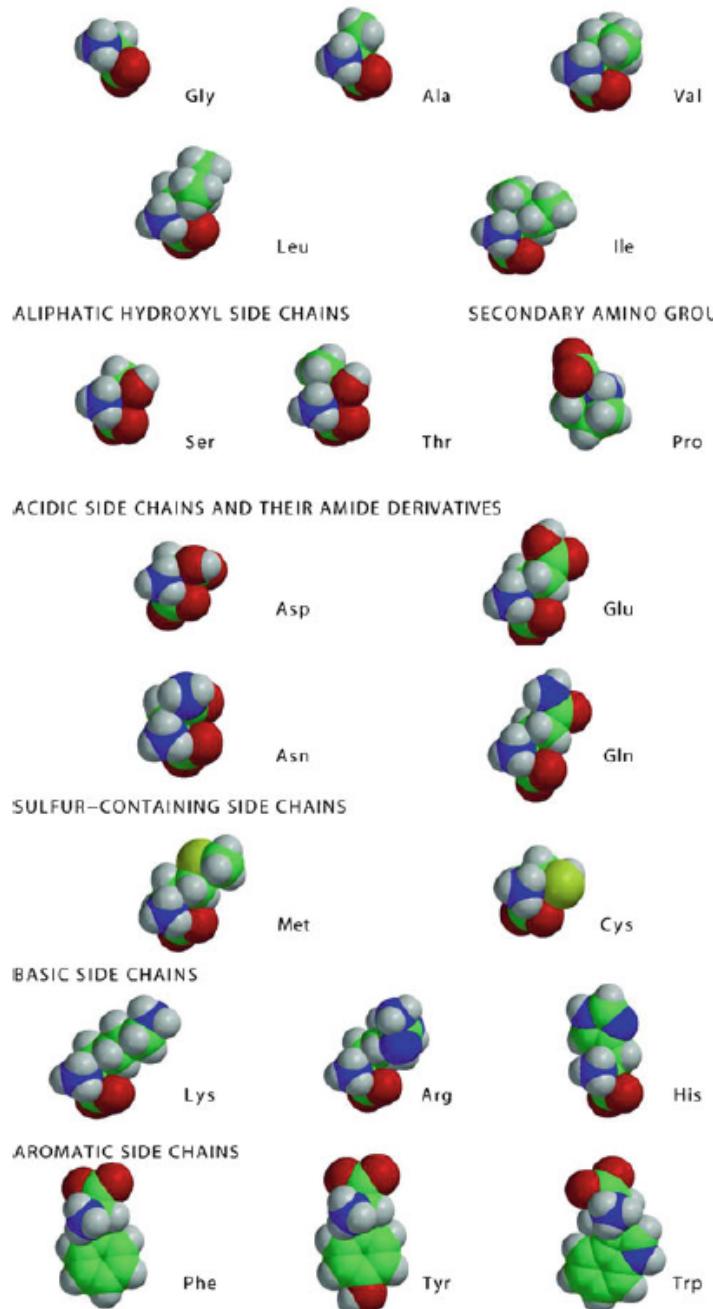


Figure 3.3. Formation of a dipeptide by joining two amino acids.

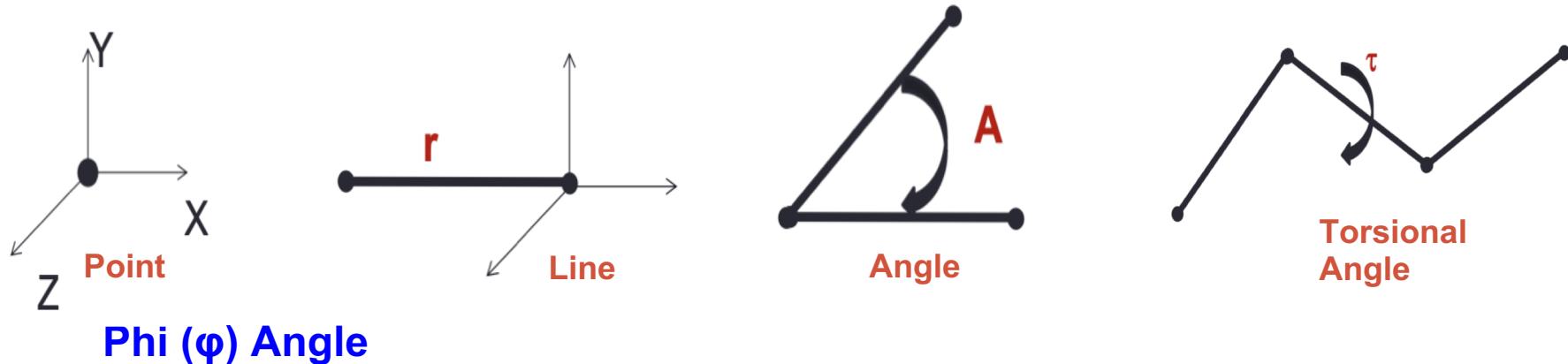
ALIPHATIC SIDE CHAINS



Amino Acid	Freq. [%]
Alanine (Ala, A)	8.1
Arginine (Arg, R)	5.1
Asparagine (Asp, D)	5.2
Aspartic acid (Asn, N)	4.0
Cysteine (Cys, C)	1.2
Glutamine (Gln, Q)	3.8
Glutamic acid (Glu, E)	6.5
Glycine (Gly, G)	7.2
Histidine (His, H)	2.2
Isoleucine (Ile, I)	6.8
Leucine (Leu, L)	10.3
Lysine (Lys, K)	5.9
Methionine (Met, M)	2.5
Phenylalanine (Phe, F)	4.2
Proline (Pro, P)	4.3
Serine (Ser, S)	6.2
Threonine (Thr, T)	5.1
Tryptophan (Trp, W)	1.1
Tyrosine (Tyr, Y)	3.2
Valine (Val, V)	6.9

Figure 3.7. Space-filling models of the 20 amino acids.

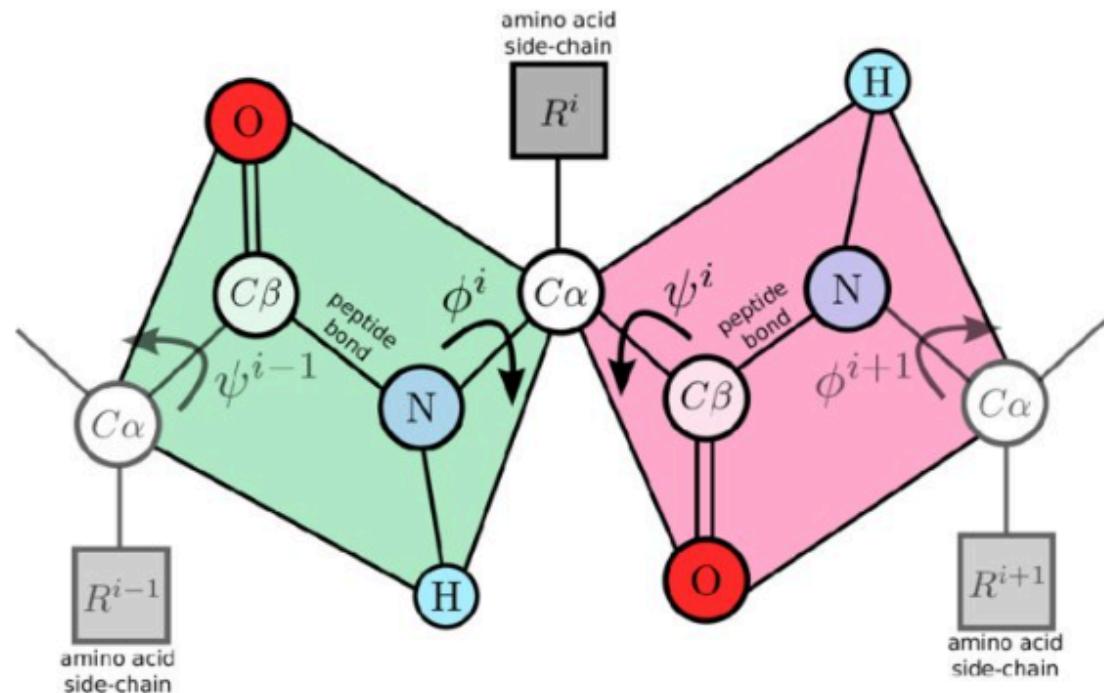
Torsional Angle



Phi (ϕ) Angle

Definition: The phi (ϕ) angle is the rotation around the N-C α bond.

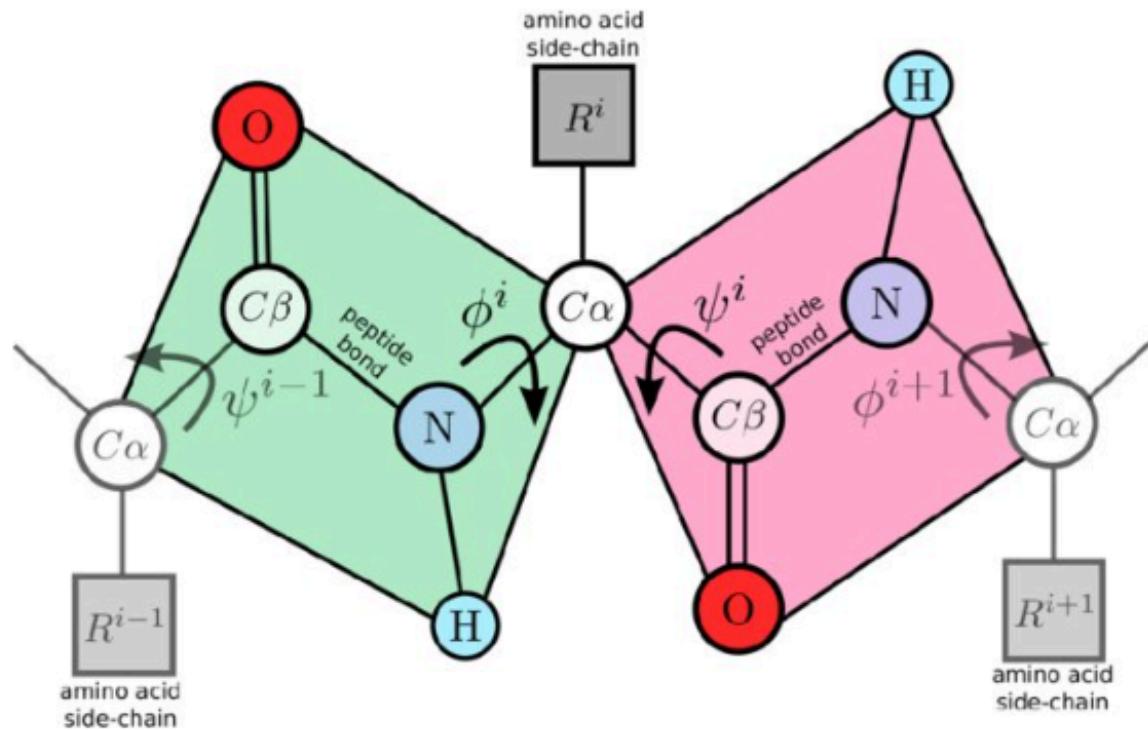
Measurement: It is measured between the plane formed by the atoms C-N-C α -C and the plane formed by the atoms N-C α -C-N.



Torsional Angles

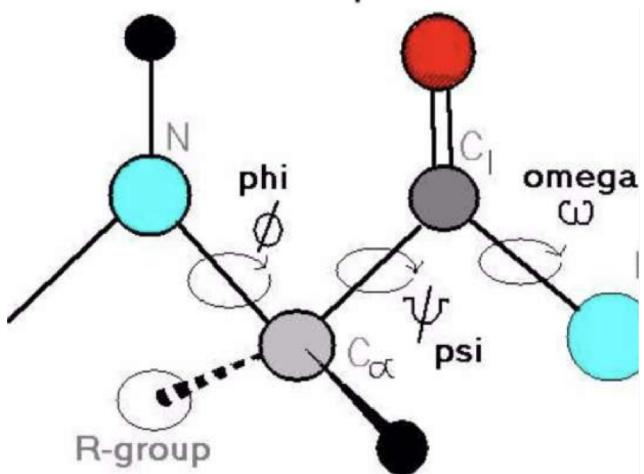
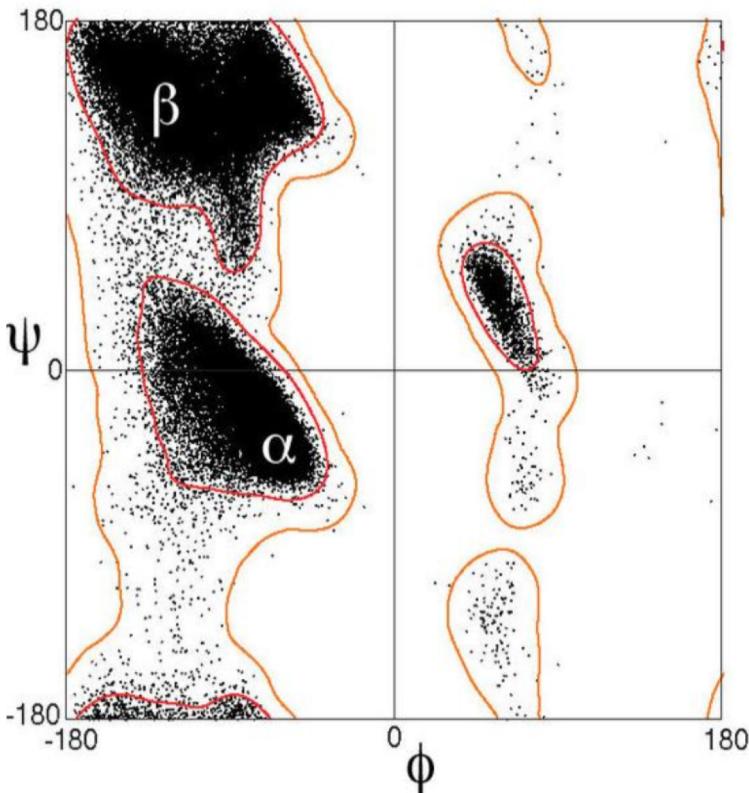
Psi (ψ) Angle

- The psi (ψ) angle is the rotation around the $C\alpha$ -C bond.
- Measurement:** It is measured between the plane formed by the atoms N- $C\alpha$ -C-N and the plane formed by the atoms C- $C\alpha$ -N-C.



Ramachandran Plot

- A Ramachandran plot is a useful way to visualise dihedral angles ϕ (phi) and ψ (psi) of amino acid residues in protein structures.
- These plots help in understanding the conformational constraints of proteins.
- The Ramachandran plot provides an overview of the allowed and disallowed regions of torsion angle values.
- These regions are important indicators of the quality of protein three-dimensional structures.
- Torsion angles, specifically ϕ , ψ , are among the most critical local structural parameters that control protein folding.

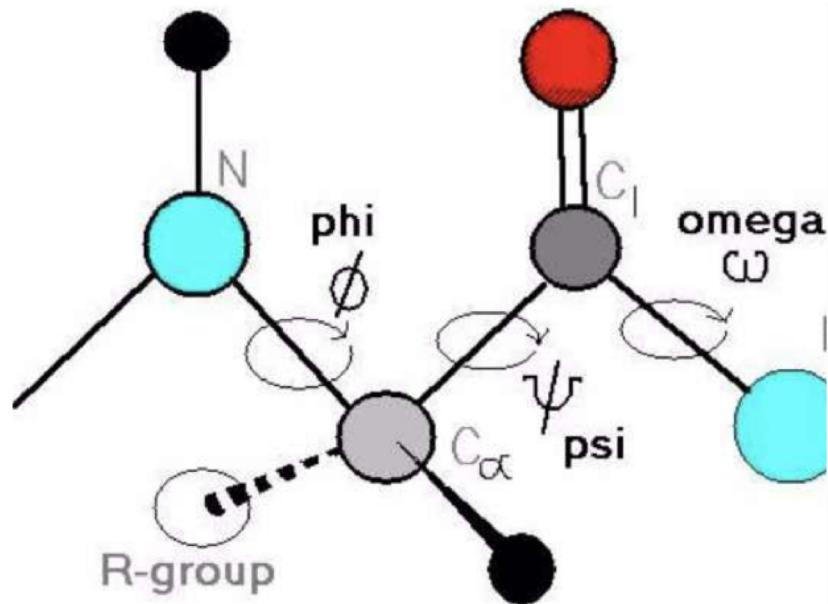


Torsional Angles in Ramachandran Plot

Describing the Protein Backbone Using Torsion Angles

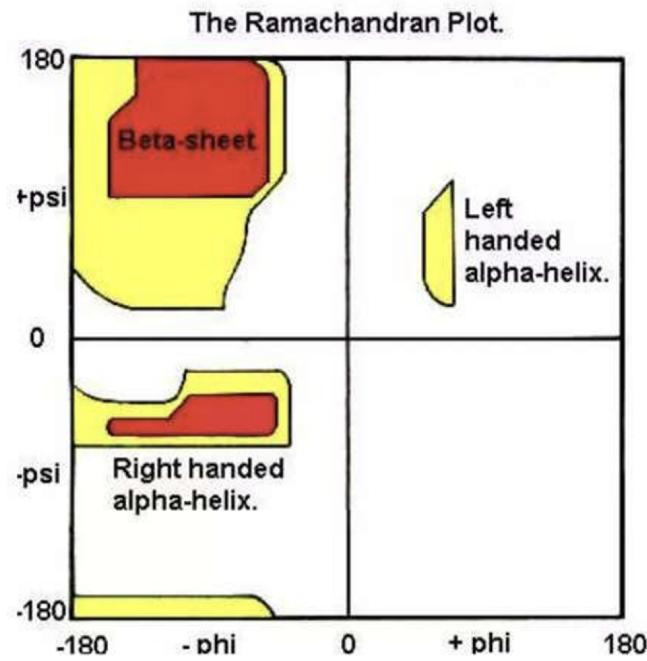
The protein backbone can be described in terms of the ϕ (phi), ψ (psi) torsion angles of the bonds:

- **Phi (ϕ) angle:** The angle around the -N-CA- bond, where 'CA' is the alpha-carbon.
- **Psi (ψ) angle:** The angle around the -CA-C- bond.



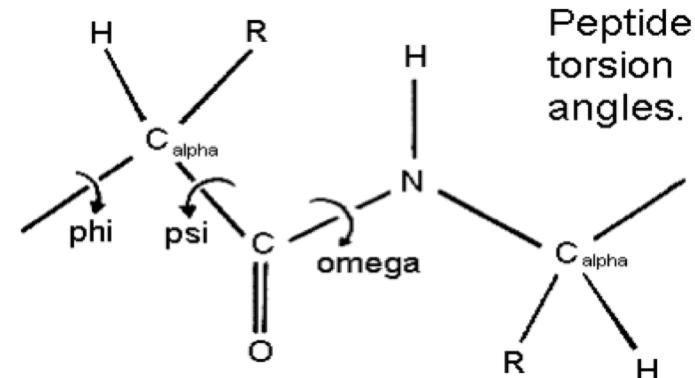
Understanding the Ramachandran Plot and Protein Folding

- If we could predict the Ramachandran angles (ϕ and ψ) for a specific protein, we could foresee its 3D structure.
- This is because these angles determine the flexibility required for the polypeptide backbone to adopt a specific fold, as the ω angle is essentially flat and fixed at 180 degrees.
- This rigidity is due to the partial double-bond character of the peptide bond.

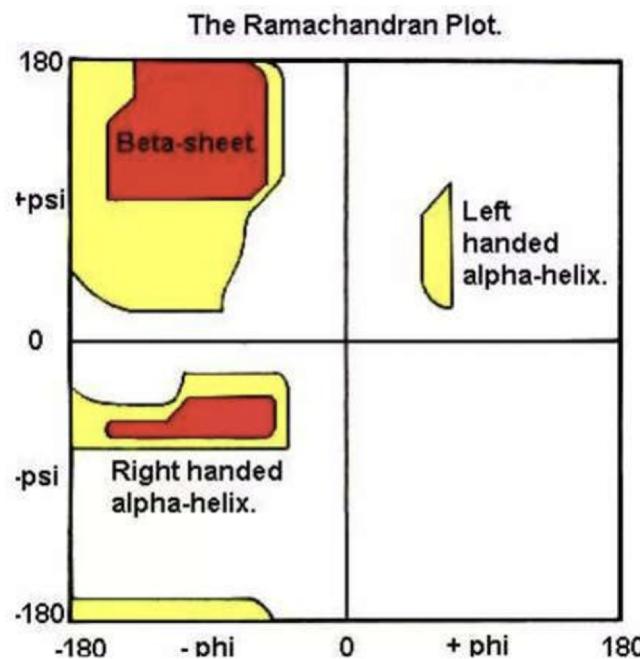


Understanding the Ramachandran Plot and Protein Folding (contd.)

- This character restricts rotation around the C-N bond, placing two successive alpha-carbons, along with carbon, oxygen, nitrogen, and hydrogen, in the same plane.

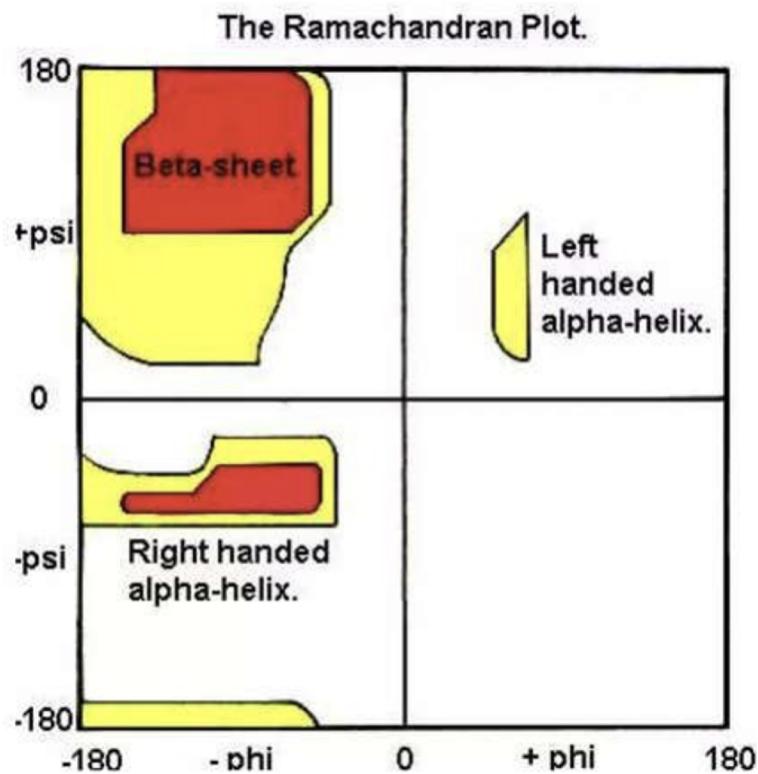


- Therefore, the rotation of the protein chain is essentially the rotation of the peptide bond planes relative to each other.
- By understanding and predicting the ϕ and ψ angles, which are plotted in the Ramachandran plot, one can predict the possible conformations and the overall 3D folding of the protein.



Understanding Disallowed Regions in the Ramachandran Plot

- ❖ In the Ramachandran plot, the white areas correspond to conformations where atoms in the polypeptide come closer than the sum of their Van der Waals radii.
 - These regions are sterically disallowed for all amino acids except glycine.
 - Glycine is unique because it lacks a side chain, allowing it to occupy regions in the plot that other amino acids cannot.



Region Classification

□ Alpha Helices

- Typically found in the regions where ϕ is around -60° and ψ is around -45° .

□ Beta Sheets

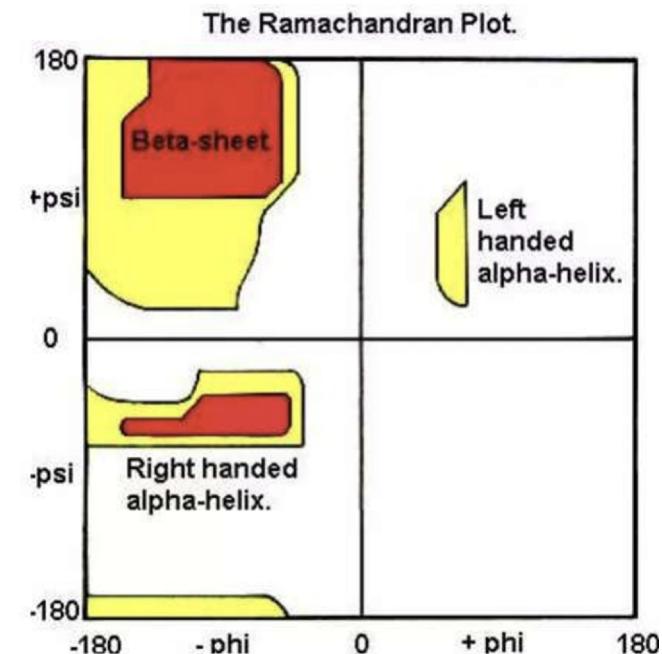
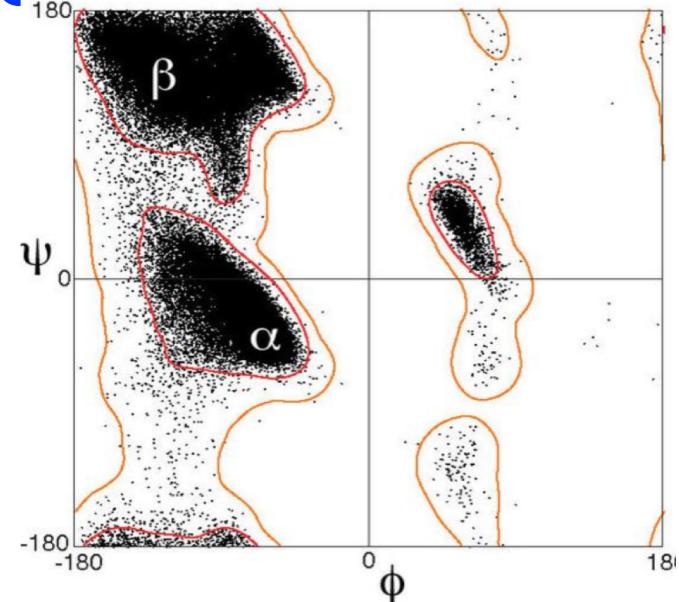
- Found in the regions where ϕ is around -135° and ψ is around 135° .

□ Left-handed Helices

- Less common, located in the region where ϕ is around 60° and ψ is around 45° .

□ Disallowed Regions

- Areas where steric hindrance makes certain combinations of ϕ and ψ angles energetically unfavorable.



Objectives

- Predicting dihedral angles for given amino acid sequences.
- Classifying protein structures based on their Ramachandran plot distributions.
- Analyzing the impact of mutations on protein conformation.

Evaluating Dihedral Angles

Atomic coordinates

$$\mathbf{a}_1 = (a_1x, a_1y, a_1z)$$

$$\mathbf{a}_2 = (a_2x, a_2y, a_2z)$$

$$\mathbf{a}_3 = (a_3x, a_3y, a_3z)$$

$$\mathbf{a}_4 = (a_4x, a_4y, a_4z)$$

Bond vectors

$$\mathbf{b}_{12} = (a_2x - a_1x, a_2y - a_1y, a_2z - a_1z)$$

$$\mathbf{b}_{23} = (a_3x - a_2x, a_3y - a_2y, a_3z - a_2z)$$

$$\mathbf{b}_{34} = (a_4x - a_3x, a_4y - a_3y, a_4z - a_3z)$$

Normal vectors

$$\mathbf{n}_{123} = \mathbf{b}_{12} \times \mathbf{b}_{23}$$

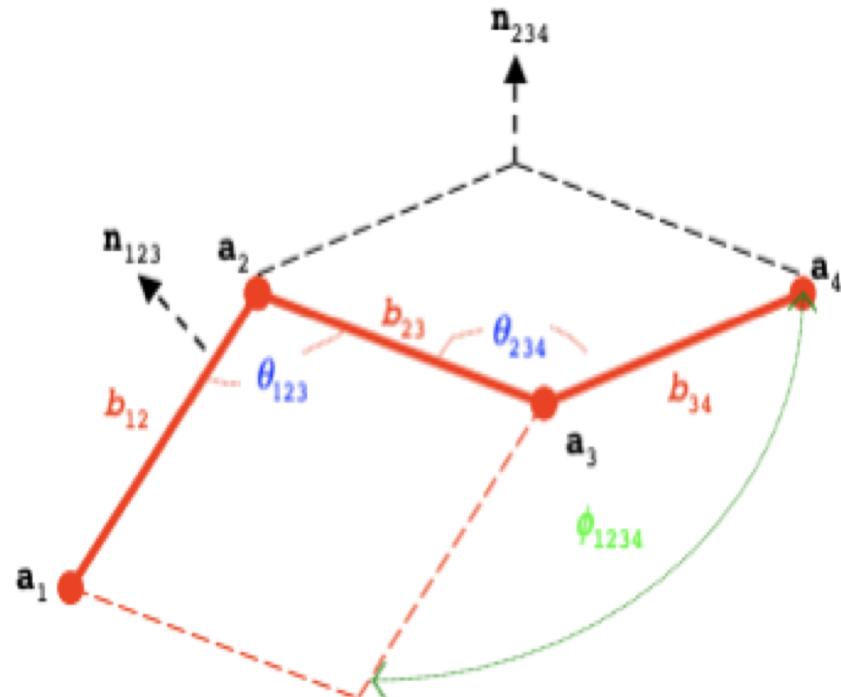
$$\mathbf{n}_{234} = \mathbf{b}_{23} \times \mathbf{b}_{34}$$

$$\mathbf{n}_{123} \cdot \mathbf{n}_{234} = \cos \varphi_{1234}$$

Sign determination

$$(\mathbf{n}_{123} \times \mathbf{n}_{234}) \cdot \mathbf{b}_{23} > 0 \Rightarrow \varphi_{1234} > 0$$

$$(\mathbf{n}_{123} \times \mathbf{n}_{234}) \cdot \mathbf{b}_{23} < 0 \Rightarrow \varphi_{1234} < 0$$



Evaluate dihedral angles for this—

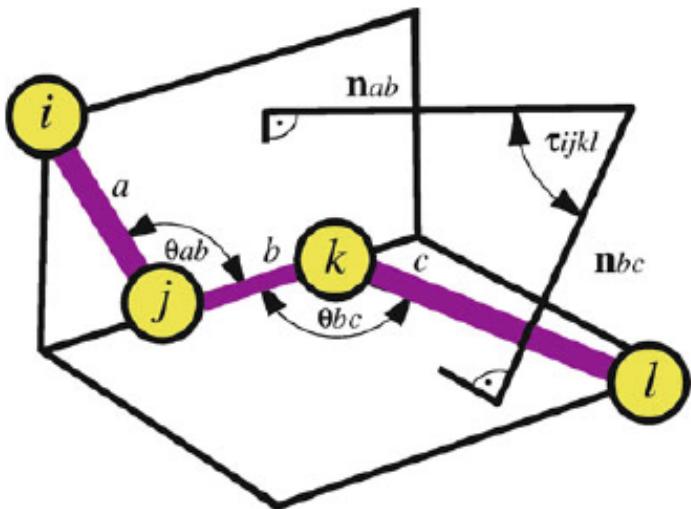
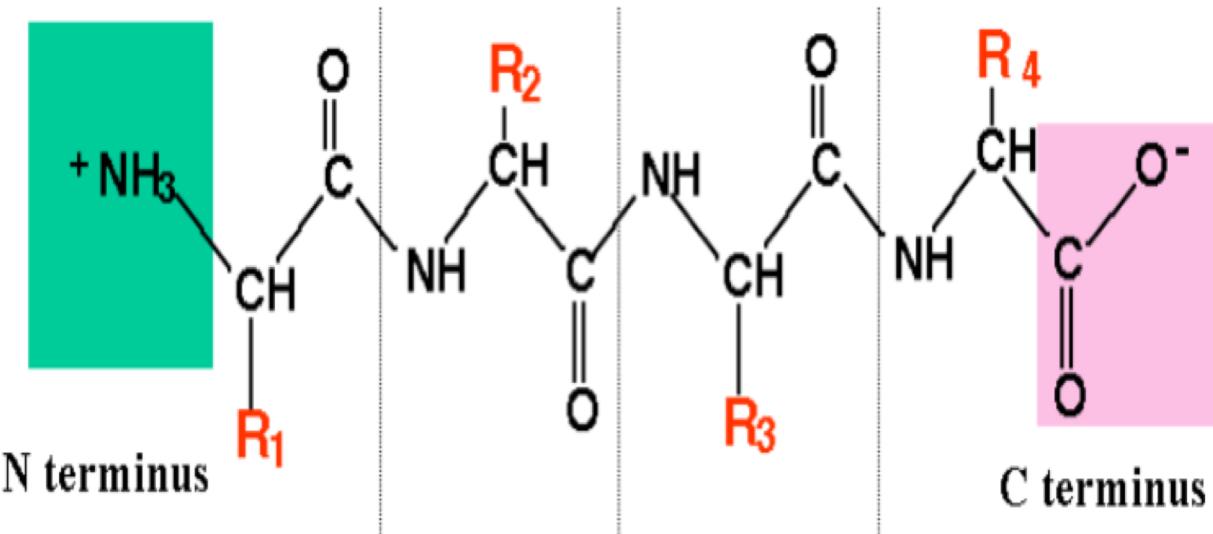


Figure 3.14. Definition of a dihedral angle $\tau_{ijkl} = \cos^{-1}(\mathbf{n}_{ab} \cdot \mathbf{n}_{bc})$, the angle between the two normals spanned by atoms i, j, k and j, k, l .

N	11.751	37.846	29.016
CA	12.501	39.048	28.539
C	13.740	38.628	27.754
N	14.235	39.531	26.906
CA	15.552	39.410	26.282
C	16.616	38.913	27.263
N	16.789	39.630	28.369
CA	17.791	39.281	29.375
C	17.598	37.844	29.863
N	16.368	37.519	30.261
CA	16.004	36.186	30.742
C	16.371	35.097	29.741

Ψ angle determination (From 1st residue)

Bond vectors

$$\mathbf{b}_{12} = (\mathbf{a}_2 - \mathbf{a}_1) = 0.75, 1.202, -0.477$$

$$\mathbf{b}_{23} = (\mathbf{a}_3 - \mathbf{a}_2) = 1.239, -0.42, -0.785$$

$$\mathbf{b}_{34} = (\mathbf{a}_4 - \mathbf{a}_3) = 0.495, 0.903, -0.848$$

(a1)	N	11.751	37.846	29.016
(a2)	CA	12.501	39.048	28.539
(a3)	C	13.740	38.628	27.754
(a4)	N	14.235	39.531	26.906

Normal vectors

$$\mathbf{n}_{123} = \mathbf{b}_{12} \times \mathbf{b}_{23} = -1.14391, -0.002253, -1.804278$$

$$\mathbf{n}_{234} = \mathbf{b}_{23} \times \mathbf{b}_{34} = 1.065015, 0.662097, 1.326717$$

$$\mathbf{n}_{123} \cdot \mathbf{n}_{234} = \cos \Psi_{1234} = \frac{1.065015 \times (-1.14391) + 0.662097 \times (-0.002253) + 1.326717 \times (-1.804278)}{\sqrt{1.065015^2 + 0.662097^2 + 1.326717^2} \times \sqrt{1.14391^2 + 0.002253^2 + 1.804278^2}}$$

$$= \frac{3.613553}{2.13634 \times 1.8256}$$

$$-\cos \Psi_{1234} = -0.92653$$

$$\Psi_{1234} = \pm 157.9^\circ$$

Sign Determination

$$(n_{123} \times n_{234}) \cdot b_{23} > 0 \Rightarrow \Psi_{1234} > 0$$

$$(n_{123} \times n_{234}) \cdot b_{23} < 0 \Rightarrow \Psi_{1234} < 0$$

Here,

$$(n_{123} \times n_{234}) = 1.191618, -0.40394, -0.75498$$

$$(n_{123} \times n_{234}) \cdot b_{23} = 2.2387 > 0$$

$$\therefore \Psi_{1234} = + 157.9^\circ$$

Calculate the same for φ for next residue.

φ angle determination (From 2nd residue)

Bond vectors

$b_{12} = (a_2x - a_1x, a_2y - a_1y, a_2z - a_1z) =$	0.495, 0.903, -0.848	(a1) C	13.740	38.628	27.754
$b_{23} = (a_3x - a_2x, a_3y - a_2y, a_3z - a_2z) =$	1.317, -0.121, -0.624	(a2) N	14.235	39.531	26.906
$b_{34} = (a_4x - a_3x, a_4y - a_3y, a_4z - a_3z) =$	1.064, -0.497, 0.981	(a3) CA	15.552	39.410	26.282
		(a4) C	16.616	38.913	27.263

Normal vectors

$$n_{123} = b_{12} \times b_{23} = -0.6661, -0.8079, -1.249$$

$$n_{234} = b_{23} \times b_{34} = -0.4288, -1.9559, -0.5258$$

$$n_{123} \cdot n_{234} = \cos \varphi_{1234} = \frac{(-0.6661) \times (-0.4288) + (-0.8070) \times (-1.9559) + (-1.249) \times (-0.5258)}{\left(\sqrt{0.6661^2 + 0.8070^2 + 1.249^2} \right) \times \left(\sqrt{0.4288^2 + 1.9559^2 + 0.5258^2} \right)}$$

$$= \frac{2.5225}{1.6298 \times 2.07}$$

$$\cos \varphi_{1234} = 0.747698$$

Sign Determination

$$(n_{123} \times n_{234}) = -2.018, 0.1853, 0.9564$$

$$\varphi_{1234} = \pm 41.6^\circ$$

$$(n_{123} \times n_{234}) \cdot b_{23} = -3.2769 < 0$$

$$\therefore \varphi_{1234} < 0$$

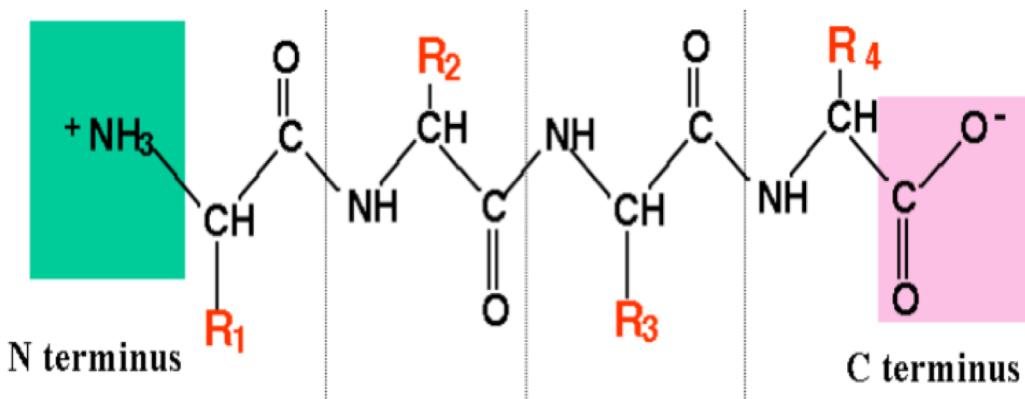
$$\therefore \varphi_{1234} = -41.6^\circ$$

Numerical

- 1.A protein's Ramachandran plot shows the following ϕ angles : -62° , -58° , -61° , -65° , -59° , 120° , -130° , 55° , 45° , 159° . Calculate the percentage of residues in each region.
- 2.Calculate the mean and standard deviation of these ϕ angles for each region.

Homework

Write a pseudo code for evaluating phi & psi angles.



N	11.751	37.846	29.016
CA	12.501	39.048	28.539
C	13.740	38.628	27.754

N	14.235	39.531	26.906
CA	15.552	39.410	26.282
C	16.616	38.913	27.263

N	16.789	39.630	28.369
CA	17.791	39.281	29.375
C	17.598	37.844	29.863

N	16.368	37.519	30.261
CA	16.004	36.186	30.742
C	16.371	35.097	29.741