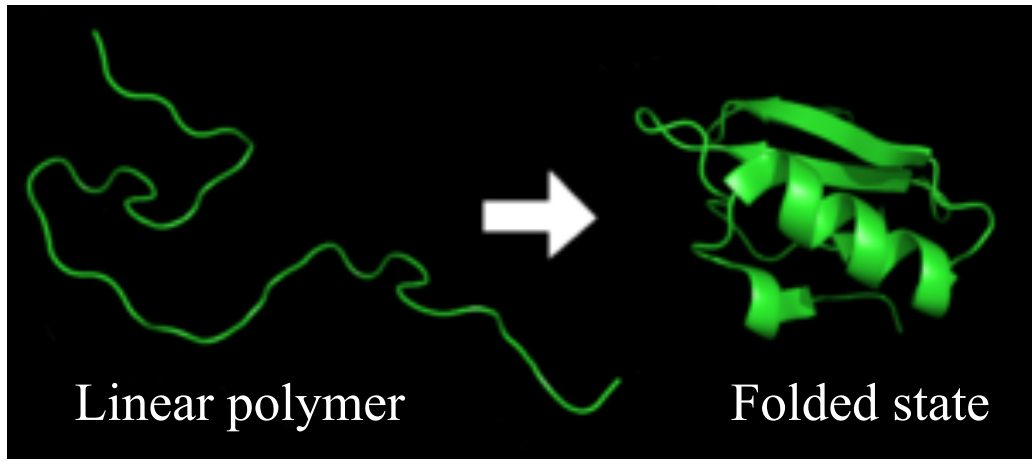
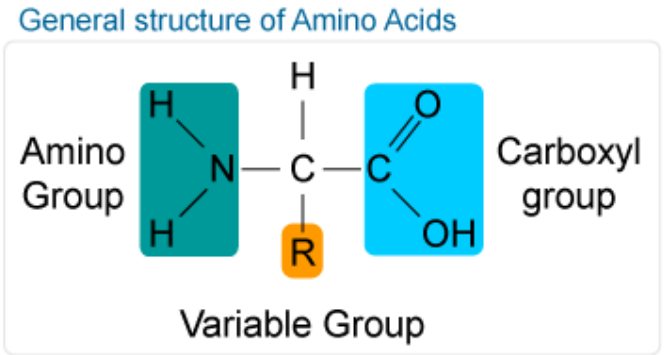
What are proteins?



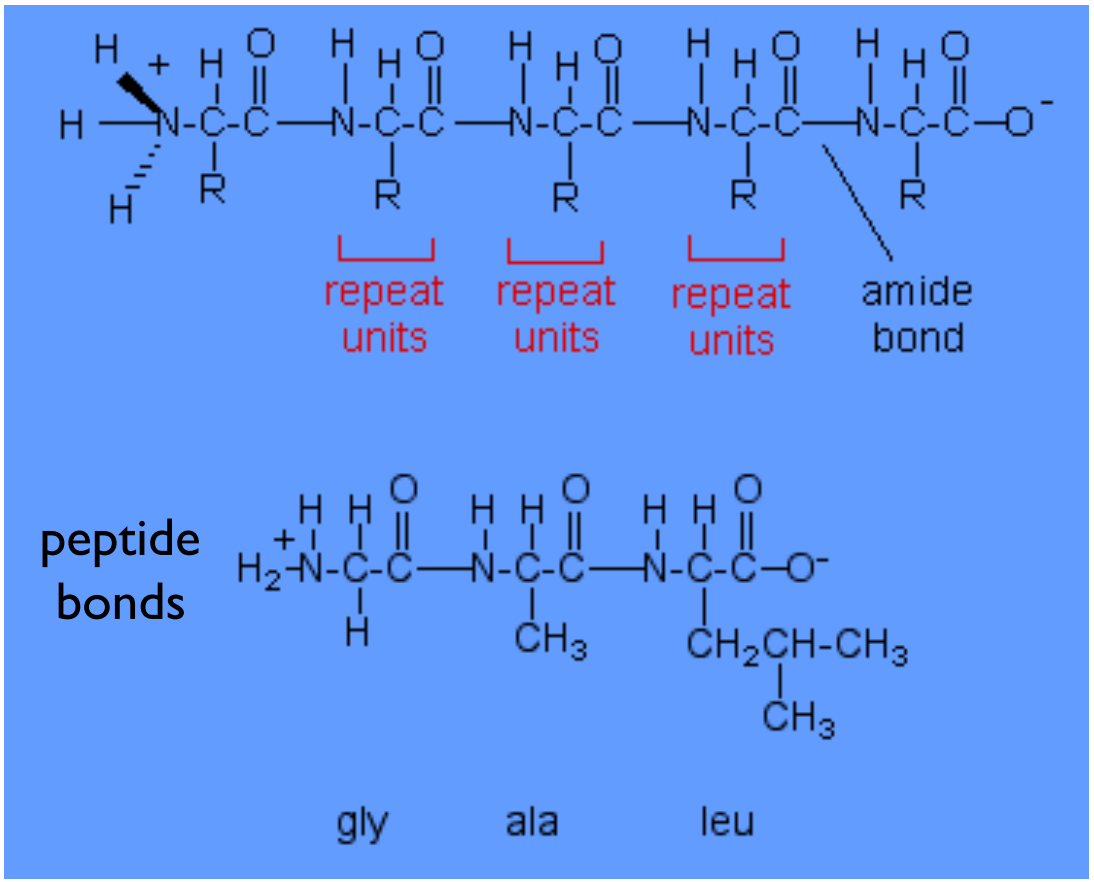
- Proteins are important; e.g. for catalyzing and regulating biochemical reactions, transporting molecules, …
- Linear polymer chain composed of tens (peptides) to thousands (proteins) of monomers
- Monomers are 20 naturally occurring amino acids
- Different proteins have different amino acid sequences
- *Structureless*, extended unfolded state
- Compact, 'unique' native folded state (with secondary and tertiary structure) required for biological function
- Sequence determines protein structure (or lack thereof)
- Proteins unfold or denature with increasing temperature or chemical denaturants

Amino Acids I



N-terminal C_α C-terminal

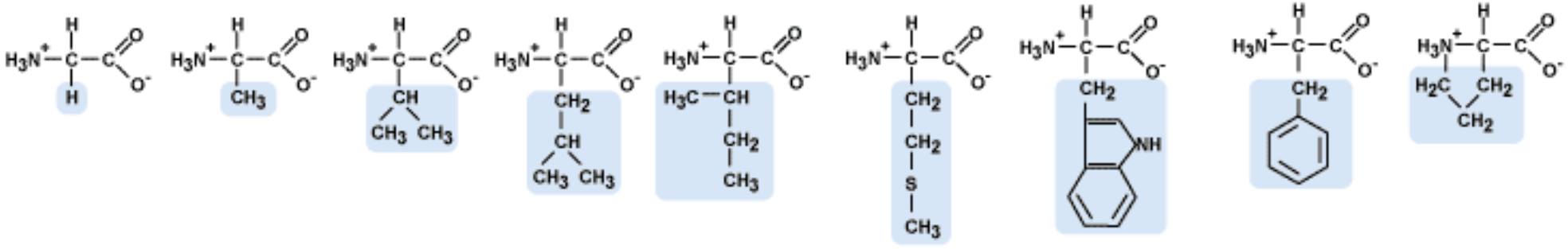
R
variable
side chain



- Side chains differentiate amino acid repeat units
- Peptide bonds link residues into polypeptides

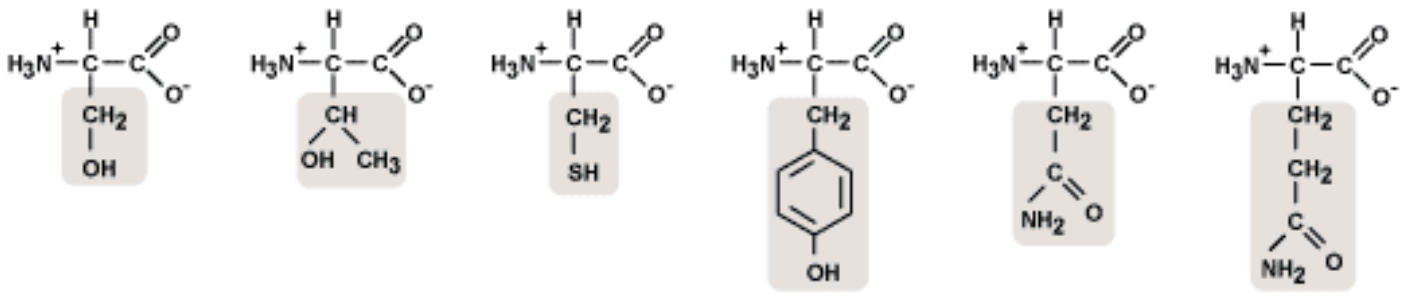
Amino Acids II

NONPOLAR



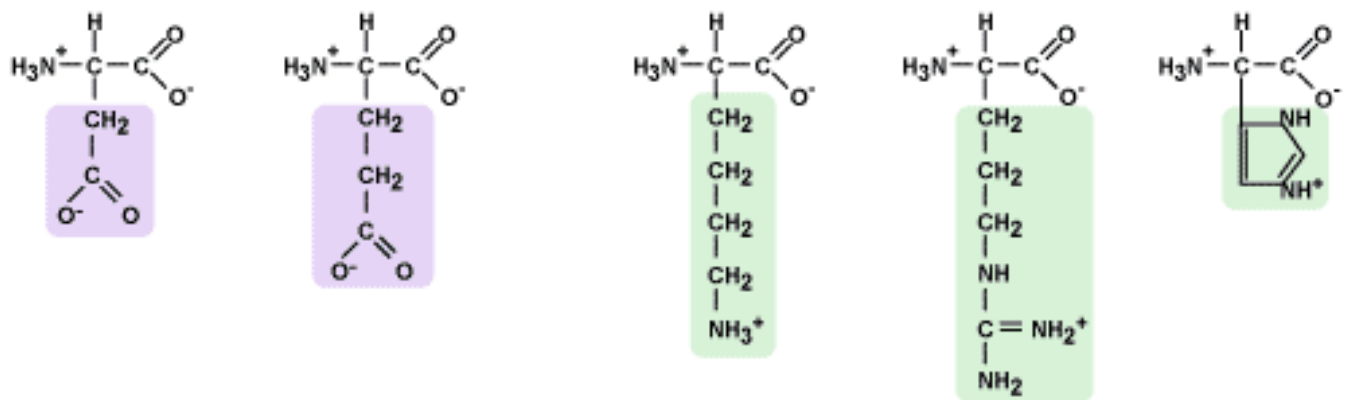
Glycine (Gly) Alanine (Ala) Valine (Val) Leucine (Leu) Isoleucine (Ile) Methionine (Met) Tryptophan (Trp) Phenylalanine (Phe) Proline (Pro)

POLAR



Serine (Ser) Threonine (Thr) Cysteine (Cys) Tyrosine (Tyr) Asparagine (Asn) Glutamine (Gln)

Electrically Charged



Acidic

Aspartic Acid (Asp) Glutamic Acid (Glu)

Lysine (Lys)

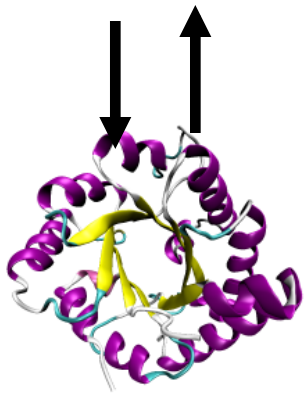
Basic

Arginine (Arg) Histidine (His)

The Protein Folding Problem:

What is ‘unique’ folded 3D structure of a protein based on its amino acid sequence? Sequence → Structure

Lys–Asn–Val–Arg–Ser–Lys–Val–Gly–Ser–Thr–Glu–Asn–Ile–Lys– His–Gln–Pro– Gly–Gly–Gly–…



Driving Forces

- Folding: hydrophobicity, hydrogen bonding, van der Waals interactions, …
- Unfolding: increase in conformational entropy, electric charge…

inside H (hydrophobic)

outside P (polar)

Hydrophobicity index

At pH 2[A]		At pH 7[B]	
Very Hydrophobic			
Leu	100	Phe	100
Ile	100	Ile	99
Phe	92	Trp	97
Trp	84	Leu	97
Val	79	Val	76
Met	74	Met	74
Hydrophobic			
Cys	52	Tyr	63
Tyr	49	Cys	49
Ala	47	Ala	41
Neutral			
Thr	13	Thr	13
Glu	8	His	8
Gly	0	Gly	0
Ser	-7	Ser	-5
Gln	-18	Gln	-10
Asp	-18		
Hydrophilic			
Arg	-26	Arg	-14
Lys	-37	Lys	-23
Asn	-41	Asn	-28
His	-42	Glu	-31
Pro	-46	Pro	-46 (used pH 2)
		Asp	-55

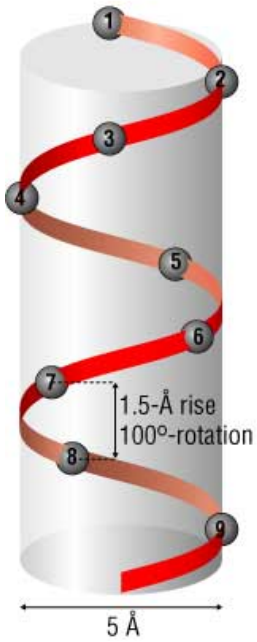
[A]pH 2 values: Normalized from Sereda et al., J. Chrom. 676: 139-153 (1994).
[B]pH 7 values: Monera et al., J. Pept. Sci. 1: 319-329 (1995).

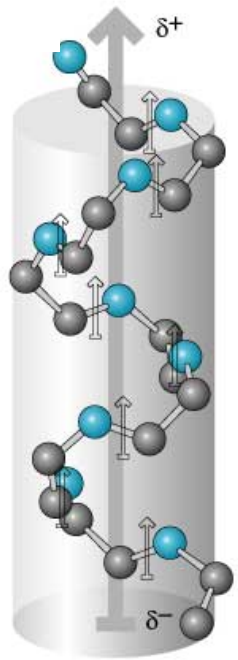
k_bT 5-12 k_bT > 150 k_bT

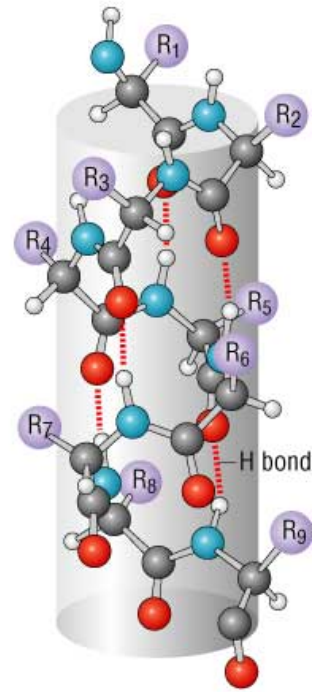
Van der Waals interactions – hydrogen bonds – covalent bonds

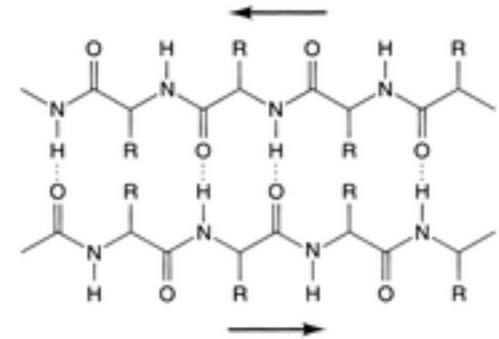
Secondary Structure: Loops, α-helices, β-strands/sheets

α-helix

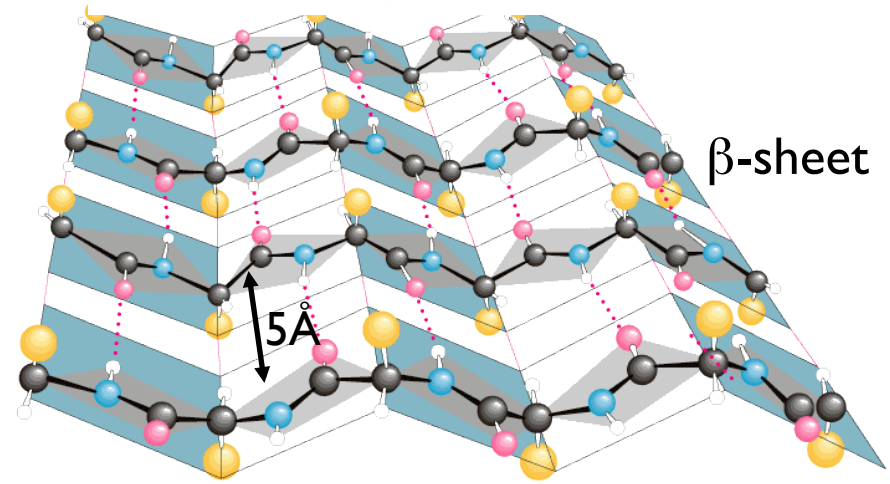








β-strand

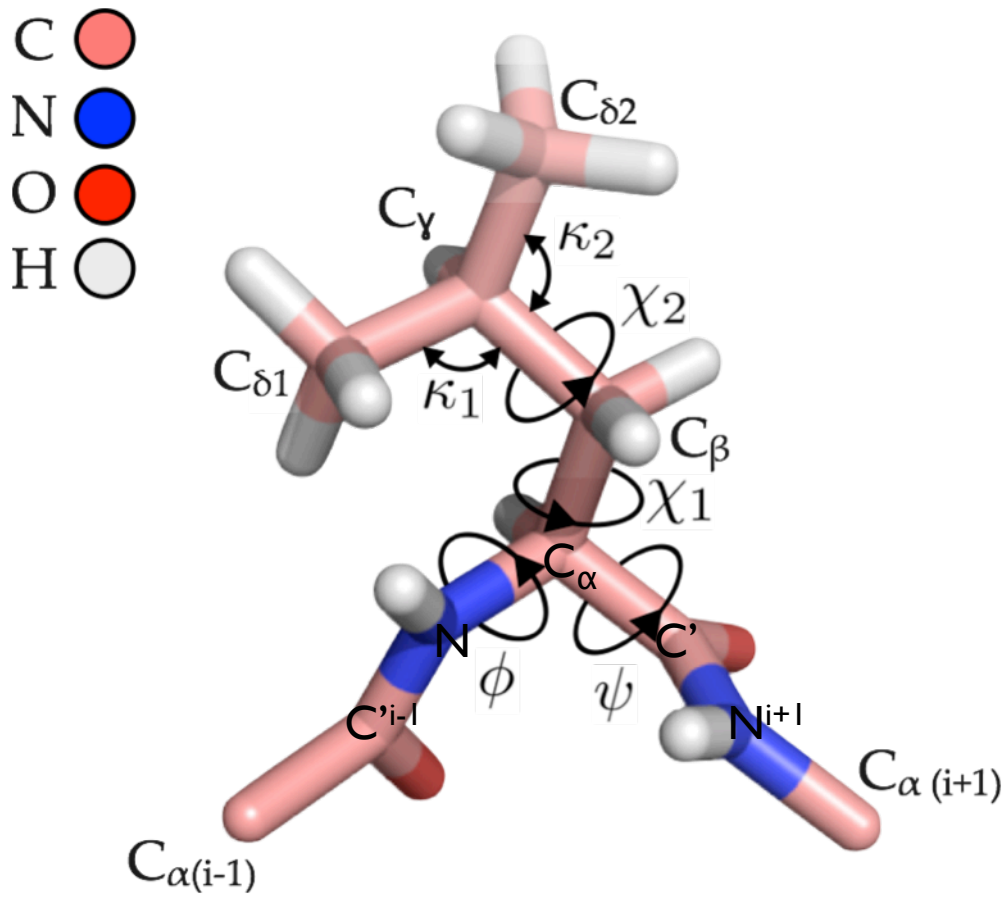


β-sheet

- Right-handed; three turns
- Vertical hydrogen bonds between NH_2 (teal/white) backbone group and C=O (grey/red) backbone group four residues earlier in sequence
- Side chains (R) on outside; point upwards toward NH_2
- Each amino acid corresponds to 100°, 1.5Å, 3.6 amino acids per turn
- $(\phi,\psi)=(-60°,-45°)$
- α-helix propensities: Met, Ala, Leu, Glu

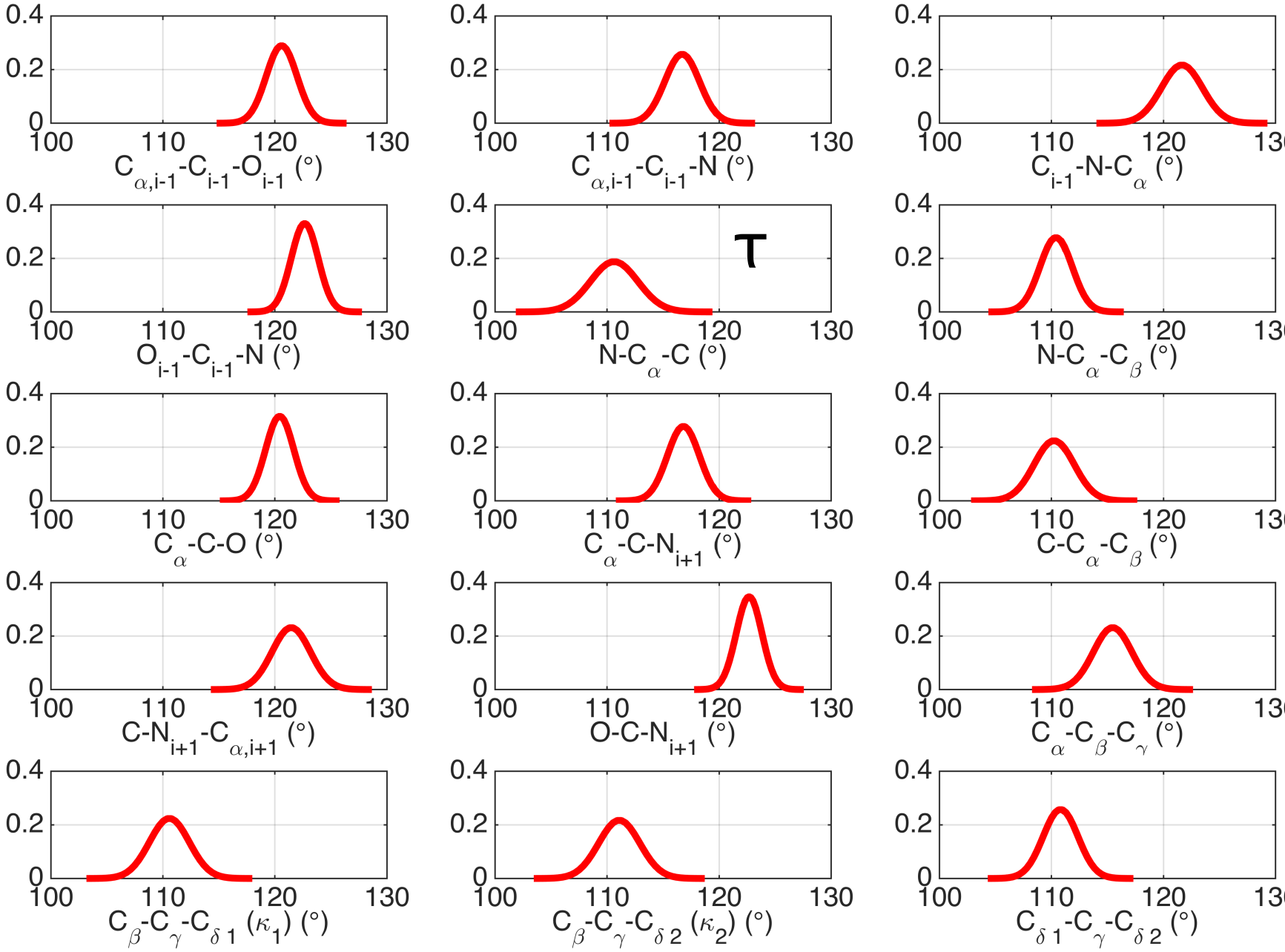
- 5-10 residues; peptide backbones fully extended
- NH (blue/white) of one strand hydrogen-bonded to C=O (black/red) of another strand
- C_α ,side chains (yellow) on adjacent strands aligned; side chains along single strand alternate up and down
- $(\phi,\psi)=(-135°,135°)$
- β-strand propensities: Val, Thr, Tyr, Trp, Phe, Ile

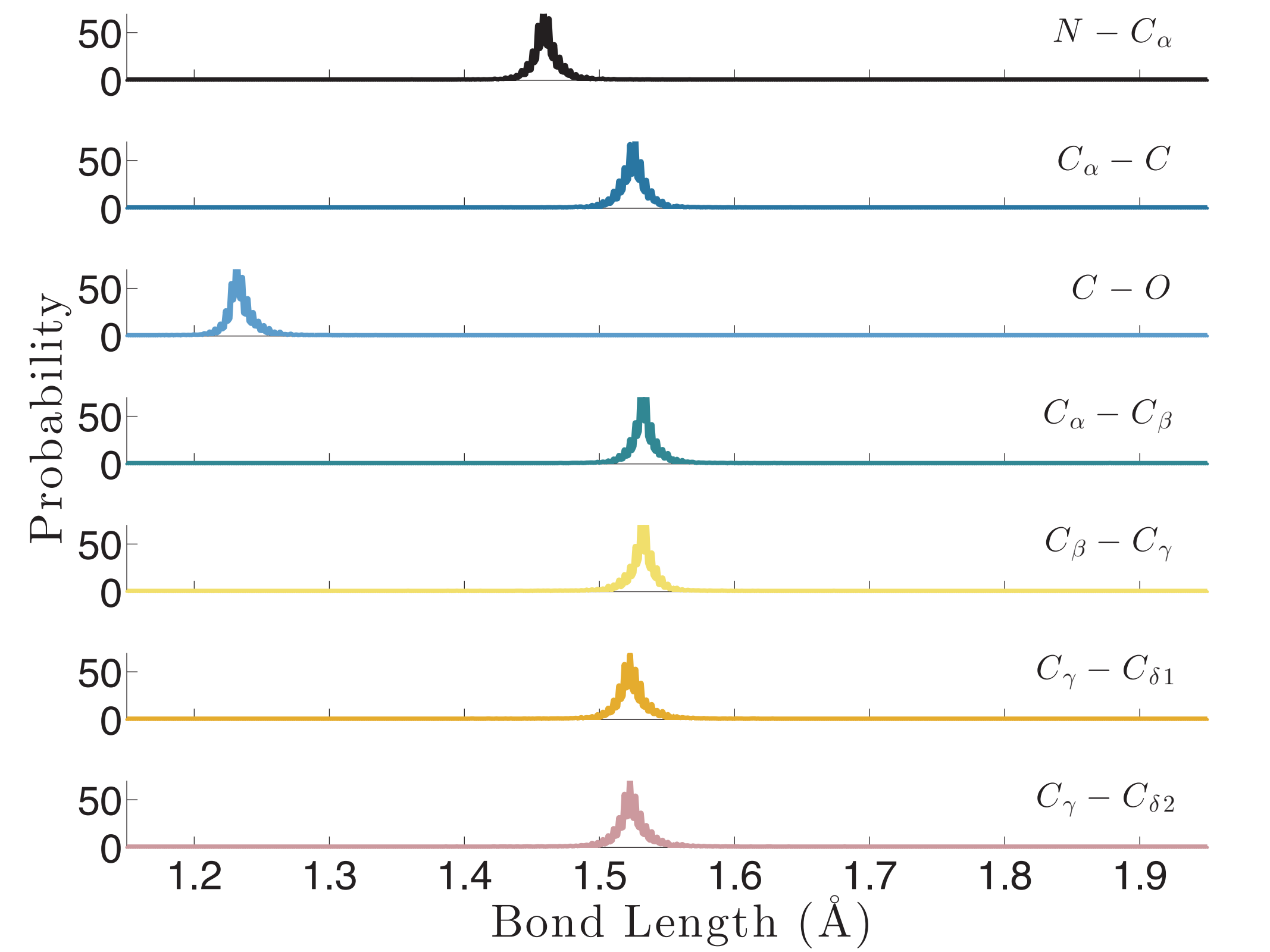
Bond Lengths and Angles

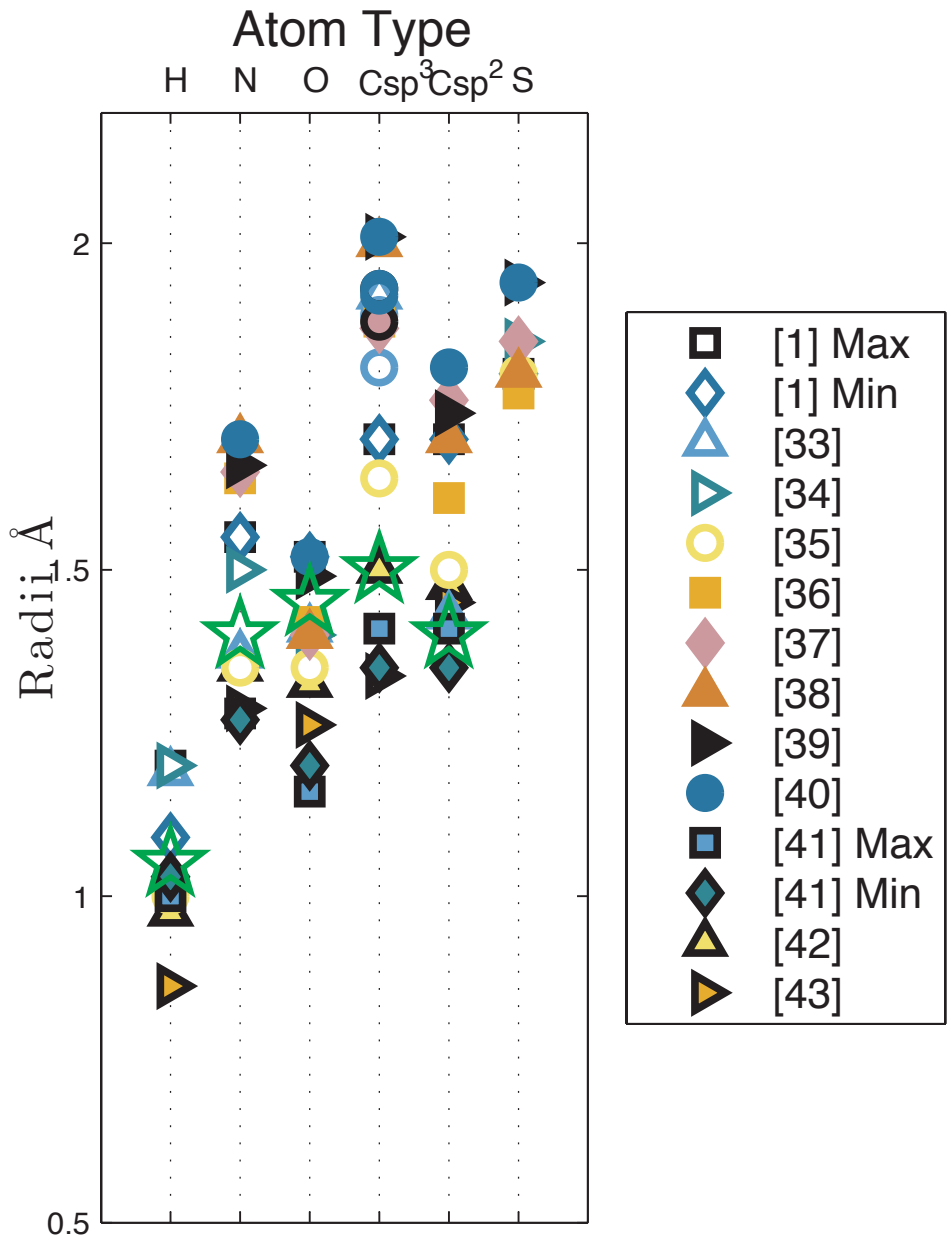


Leu dipeptide

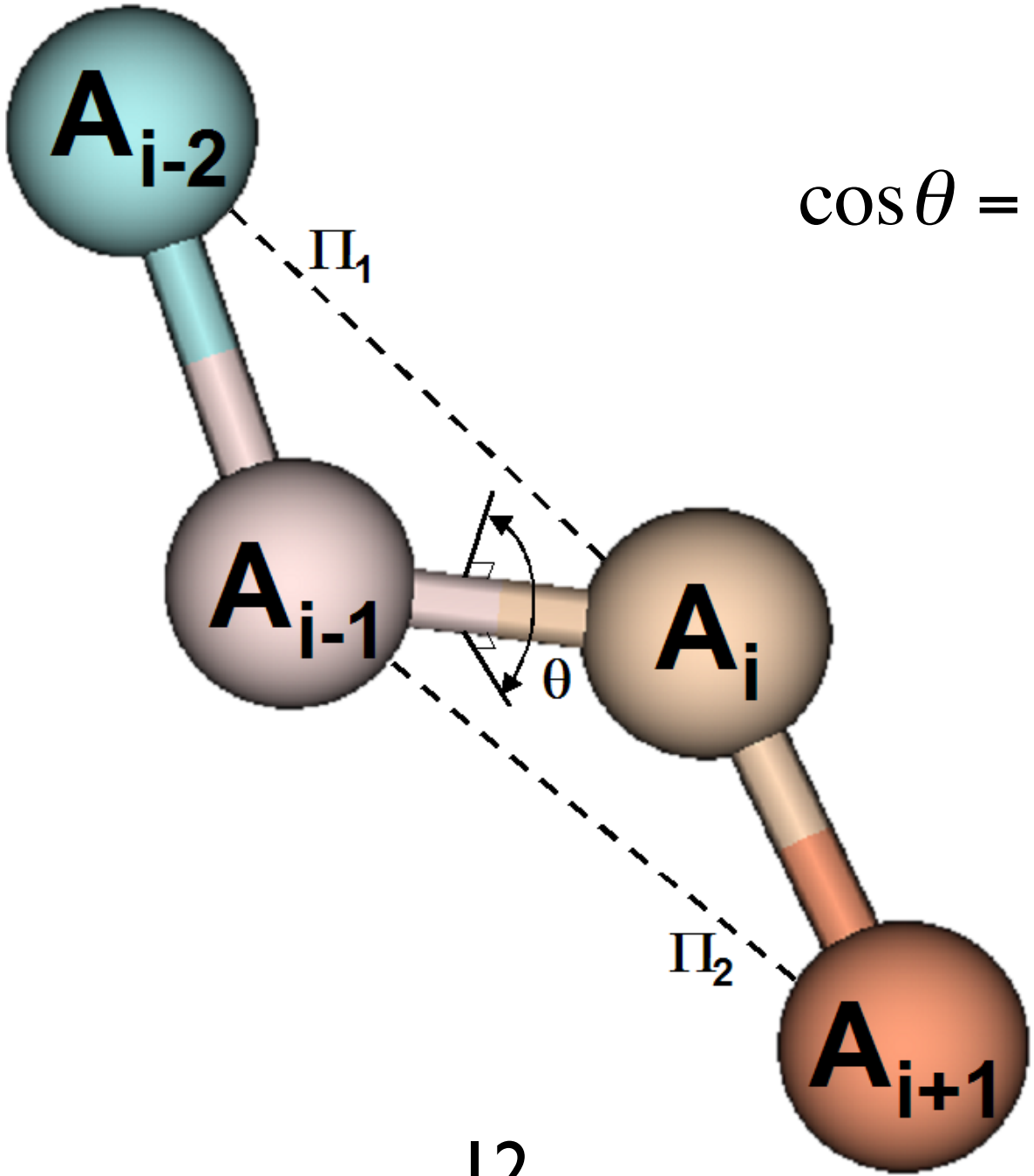
Leu


0.4
0.2
0
100
110
120
130
$C_{\alpha,i-1}$-C_{i-1}-O_{i-1} (°)
$C_{\alpha,i-1}$-C_{i-1}-N (°)
C_{i-1}-N-C_{α} (°)
O_{i-1}-C_{i-1}-N (°)
τ
N-C_{α}-C (°)
N-C_{α}-C_{β} (°)
C_{α}-C-O (°)
C_{α}-C-N_{i+1} (°)
C-C_{α}-C_{β} (°)
C-N_{i+1}-$C_{\alpha,i+1}$ (°)
O-C-N_{i+1} (°)
C_{α}-C_{β}-C_{γ} (°)
C_{β}-C_{γ}-$C_{\delta 1}$ (κ_1) (°)
C_{β}-C_{γ}-$C_{\delta 2}$ (κ_2) (°)
$C_{\delta 1}$-C_{γ}-$C_{\delta 2}$ (°)


Probability
50
0
$N - C_\alpha$
$C_\alpha - C$
$C - O$
$C_\alpha - C_\beta$
$C_\beta - C_\gamma$
$C_\gamma - C_{\delta 1}$
$C_\gamma - C_{\delta 2}$
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
Bond Length (Å)


Atom Type
H
N
O
Csp3
Csp2
S
Radii, Å
2
1.5
1
0.5
[1] Max
[1] Min
[33]
[34]
[35]
[36]
[37]
[38]
[39]
[40]
[41] Max
[41] Min
[42]
[43]

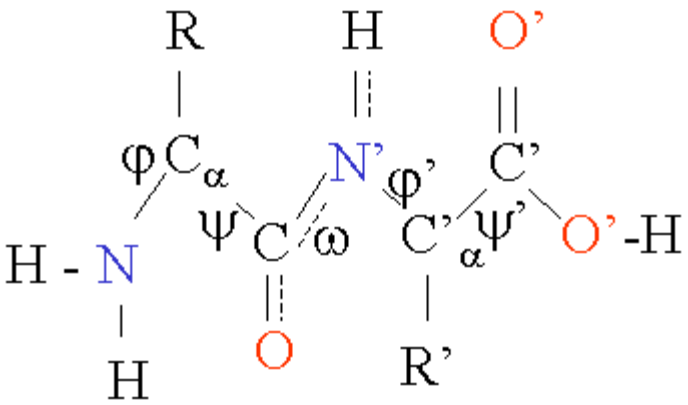
Backbonde Dihedral Angles



$$\cos\theta = \hat{\pi}_1 \bullet \hat{\pi}_2$$

Ramachandran Plot: Determining Steric Clashes

Backbone dihedral angles

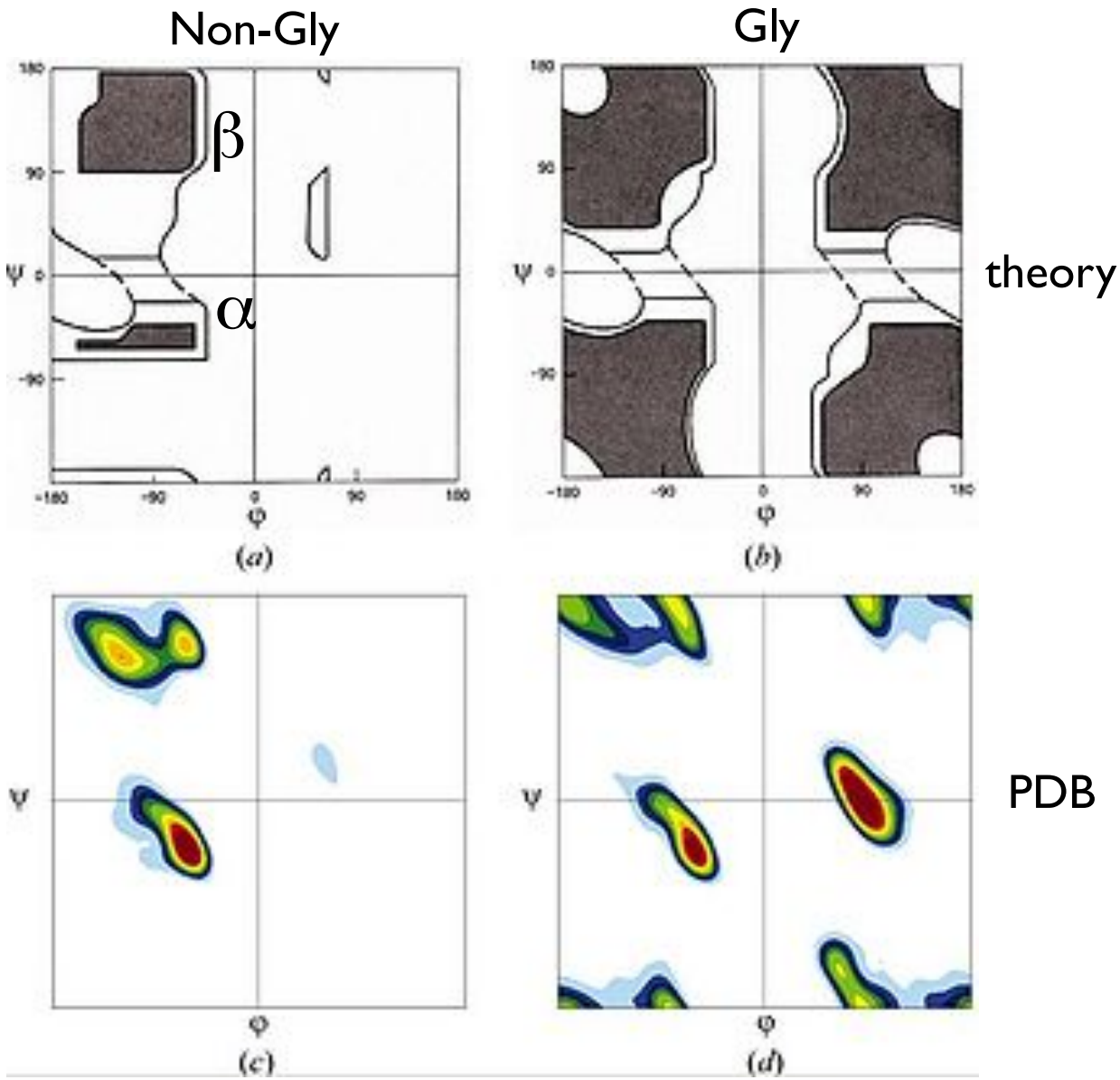


4 atoms define dihedral angle:

$C_\alpha CN'C'_\alpha$ ϕ

$C_\alpha CN'C'_\alpha$ ω=0,180°

$NC_\alpha CN'$ ψ

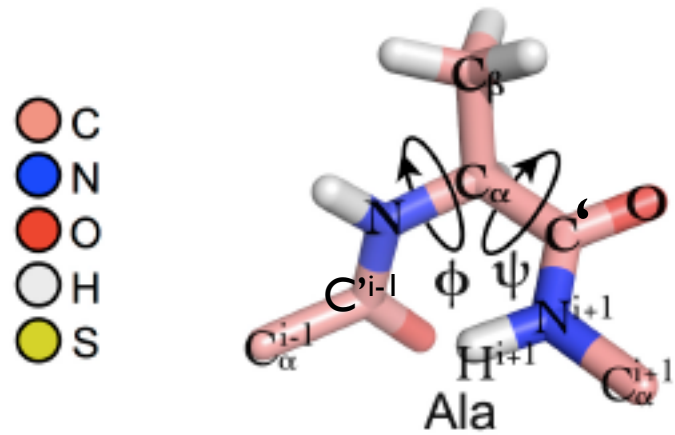


vdW radii

< vdW radii

backbone flexibility

G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, "Stereochemistry of polypeptide chain Configurations," J. Mol. Biol. 7 (1963) 95.



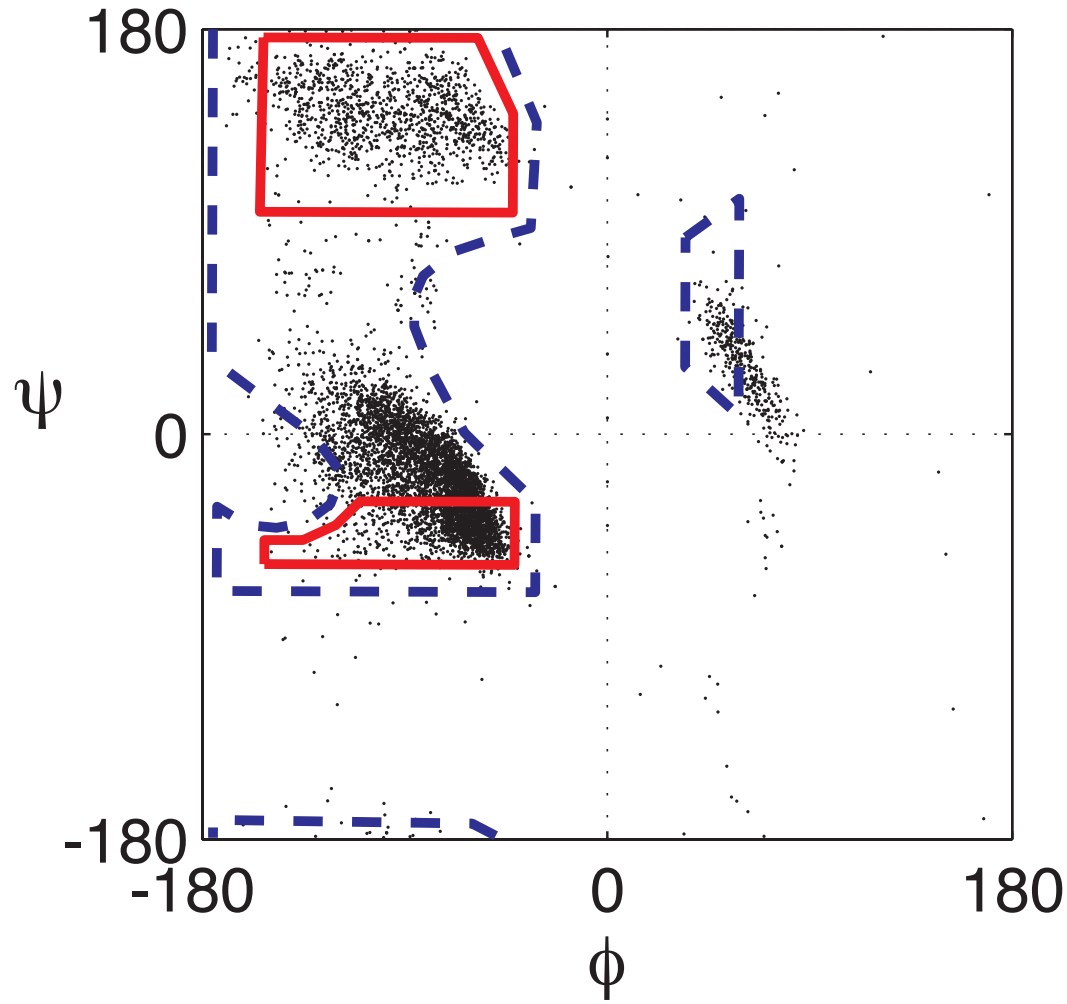
ϕ: $C'^{i-1}NC_\alpha C'$

ψ: $NC_\alpha C'N^{i+1}$

ω_1: $C^{i-1}{}_\alpha C'^{i-1}NC_\alpha$

ω_2: $C_\alpha C'N^{i+1}C^{i+1}{}_\alpha$

Backbone dihedral angles from PDB



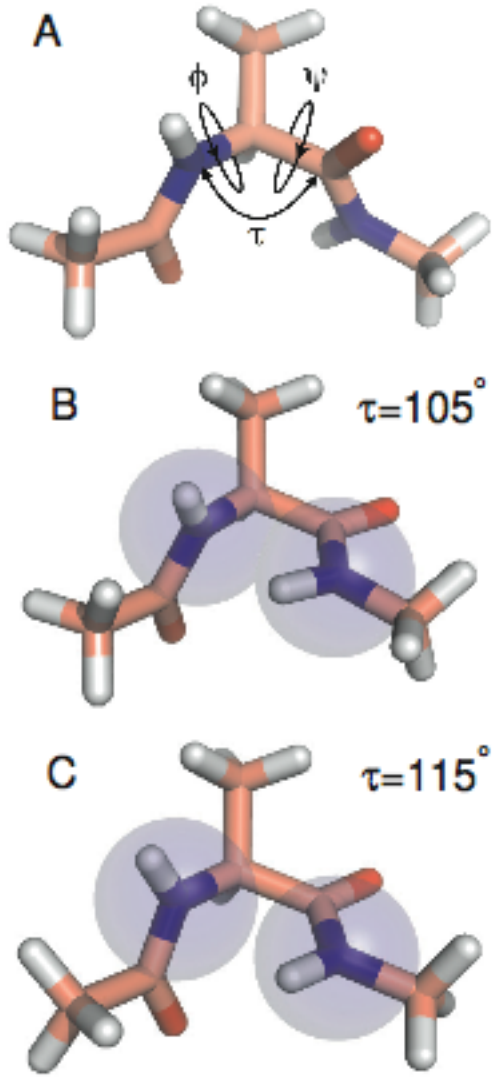


Figure 1: Stick representation of an alanyl dipeptide mimetic. Atom types are color-coded: carbon=pink, nitrogen=blue, oxygen=red, hydrogen=white. **A**: The backbone dihedral angles ϕ and ψ and the bond angle τ are indicated. **B**: $\tau = 105°$, $\phi = -90°$, $\psi = 0°$ (i.e. bridge region values of ϕ and ψ). Blue-shaded spheres indicate steric overlap between main-chain nitrogens for this value of τ. **C**: $\tau = 115°$, $\phi = -90°$, $\psi = 0°$ (i.e. bridge region values of ϕ and ψ). Blue-shaded spheres indicate no steric overlap between main-chain nitrogens for this value of τ.

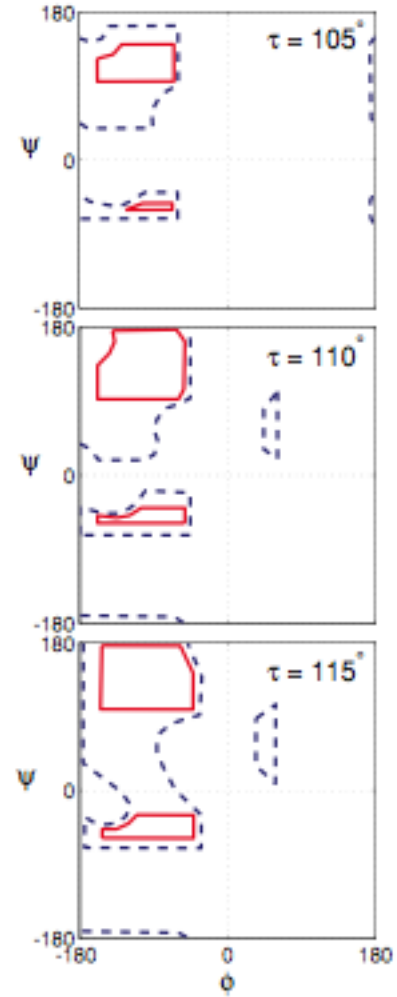
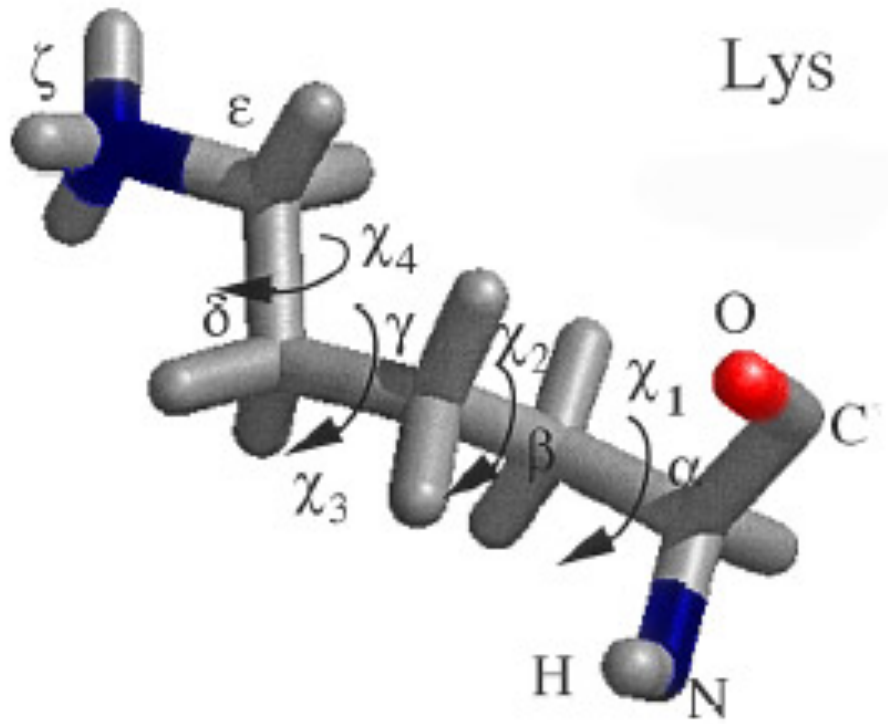


Figure 2: Ramachandran plots of allowed ϕ/ψ combinations for 3 values of τ [2]. The solid red lines enclose the 'normally allowed' ϕ/ψ combinations and the dashed blue line indicates the 'outer limit'.

Side-Chain Dihedral Angles

χ_4: Lys, Arg

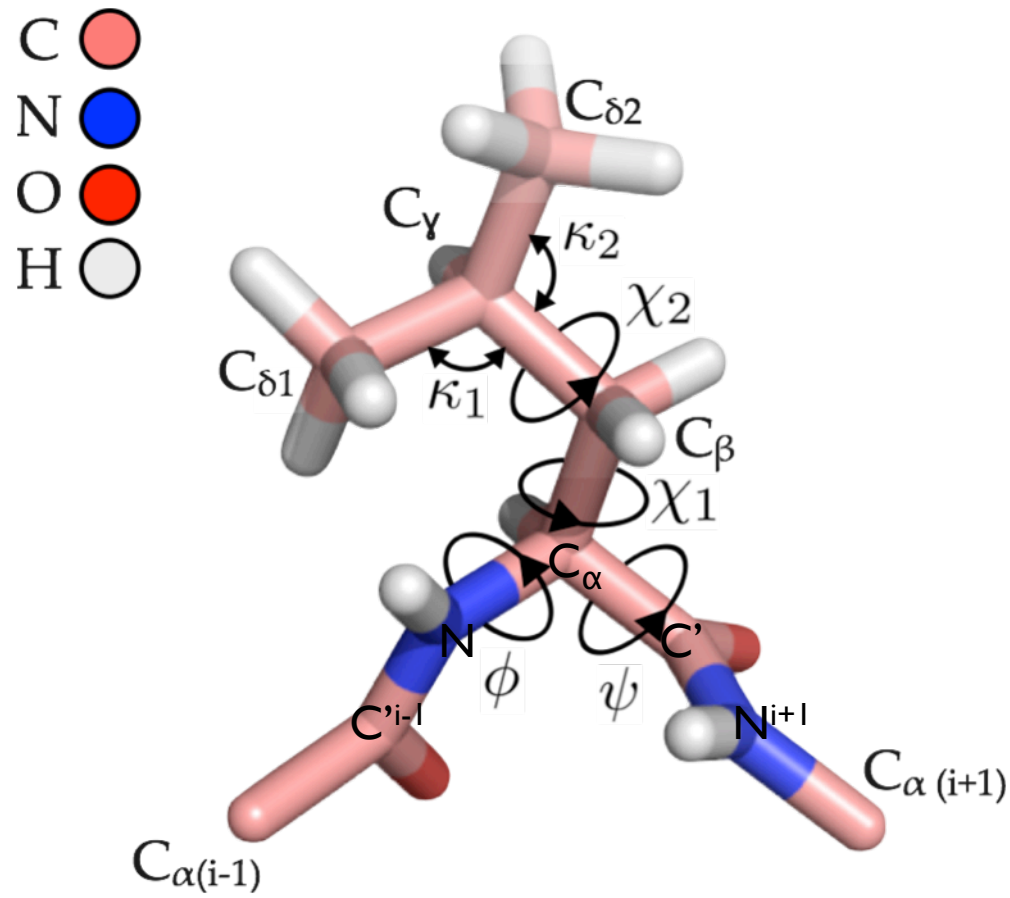
χ_5: Arg



Side chain: C_α-CH_2-CH_2-CH_2-CH_2-NH_3

Use $NC_\alpha C_\beta C_\gamma C_\delta C_\varepsilon N_\zeta$ to define χ_1, χ_2, χ_3, χ_4

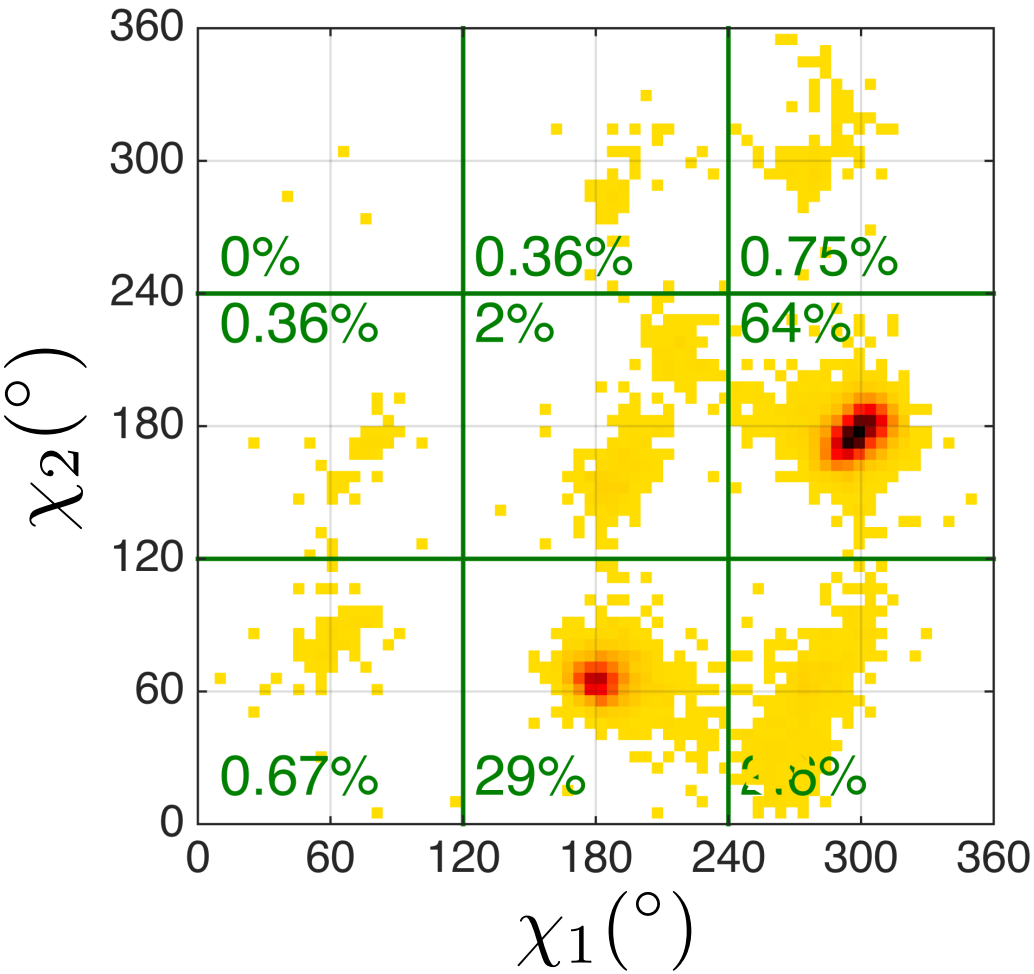
Side Chain Dihedral Angles



Leu dipeptide

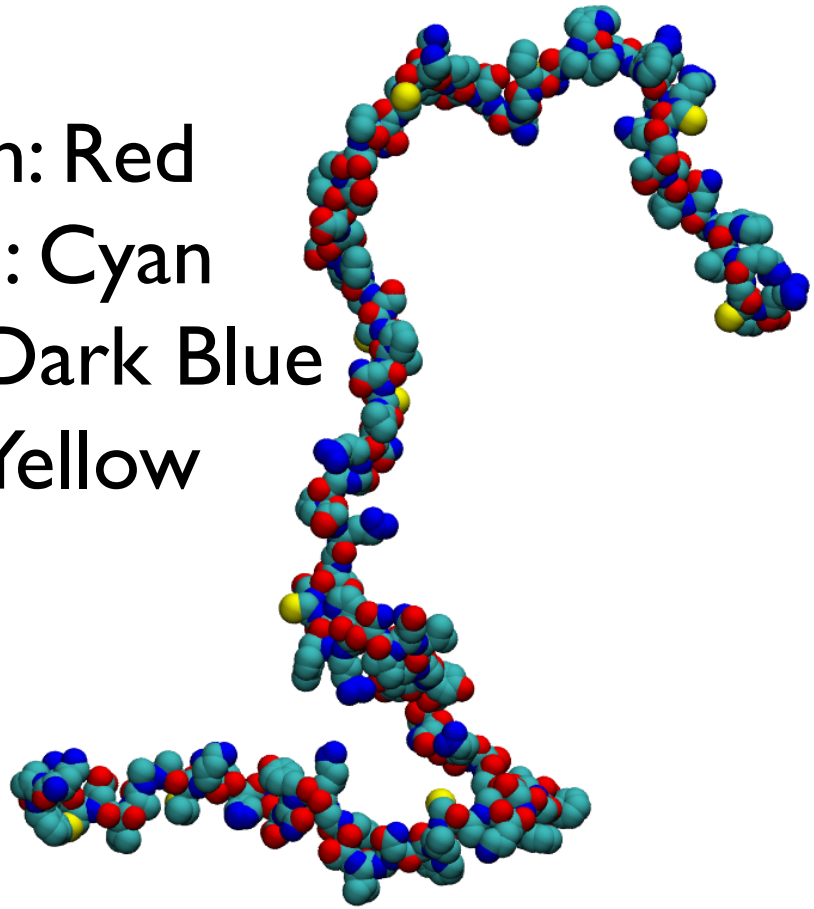
Sidechain Dihedral Angle Distributions for Leu

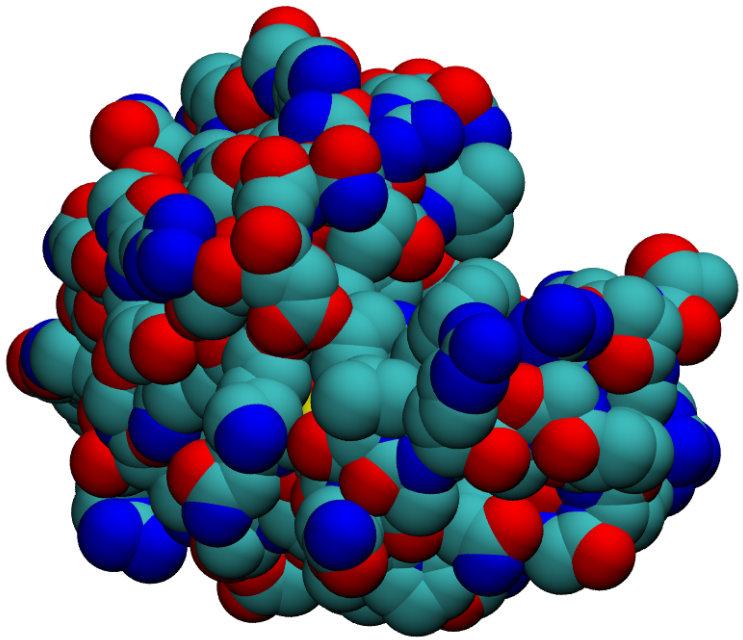
PDB



Folding Transition

Oxygen: Red
Carbon: Cyan
Nitrogen: Dark Blue
Sulfur: Yellow





$T > T_m$

$T < T_m$

Why do proteins fold (correctly & rapidly)??

Levinthal's paradox:

For a protein with N amino acids, number of backbone conformations/minima

$$N_c \sim \mu^{2N}$$

μ = # allowed dihedral angles

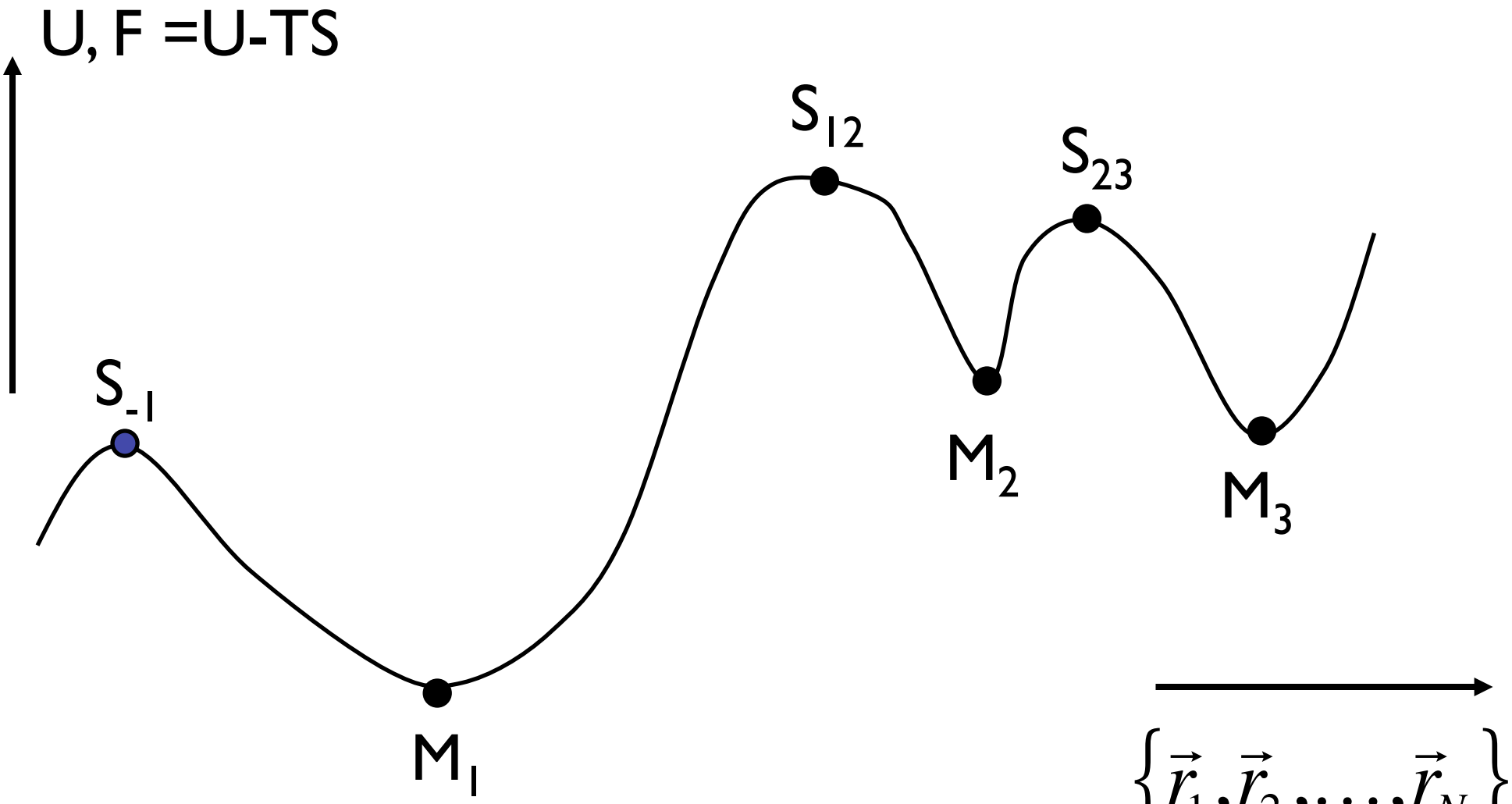
How does a protein find the global optimum w/o global search? Proteins fold much faster.

$N_c \sim 3^{200} \sim 10^{95}$

$\tau_{fold} \sim N_c \, \tau_{sample} \sim 10^{83}$ s vs $\tau_{fold} \sim 10^{-6}$-10^{-3} s

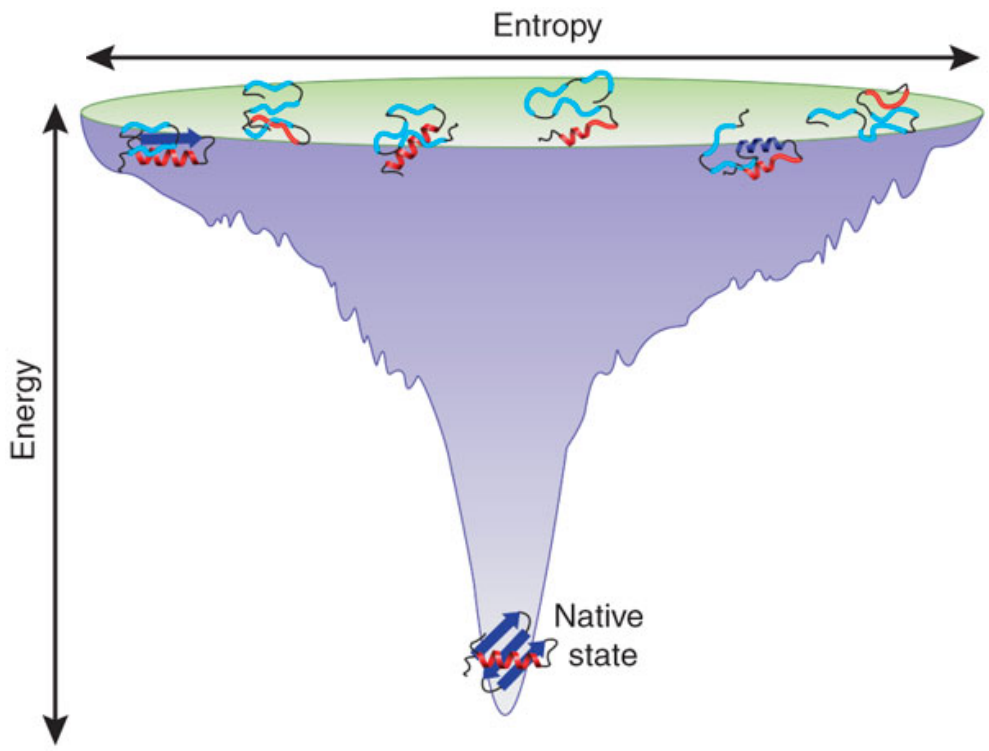
$\tau_{universe} \sim 10^{17}$ s

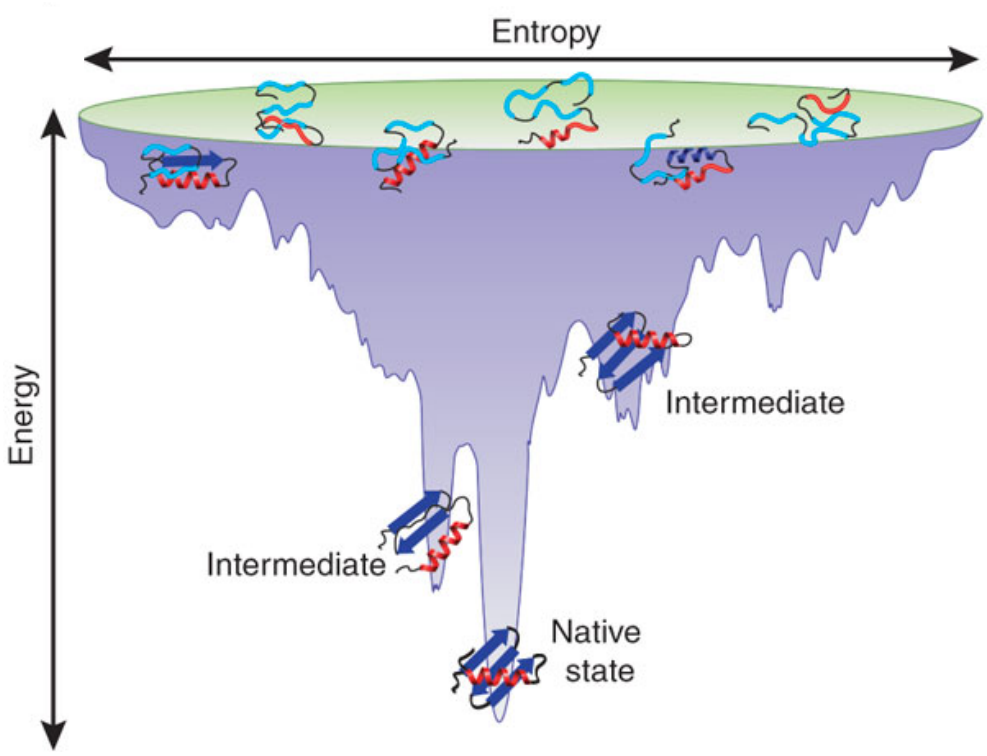
Energy Landscape



$$\vec{\nabla} U = 0 \quad \begin{cases} \nabla^2 U > 0 & \text{minimum} \\ \nabla^2 U = 0 & \text{saddle point} \\ \nabla^2 U < 0 & \text{maximum} \end{cases}$$

Roughness of Energy Landscape





smooth, funneled

rough

(Wolynes et. al. 1997)