Molecular Modeling 2018

Midterm review slides

Lecture 1 slide 24

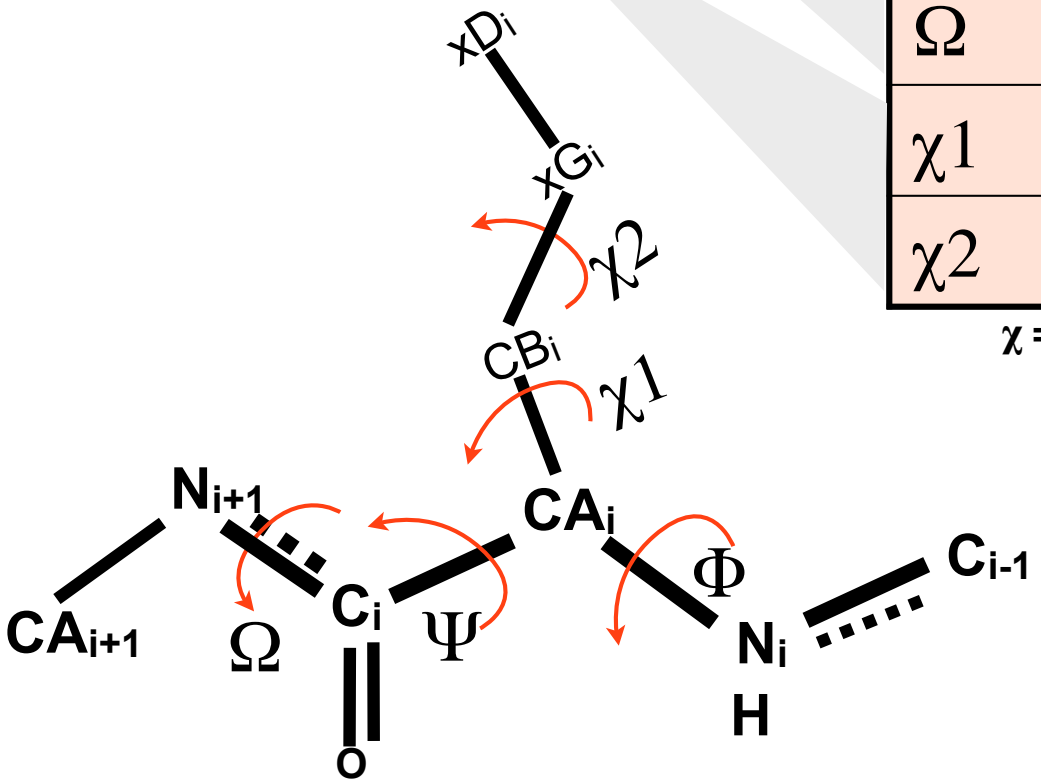
Torsion angles

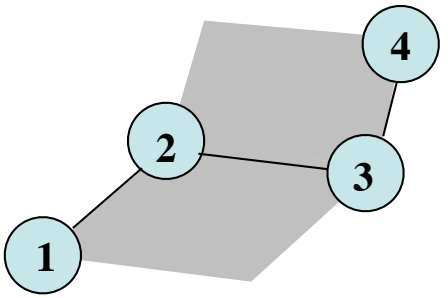
Protein flexibility is due to rotations around single bonds, backbone and side chain.

4 atoms define two planes

Angle	atom1	atom2	atom3	atom4
Φ	C_{i-1}	N_i	CA_i	C_i
Ψ	N_i	CA_i	C_i	N_{i+1}
Ω	CA_i	C_i	N_{i+1}	CA_{i+1}
χ1	C_i	CA_i	CB_i	xG_i
χ2	CA_i	CB_i	xG_i	xD_i

χ = chi



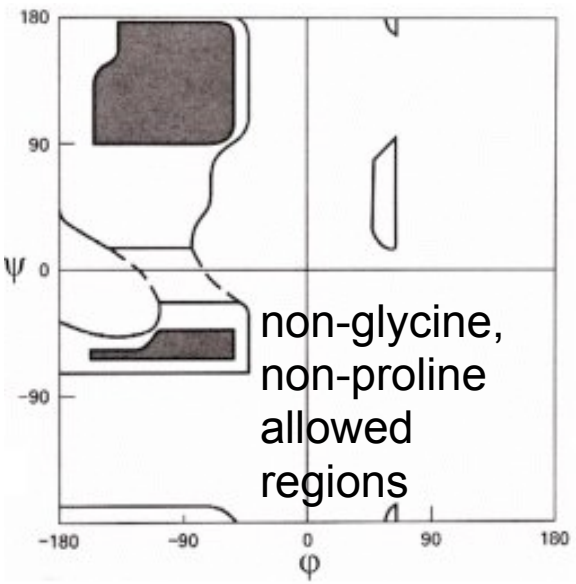


Lecture 1 slide 26

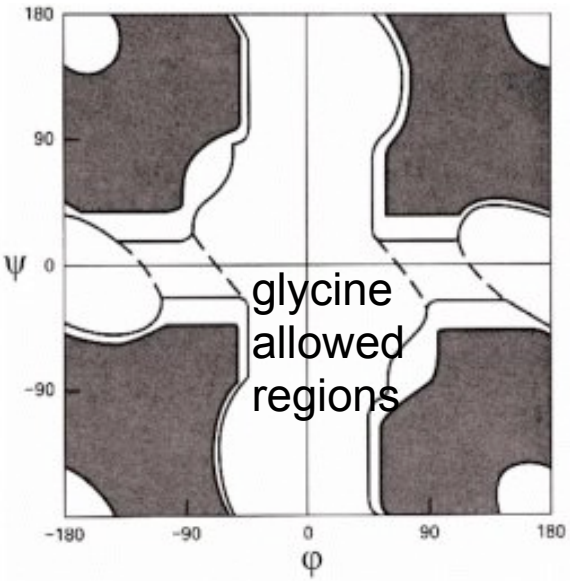
Ramachandran Plot maps allowable phi, psi regions

Ramachandran used a physical model of dipeptides to determine the allowed (dark) and disallowed (white)combinations of phi and psi backbone angles.
The observed frequencies roughly agree with R's allowed regions.

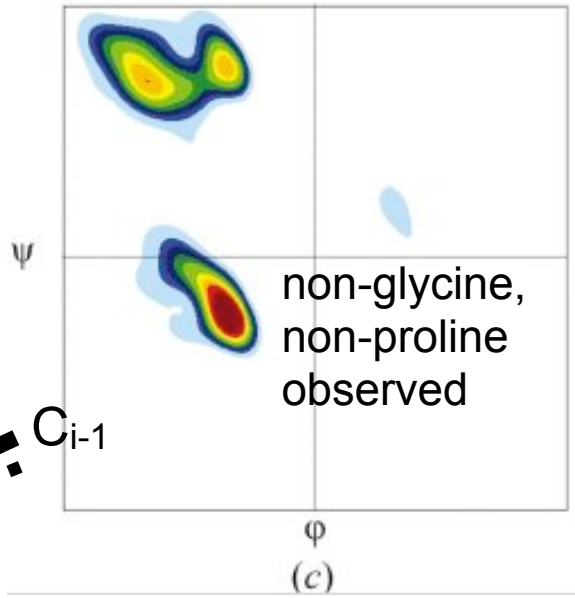
Ramachandran & Sasisekharan (1968)



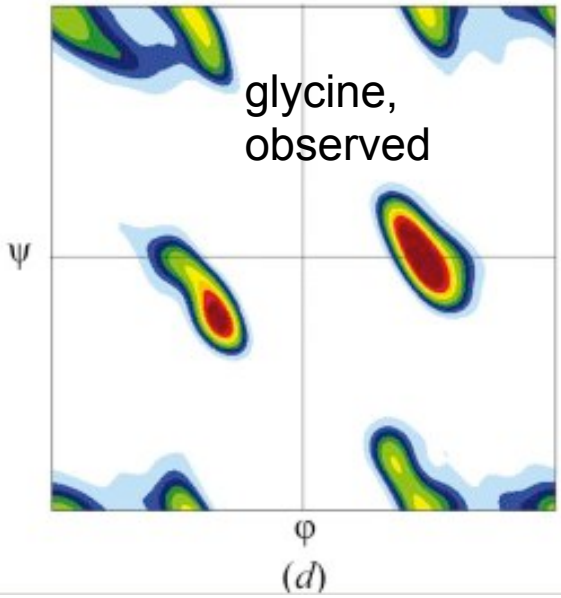
(a)



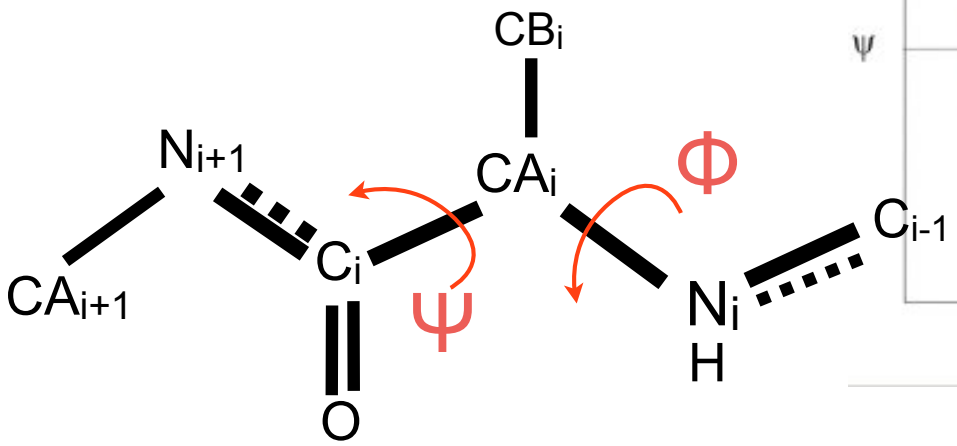
(b)



(c)



(d)



Lecture 2 slide 12

Structure quality: resolution

- Resolution = *d* in Bragg's Law. $n\lambda=2d\sin\theta$. Lower *d* is higher resolution.
- "Resolution" = resolution limit = the lowest *d* observed = the highest scattering angle observed.

Resolution	**quality**
> 4Å	nearly worthless, shows blobs of density
3-4Å	medium. Shows backbone and some sidechains.
2-3Å	typical good structure, all sidechains visible
1.5-2Å	high resolution. Atom positions known within 0.1Å rmsd.
< 1.5Å	ultra high resolution! Hydrogens sometimes visible.

Lecture 3 slide 6

SCOP fold jargon

example: α/β proteins: flavodoxin-like

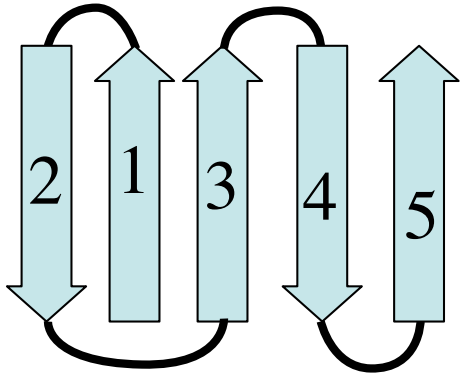
SCOP Description: 3 layers, α/β/α; parallel beta-sheet of 5 strand, order 21345

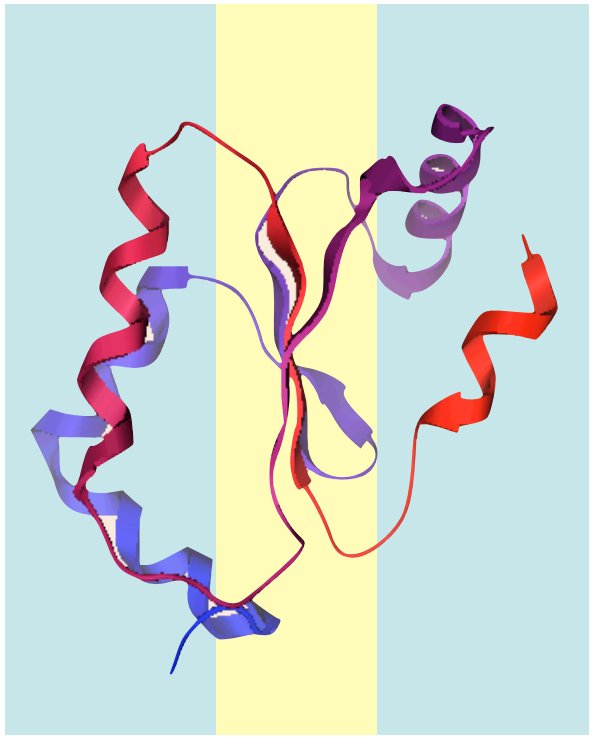
Note the term: "*layers*"

Rough arrangements of secondary structure elements.

Note the term: "*order*"

The sequential order of beta strands in a beta sheet.





α layer

α layer

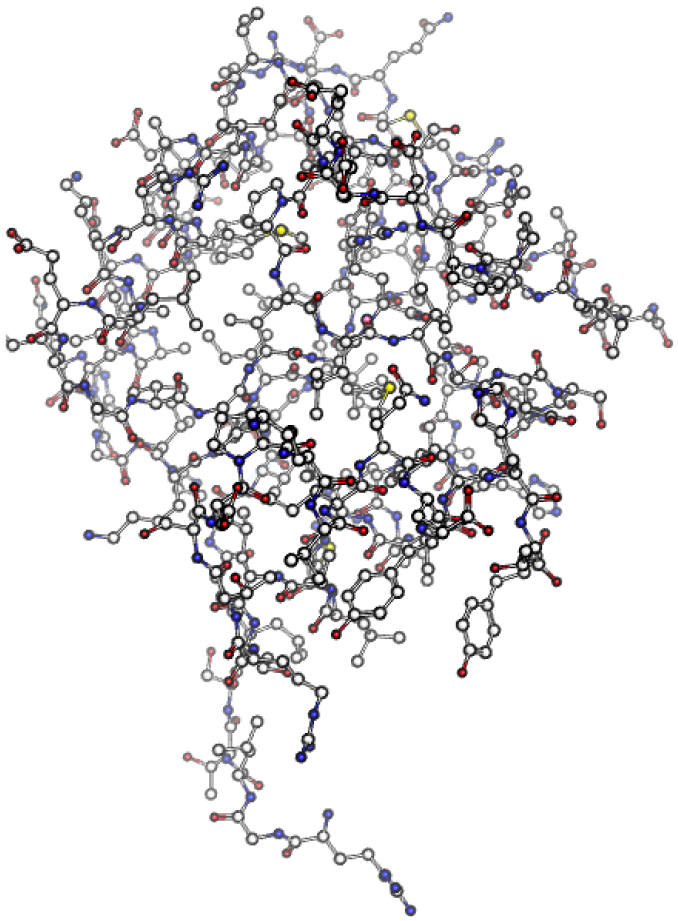
β layer

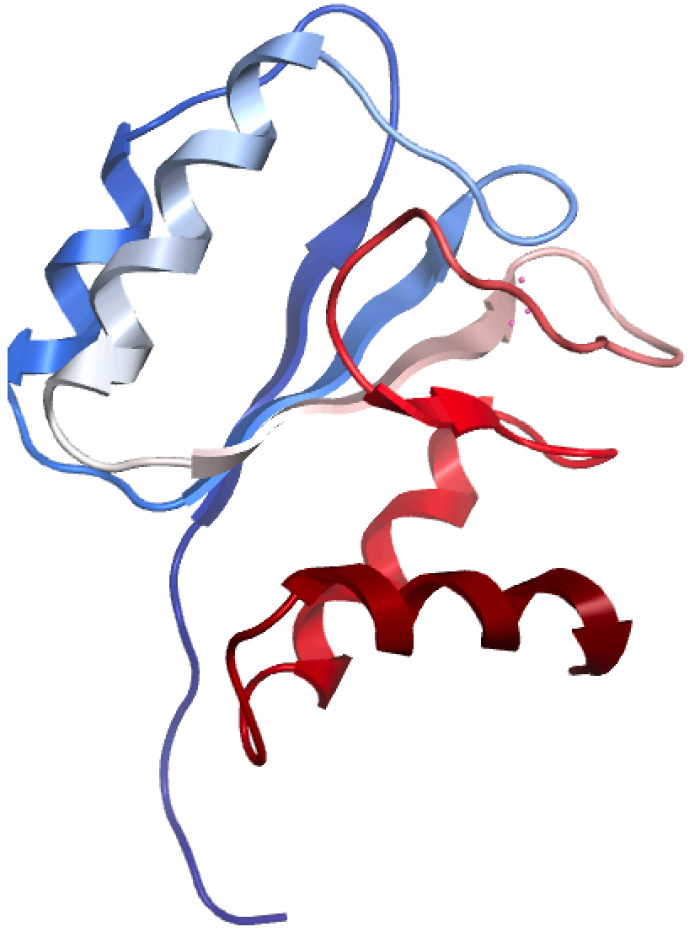
Lecture 3 slide 12

How to draw TOPS

On course website, find the link "**TOPS practice**" (tops_practice.moe)

Save it. Open it in moe.



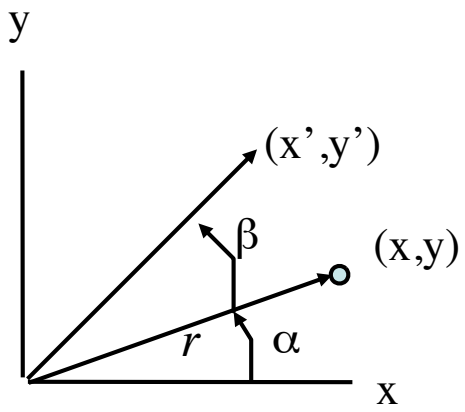


Lecture 4 slide 7

A rotation matrix

$$x = r\cos\alpha$$

$$y = r\sin\alpha$$



$$x' = r\cos(\alpha+\beta)$$
$$= r(\cos\alpha\cos\beta - \sin\alpha\sin\beta)$$
$$= (r\cos\alpha)\cos\beta - (r\sin\alpha)\sin\beta$$
$$= x\cos\beta - y\sin\beta$$

$$y' = r\sin(\alpha+\beta)$$
$$= r(\sin\alpha\cos\beta + \sin\beta\cos\alpha)$$
$$= (r\sin\alpha)\cos\beta + (r\cos\alpha)\sin\beta$$
$$= y\cos\beta + x\sin\beta$$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} r\cos\alpha \\ r\sin\alpha \end{pmatrix} = \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

rotation matrix is the same for any r, any α.

Lecture 4 slide 17

RMSD

Root Mean Square Deviation in superimposed coordinates is the standard measure of **structural difference**.

$$\sqrt{\frac{\sum_{i=1,N}(\vec{v1}_i - \vec{v2}_i)^2}{N}}$$

Where $\vec{v1}_i$ and $\vec{v2}_i$ are the ***equivalent**** ***coordinates*** from molecules 1 and 2, respectively.

*Equivalent as defined by an **alignment**.

Chicken/Egg

- Least squares superposition defines the alignment.
- The alignment defines the least squares superposition.

Lecture 5 slide 7

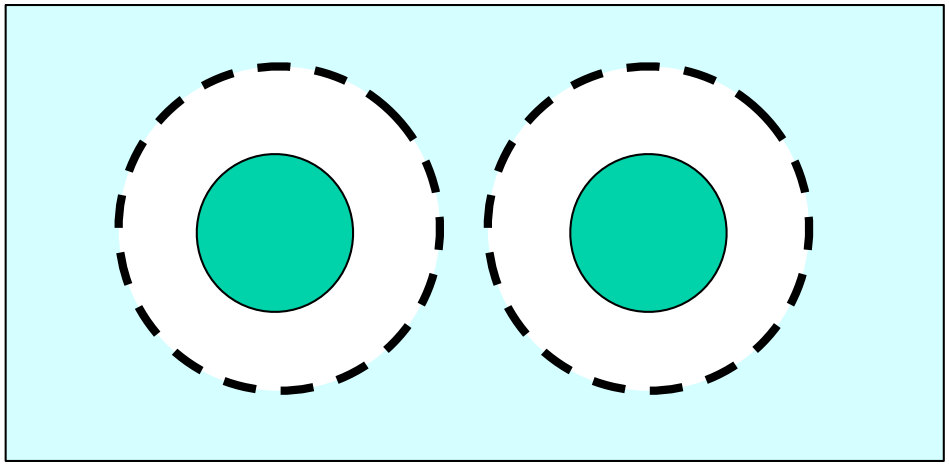
What is energy?

- **Energy** (G) is a measure of the **probability** of the **state of the system**. Energy is the **negative log of the probability ratio, times temperature**.
- $\Delta G = -RT \ln (A / \text{not } A)$
  or $-RT \ln(P / (1-P))$, where P = probability.
- The system = <u>the atoms</u>.
- State = <u>where the atoms are</u>.
  (This is a vague definition so we can be flexible about what the energy means.)
- Energy is always relative.
- Energy is measured between two states.
- Energy is expressed in J/mole, or kJ/mole.
- Energy breaks down into enthalpy (H) and entropy (S).
  $\Delta G = \Delta H - T\Delta S$.
- Energy also breaks down to **potential** energy and **kinetic** energy.

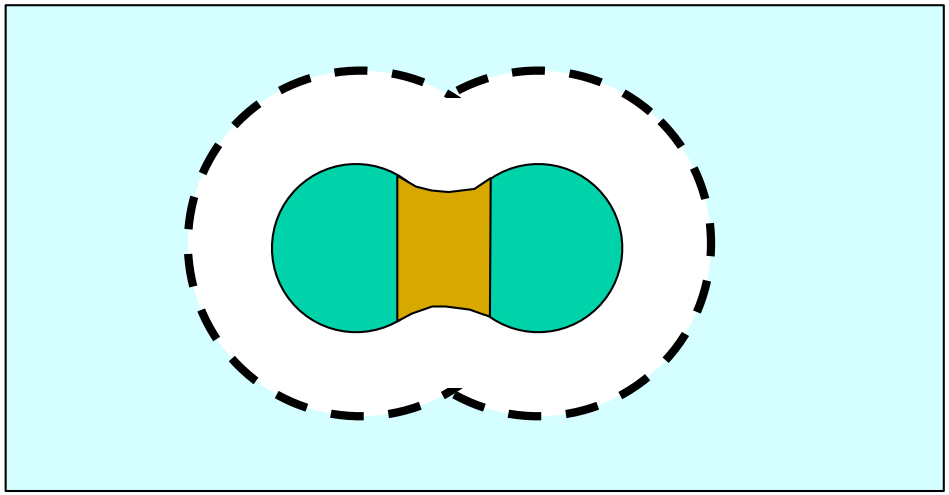
Lecture 5 slide 20

The Hydrophobic Effect

Solvent accessible surface (dashed line) around non-polar atoms contains "high energy waters" because those waters lose H-bonds.



Non-polar atoms come together because it decreases the number of high energy waters. (Even at the cost of creating void space (brown).



Lecture 6 slide 16

The rule: similar sequence means similar structure

sequences that…

…have a common ancestor

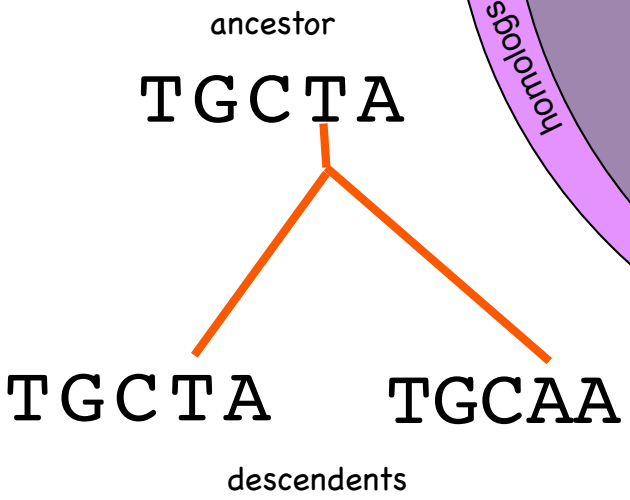
…superimpose in space

..Venn diagram..

most homologs are superposable

homologs that don't superpose

results of convergent evolution

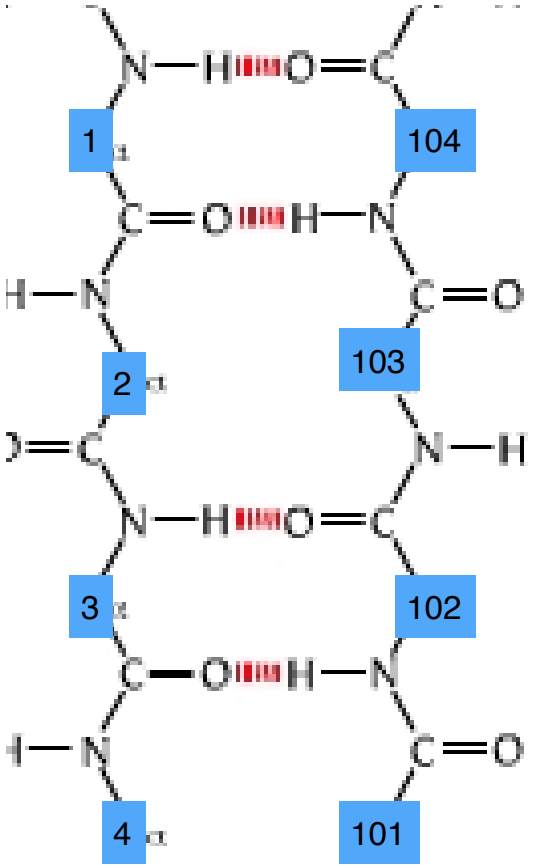


Lecture 6 slide 10

Secondary structure using matrices: antiparallel sheet

0	1	0
1	0	0

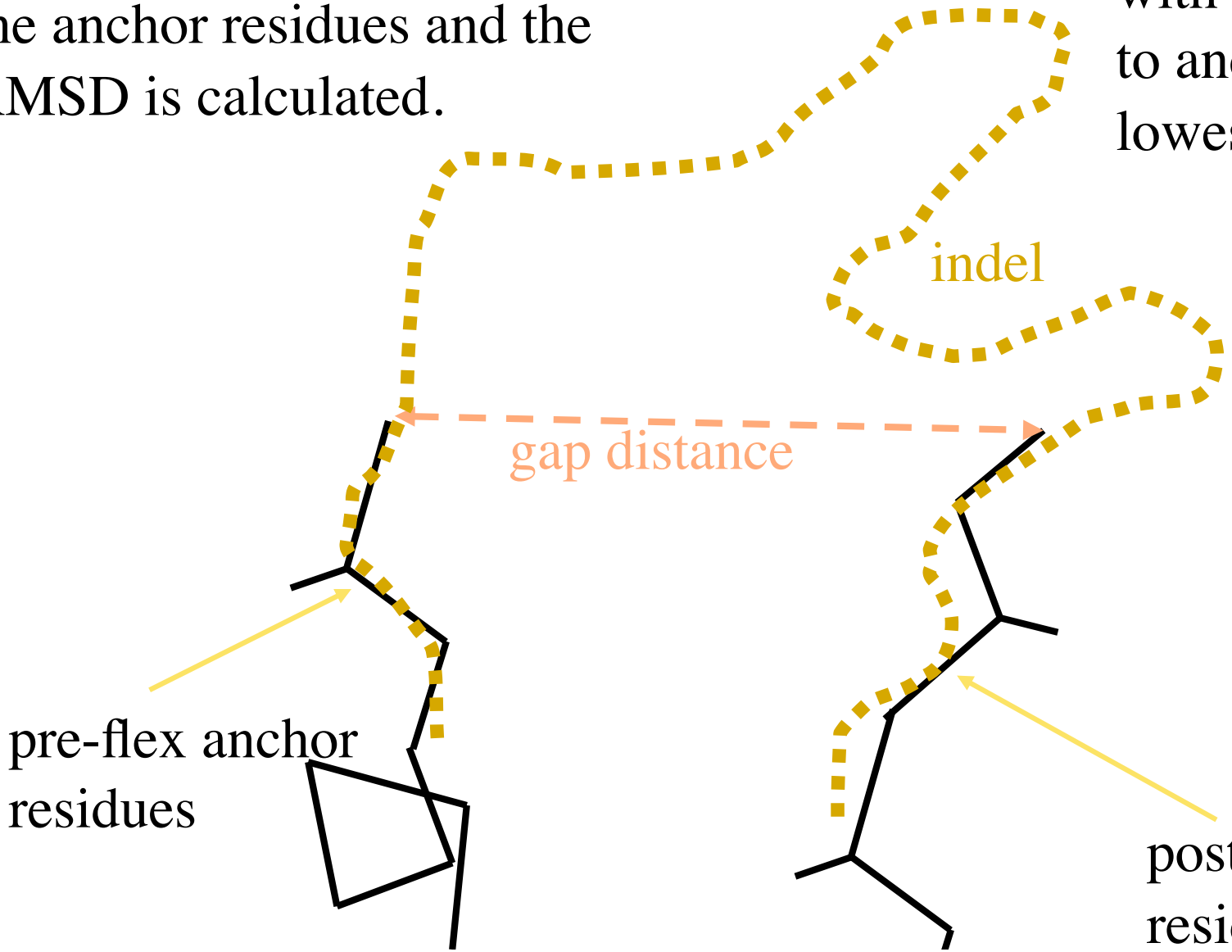
0	1	-2
1	0	+2



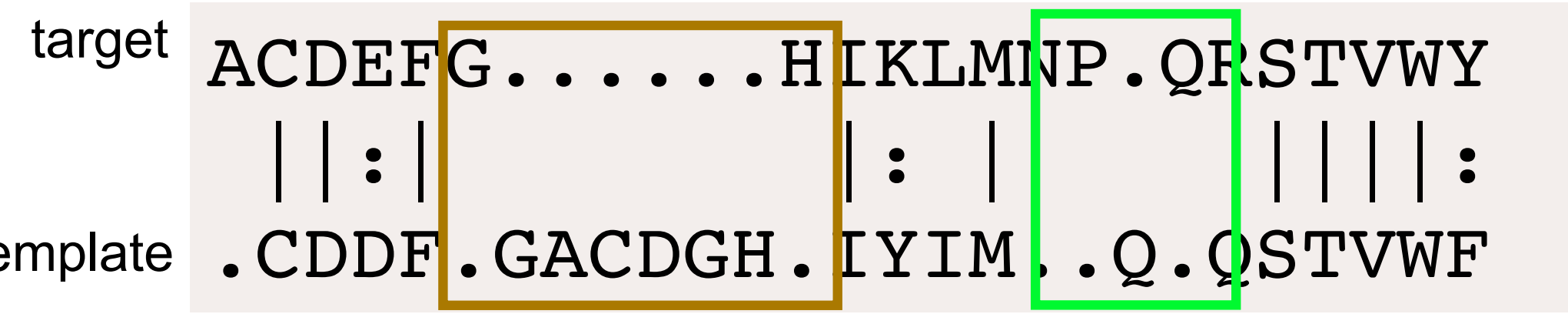
Automated Loop Search

Loops of the right length in the database are superimposed on the anchor residues and the RMSD is calculated.

MOE keeps the loops with the best RMSDs to anchors, and lowest energy.



Telling MOE how to anchor a better loop search



Align F to F, I to I, delete GACDGH and add 2-residue loop GH from a loop search.

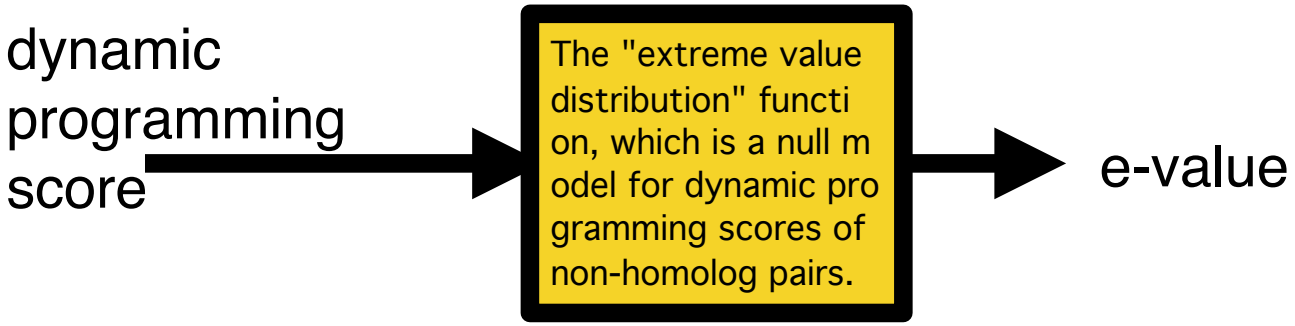
Align M to M, R to $Q_{(2)}$, delete Q and add a 3-residue loop NPQ from a loop search.

2-for-6
instead of 0 for 4.

3-for-1
instead of 2 for 0.

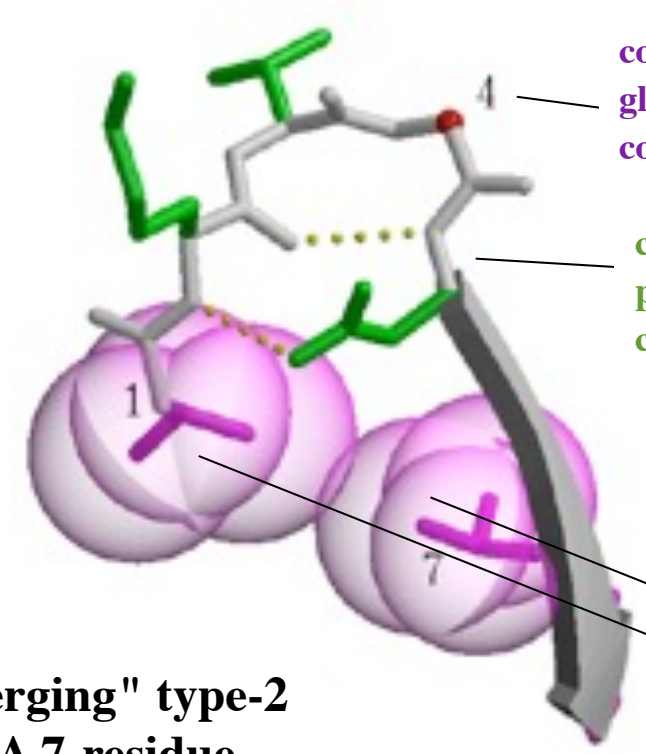
e-value

- The number of times in a database search that you will get a random, non-homologous hit with the same score or better.



http://www.bioinfo.rpi.edu/bystrc/Isites2/

Biophysics of an I-sites motif

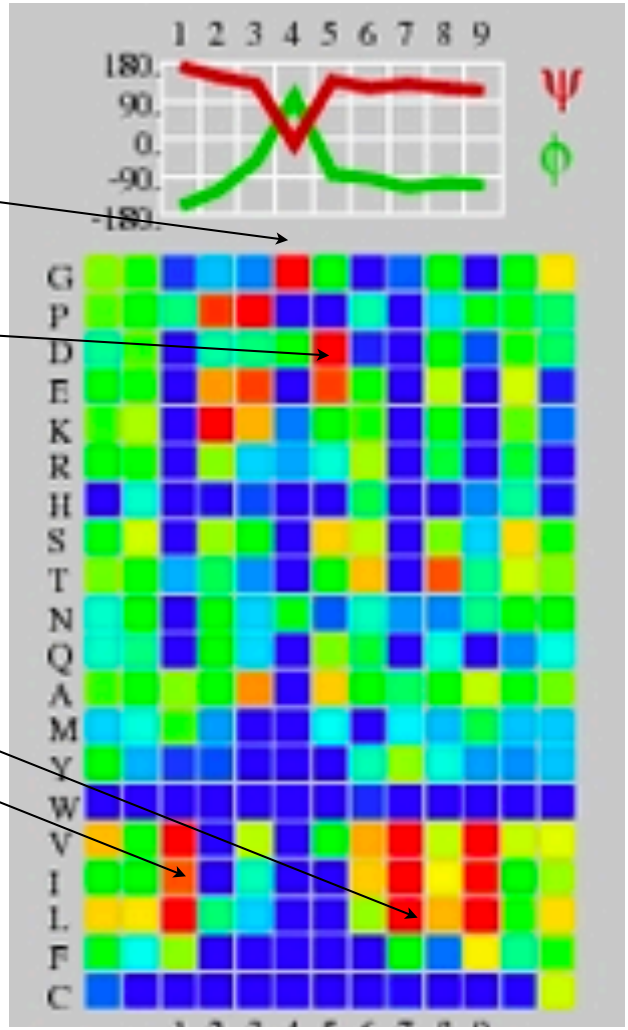


"Diverging" type-2 turn. A 7-residue peptide forms this structure! [2]

conserved glycine in α_L conformation

conserved polar side chain

conserved non-polar side chains



conserved motif backbone angles phi, psi

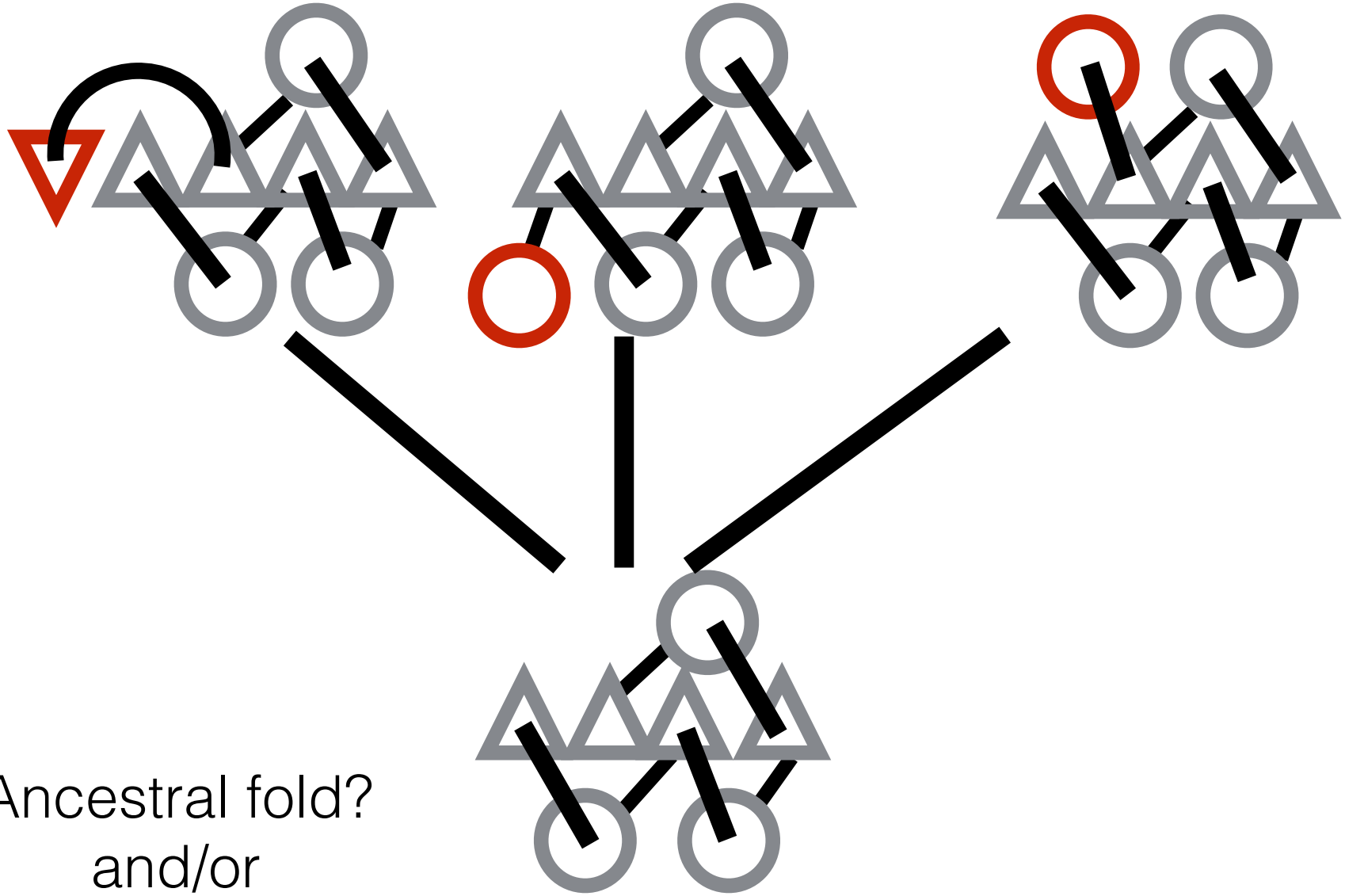
red= >1 log unit more likely than chance

blue= >1 log unit less likely than chance

[1]Bystroff C & Baker D. (1998). Prediction of local structure in proteins using a library of sequence-structure motifs. *J Mol Biol* 281, 565-77.

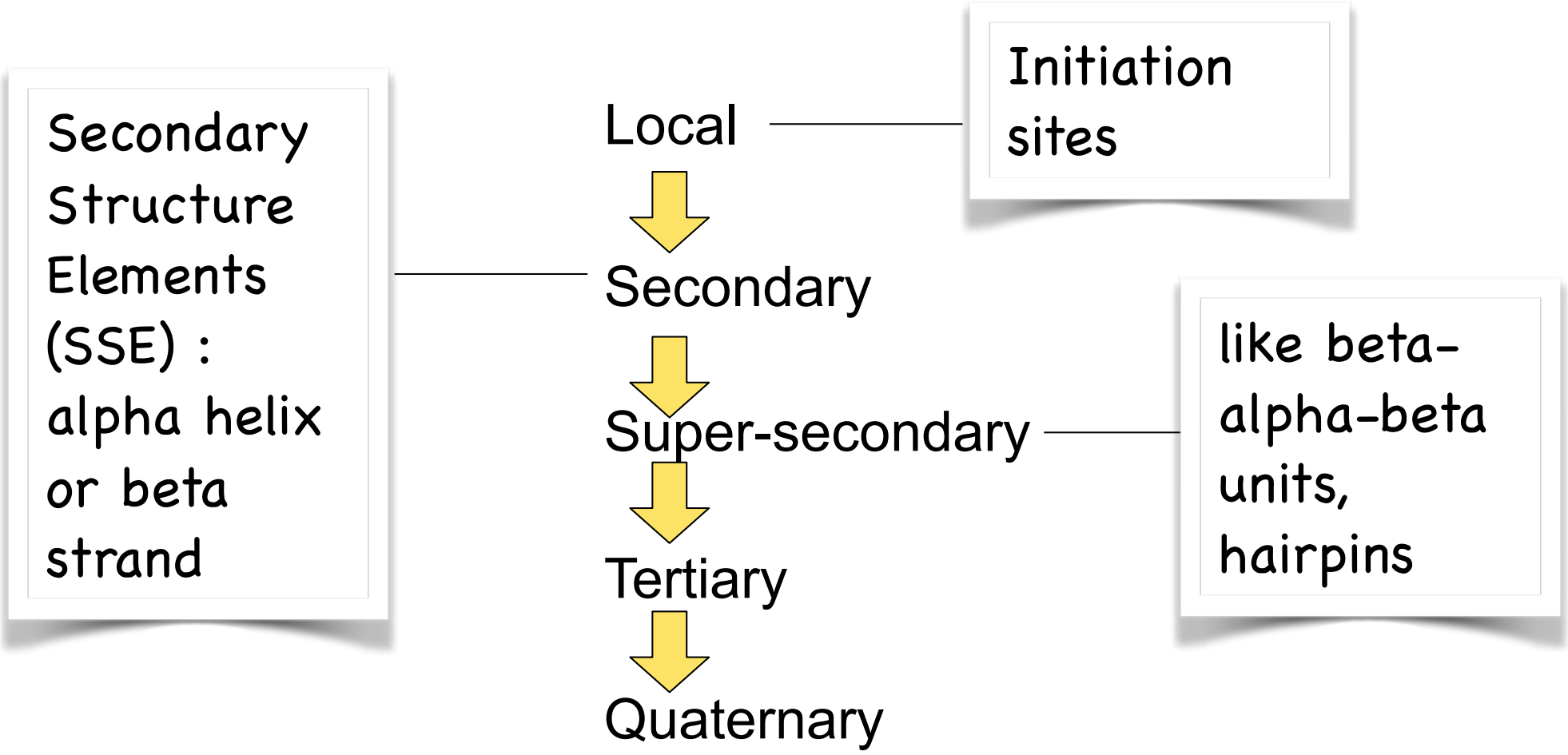
[2] Yi Q, Bystroff C, Rajagopal P, Klevit RE & Baker D. (1998). Prediction and structural characterization of an independently folding substructure in t he src SH3 domain. J Mol Biol283, 293-300.

Many proteins share common core structures (Efimov cores)



Ancestral fold?
and/or
Folding intermediate?

Folding


Local
Secondary
Super-secondary
Tertiary
Quaternary
Initiation sites
Secondary Structure Elements (SSE) : alpha helix or beta strand
like beta-alpha-beta units, hairpins

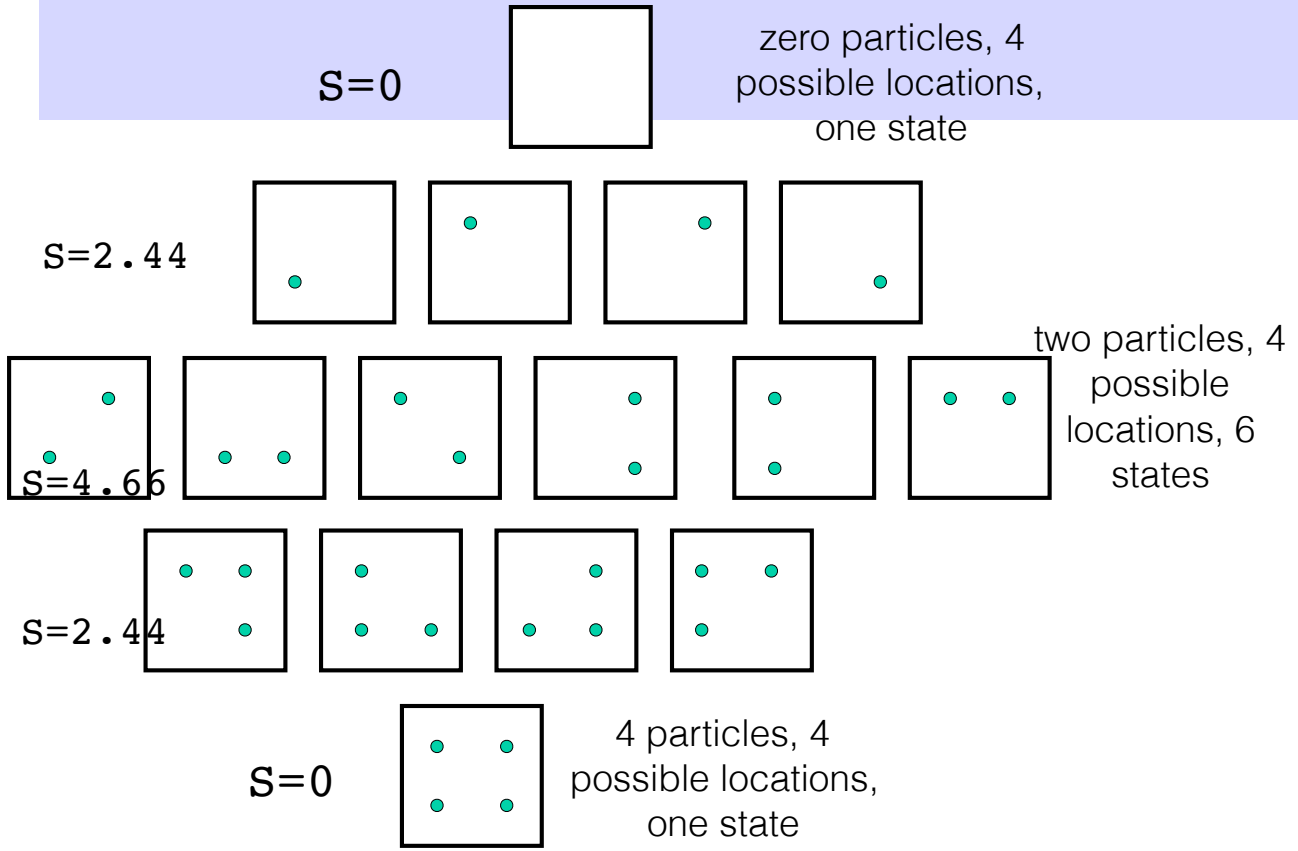
Nature abhors a vacuum

no, not this kind...

There is only **one way** to make space empty, but **many ways** to fill it.

S = $p \log p$, *where p is the number of states*.

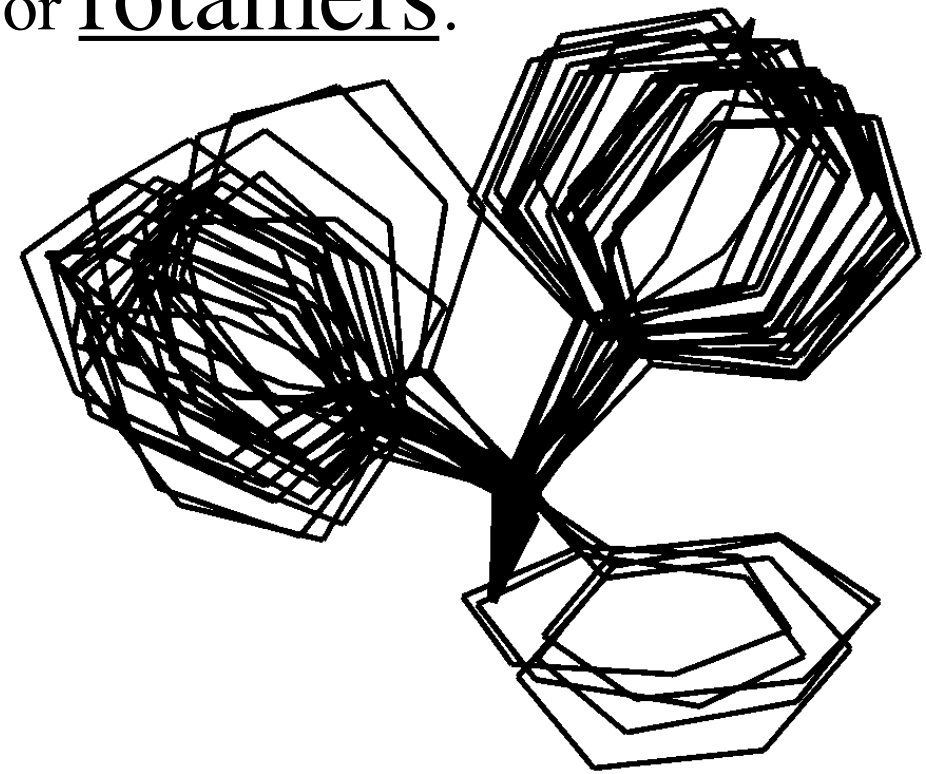
Higher entropy means more probable.



Sidechain Rotamers

Discrete approximation of the continuous space of backbone angles.

Sidechain conformations fall into descrete classes called rotational isomers, or rotamers.



A random sampling of Phenylalanine sidechains, w/ backbone superimposed