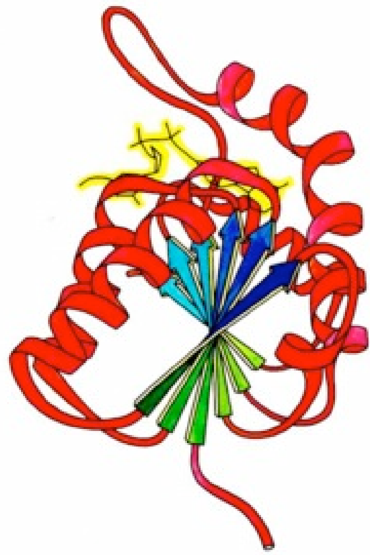
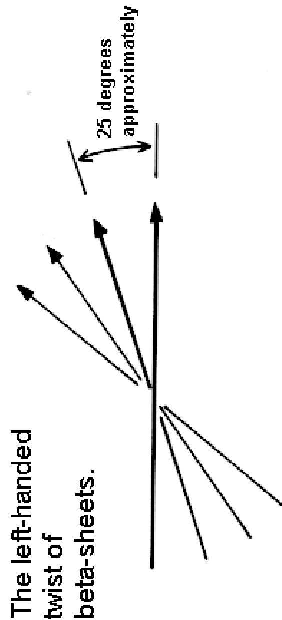
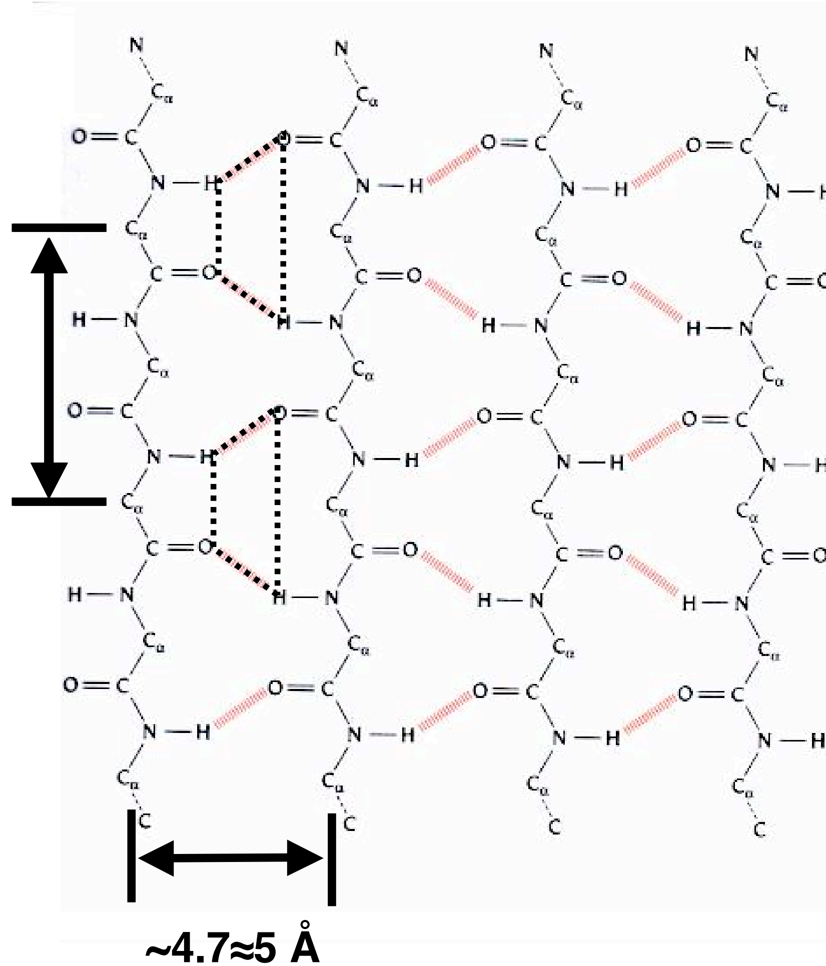
Parallel beta strands--why are they twisted?

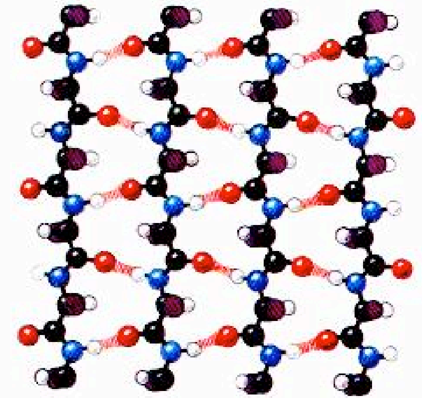


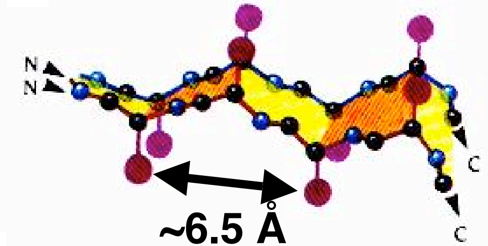
Lactate Dehydrogenase domain 1, end view



"dipeptide repeat" ~6.5 Å

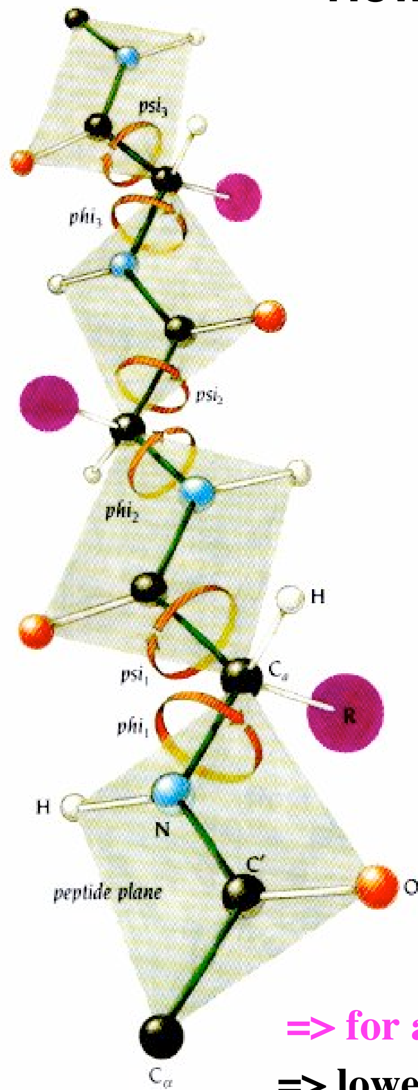


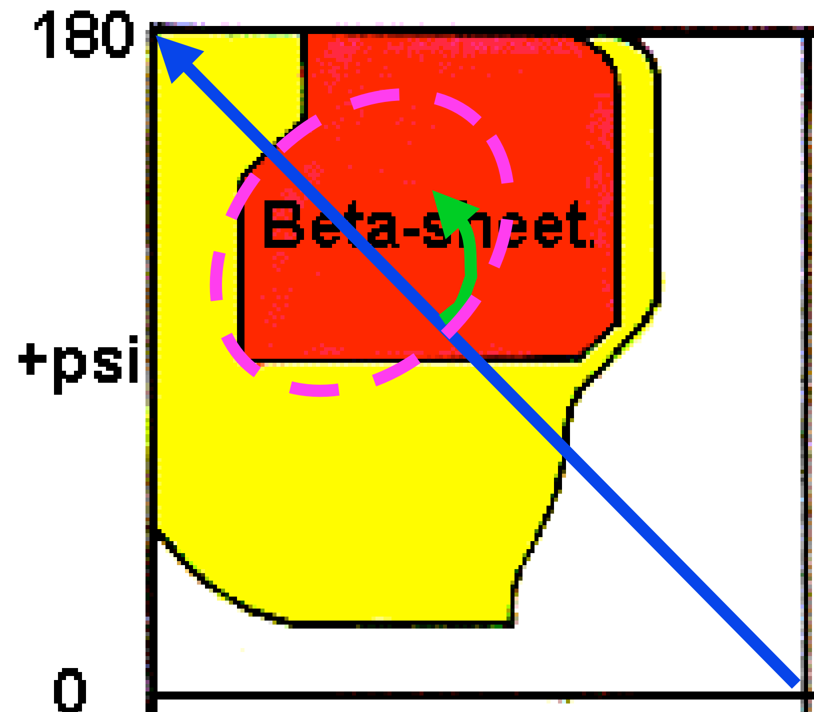




The twist of the parallel β sheet is due to the intrinsic right-handed twist of the polypeptide backbone.

How much diversity in secondary structure?





On diagonal: no twist to peptide backbone; phi=-psi

Above diagonal: right-handed twist to peptide backbone; favored by side chains larger than Ala.

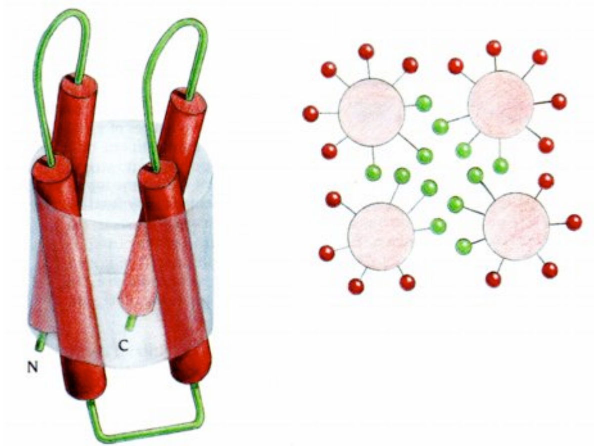
=> for a given energy, more accessible states for right-hand twist than for left-hand

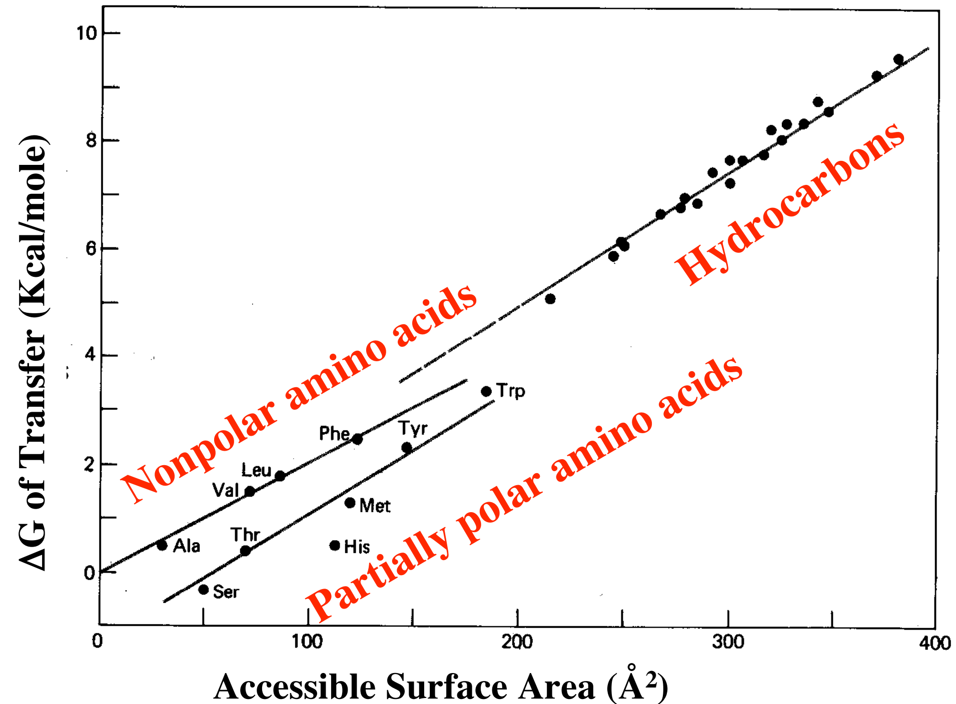
=> lowest free-energy conformations will be those in which the polypeptide backbone maximizes entropy by sampling all accessible states, which will be predominantly those with right-handed twist.

Ref: Chothia, C. "Conformation of twisted beta-pleated sheets in proteins". *J Mol Biol* 75, 295-302, (1973).

(Thanks, K.Y.)

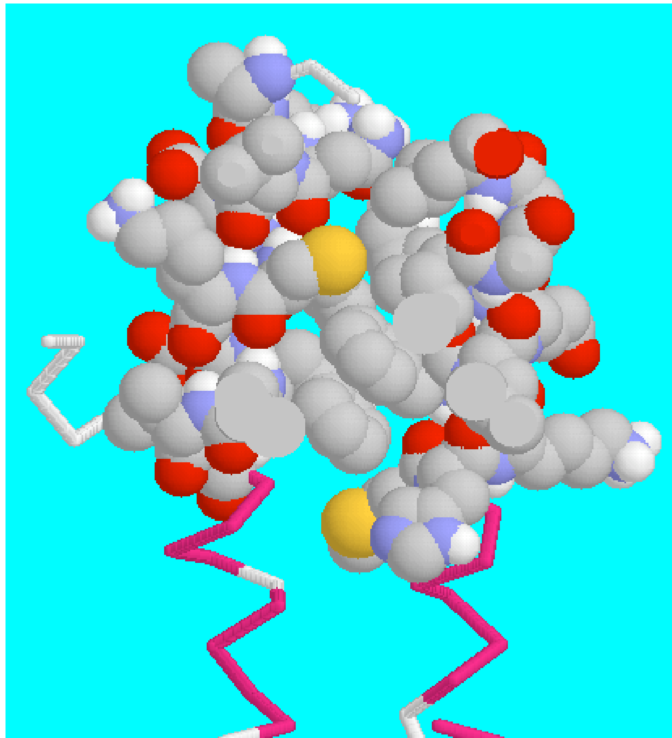
The hydrophobic contribution: quantitative view





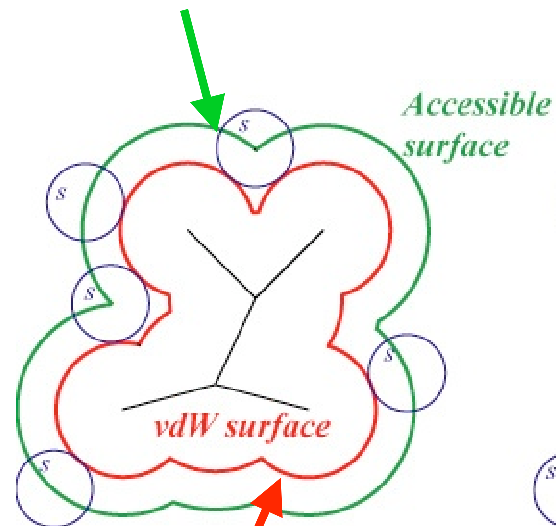
ΔG of transfer of hydrophobic side chains, aqueous -> nonpolar environment, estimated to be ~ -(0.022-0.025) Kcal/(mol $Å^2$).
Early estimates of energy of stabilization of protein by transfer of nonpolar surfaces from solvent -> interior ~0.025-0.030 Kcal/(mol $Å^2$).

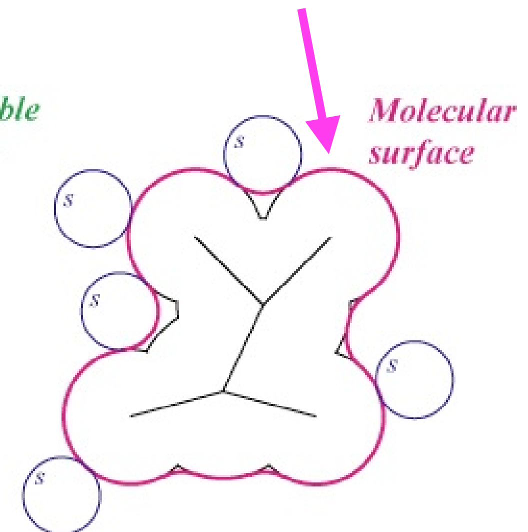
Measuring a "surface" in a protein



Probe: typically a sphere of 1.4Å radius mimicking a water molecule--center traces "Accessible surface" as it rolls on vdW surface

The "Molecular surface" is the lower envelope generated by the rolling sphere.

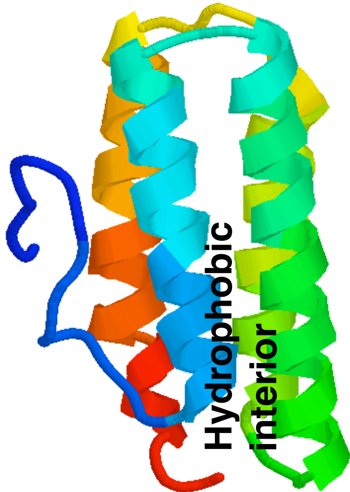




It differs from the vdW surface in that some areas are inaccessible to the rolling sphere.

Using van der Waals radii for atoms, can define "van der Waals surface" of the molecule.

Ballpark estimate of free energy change of folding nonpolars -> interior



Unfolded [U] <-> folded/native [N]
assume hydrophobic interactions are the only factor

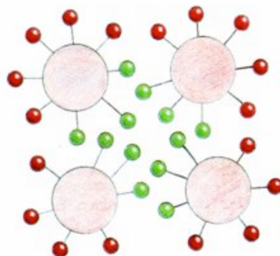
"Packing dimensions" of helices?

~25 Å length

~10 Å diameter

=> surface of helix (in <u>naïve</u> approximation)
$= \pi r^2 h \approx 2000$ Å^2

Estimate ~one third is buried =>
For four helices, ~ 2700"Å^2 buried;
0.022 Kcal/(mol Å^2) free energy of stabilization =>
ΔG folding ~ -60 Kcal/mole



At room temperature, RT≈0.6 Kcal/mole,

K_{eq} = [N]/[U] = exp (-ΔG/RT) ~exp(100) ~10^{44}

?????????!!!!!!!!!!!!!!!!!##*&@(+*#&

From Creighton, "Proteins":

Table 7–6 Estimates of the Net Contributions of Individual Factors to Stability of an Average Protein of 100 Residues

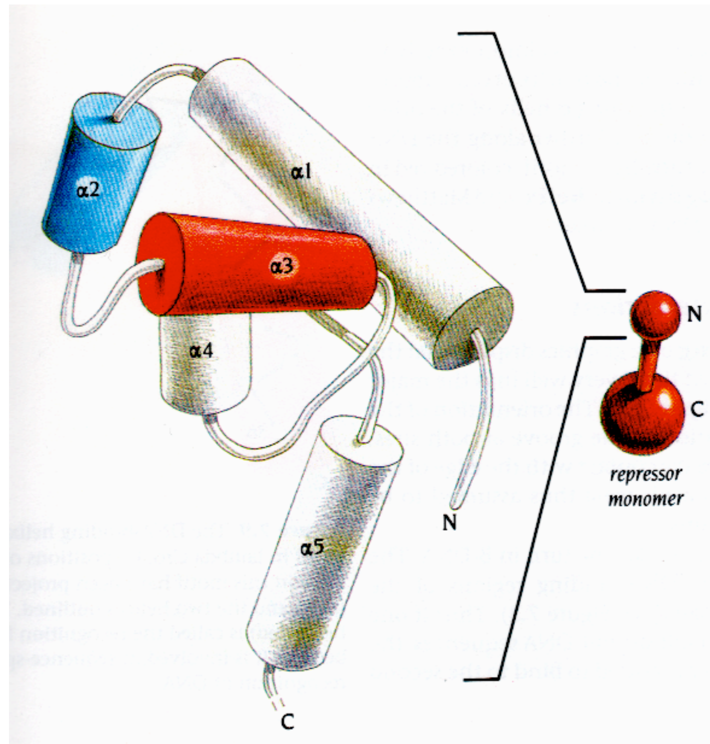
FACTOR	FREE ENERGY DIFFERENCE BETWEEN N AND U AT 25° C (kcal/mole)	CONTRIBUTION TO ENTHALPY DIFFERENCE
Conformational entropy at 25° C	+330 to +1000	negligible
Unfavorable interactions in folded state	+200	positive
Hydrophobic interactions[a]	−264	positive
Increased van der Waals interactions due to close-packing[b]	−227	negative
Required contribution of hydrogen bonds[c]	−49 to −719	negative
Observed net effect	−5 to −10	negative

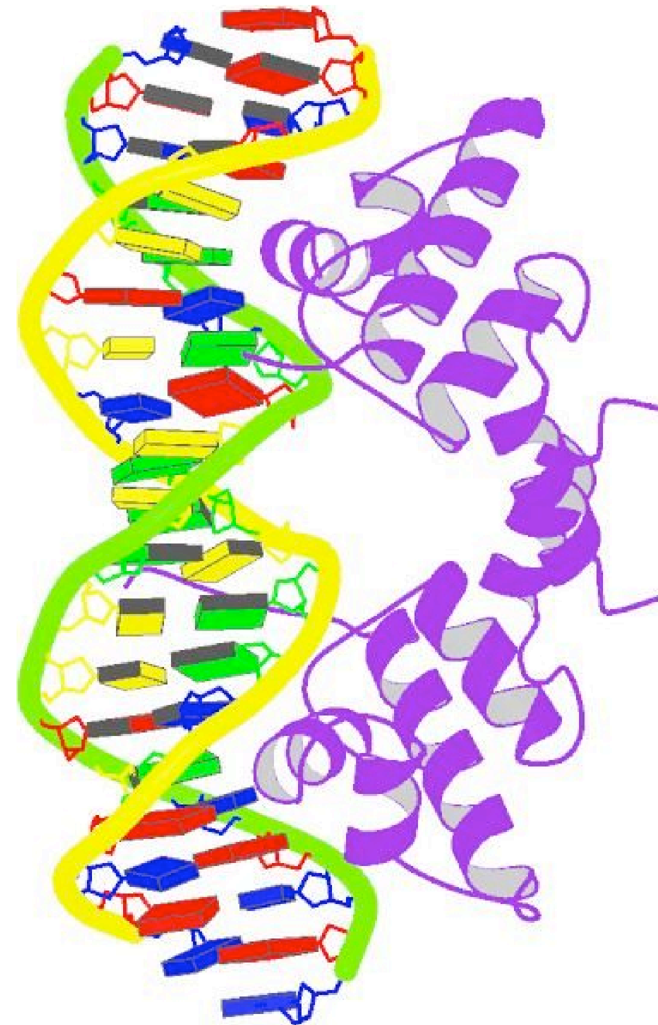
[a]The free energy of the hydrophobic interaction is proportional to the buried accessible surface area (−24 cal/mole/Å^2; Figure 4–5). The average surface area buried in a protein of molecular weight M, in Å^2, is given by $1.449\,M - 11.116\,M^{2/3}$: from D. C. Teller, Nature 260:729–731, 1976.

[b]The free energy change is estimated from the average enthalpy of fusion observed with small, nonpolar model compounds, −30 cal/g, multiplied by the fraction of accessible surface of the protein involved in close-packed interactions (given by $0.595 + .0.536\,M \times 10^{-5}$ for an average protein: C. Chothia, Nature 254:304–308, 1975.

[c]The average number of protein hydrogen bonds would be 74 (C. Chothia, Nature 254:304–308, 1975), so an average contribution of −0.7 to −9.7 kcal/mole would be required. With $K_{AB} = 10^{-2}\ \text{M}^{-1}$ for hydrogen bonding in water (Table 4–5), effective concentrations of the interacting groups of 310 to 1.4×10^9 M would be required in proteins.

Bacteriophage lambda repressor: biological activity is to regulate transcription by binding a specific DNA site. The repressor is a dimer; the DNA binding domains, by themselves, also form a dimer. Structures of the DNA binding domains, both with and without DNA, have been solved.



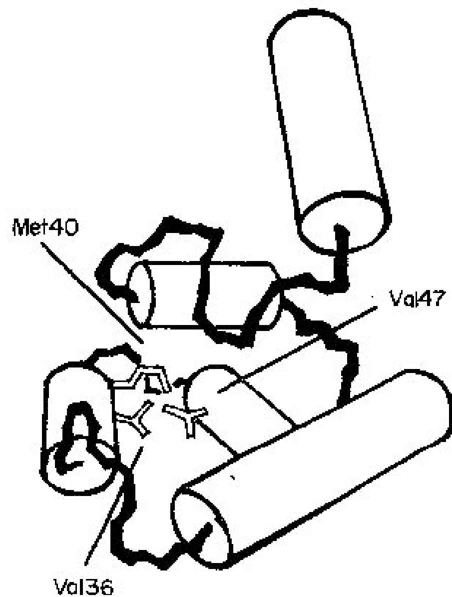


Lambda repressor used to test models for protein folding and stability.

Experiment (with several variations):

Mutagenize cluster of hydrophobic residues in the interior "core" of the protein.

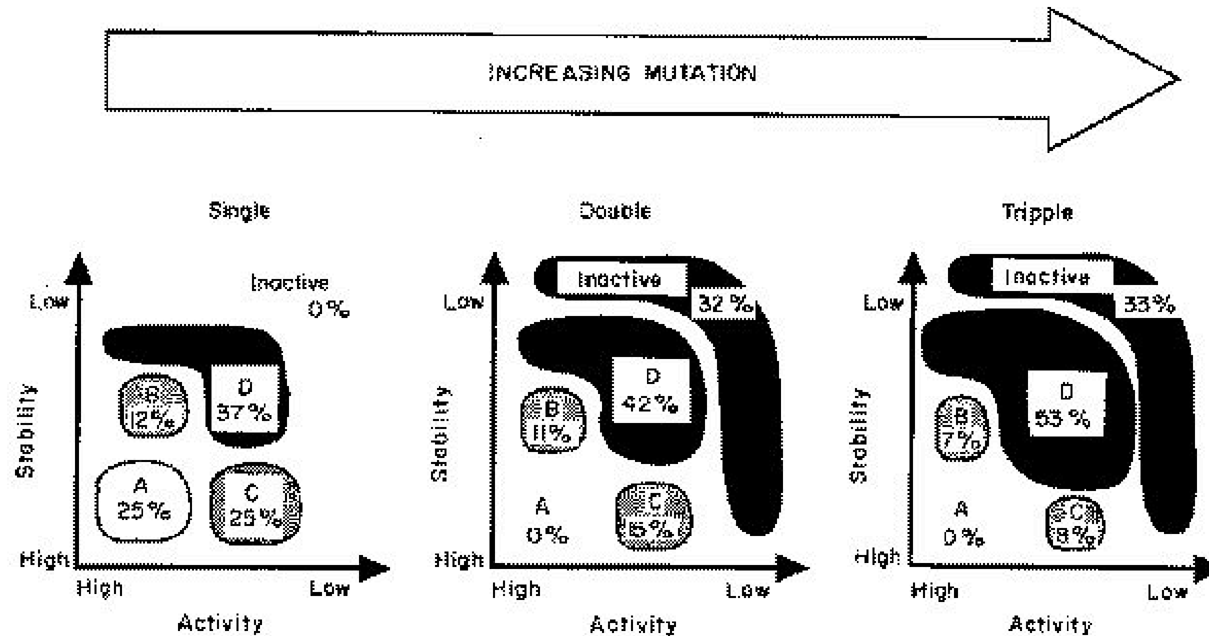
Measure *in vivo* activity, and temperature dependence of *in vivo* activity, which is related to protein stability (e.g. 26, 30, 37, 42 °C).



E.g. in one case, each of three residues (V36, M40, V47) randomly mutated to five different residues (V, L, I, M, F).
Mutants segregated into classes depending on whether they had lower activity, lower stability, or both.

Lim, W. A. and Sauer, R. T. "Alternative packing arrangements in the hydrophobic core of lambda repressor". *Nature* 339, 31-36, (1989).
Lim, W. A. and Sauer, R. T. "The role of internal packing interactions in determining the structure and stability of a protein". *J Mol Biol* 219, 359-376, (1991).

Results of mutagenesis of hydrophobic core of lambda repressor



78 of a possible 125 mutations characterized. General trend: higher level of mutation gives drift toward reduced stability, reduced activity, or both.

Main conclusions:

"…we find that the majority of sequences meeting the simple requirement that core residues remain hydrophobic are able to adopt a reasonably stable conformation with the same basic shape as wild-type λ repressor".

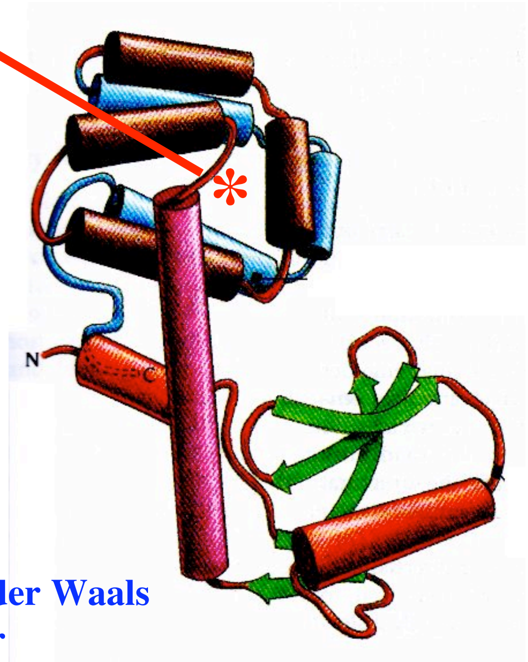
"By contrast…. packing interactions do appear to play an important role in specifying the precise structure and ligand binding properties of the protein".

Lessons on protein stability learned from T4 lysozyme

(B.W. Matthews, Adv. Prot. Chem. 46, 249-278 (1995)).

Change in T4 lysozyme stability due to Leu->Ala or Phe->Ala Mutation in nonpolar interior

Leu or Phe->Ala



Decrease in protein stability $\Delta\Delta G$(Kcal/mole)

0 -2 -4 -6 -8

0 50 100 150 200 250

Increase in cavity volume (Å^3)

L46A L121A L118A L133A F153A L99A L99A/F153A

$\Delta\Delta G$ due to "relaxation" of structure to compensate for (fill in) internal cavity.

$\Delta\Delta G$ (probably) due to loss of Van der Waals interactions in hydrophobic interior

~1.9 Kcal/mole due to loss of "hydrophobic effect", i.e. decrease of hydrophobic surface transferred from polar to nonpolar environment [~0.025 Kcal/(mole Å^2)]

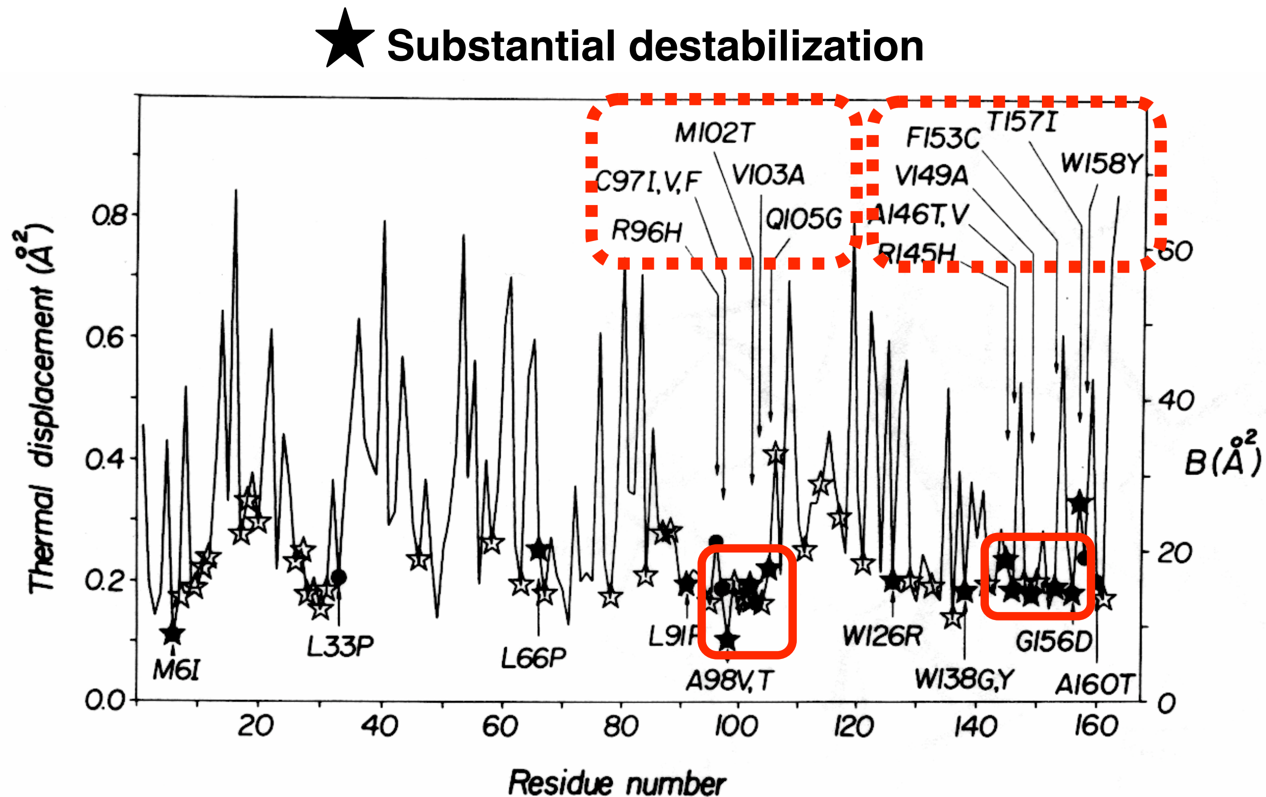
=> $|\Delta\Delta G|$ is ~ 0.025 Kcal/(mole Å^2) due to altered contribution of hydrophobic effect <u>plus</u>
~ 0.020 Kcal/(mole Å^2) for the surface area of the internal cavity after the protein has relaxed

Where are mutations most destabilizing to the structure?

--suppression of amber mutants used to introduce 13 different amino acids at all but one residue; of 2015 single amino acid substitutions, only 173 (<10%) localized to 53 residues were seriously deleterious.

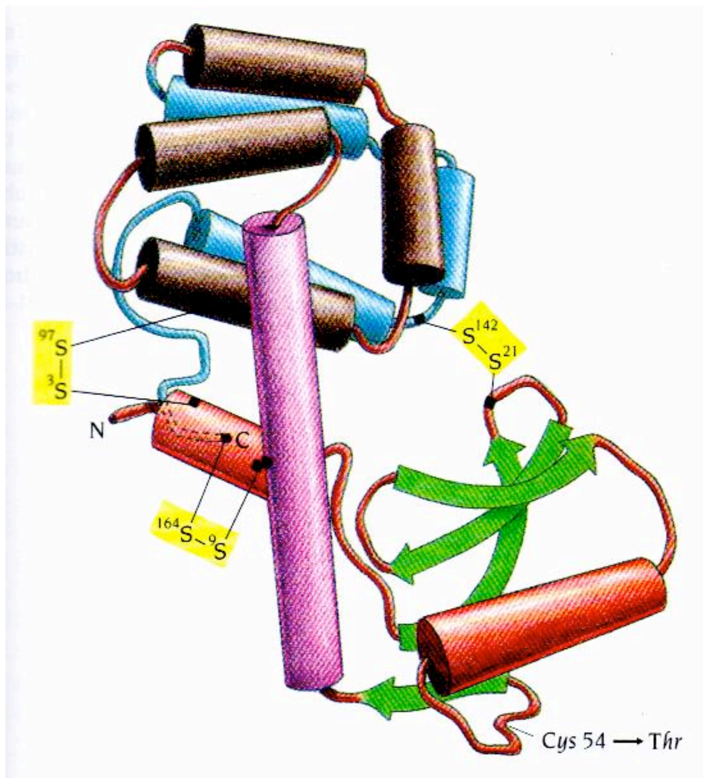
Statistical order of structure in crystal

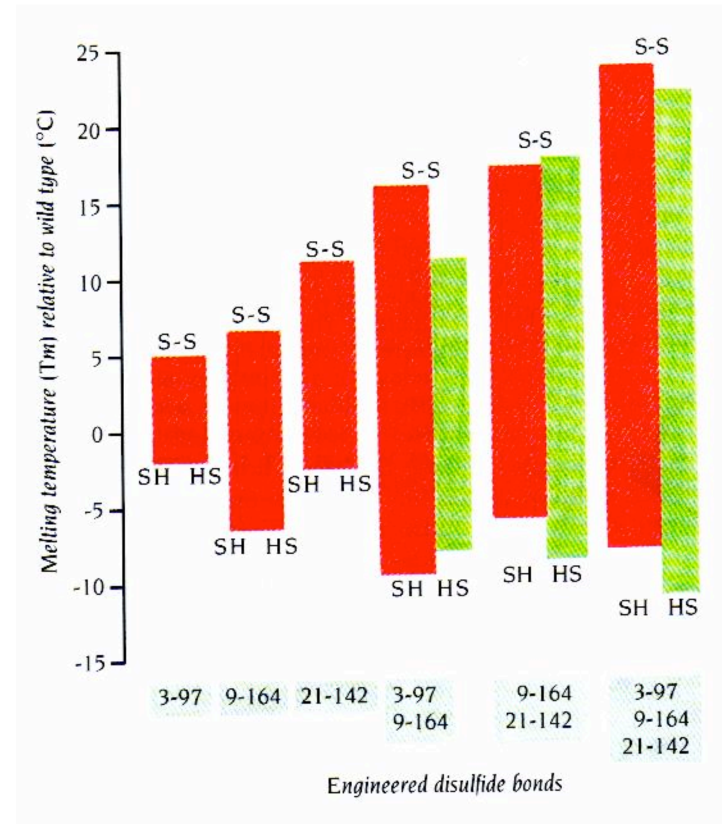
Increasing disorder -->



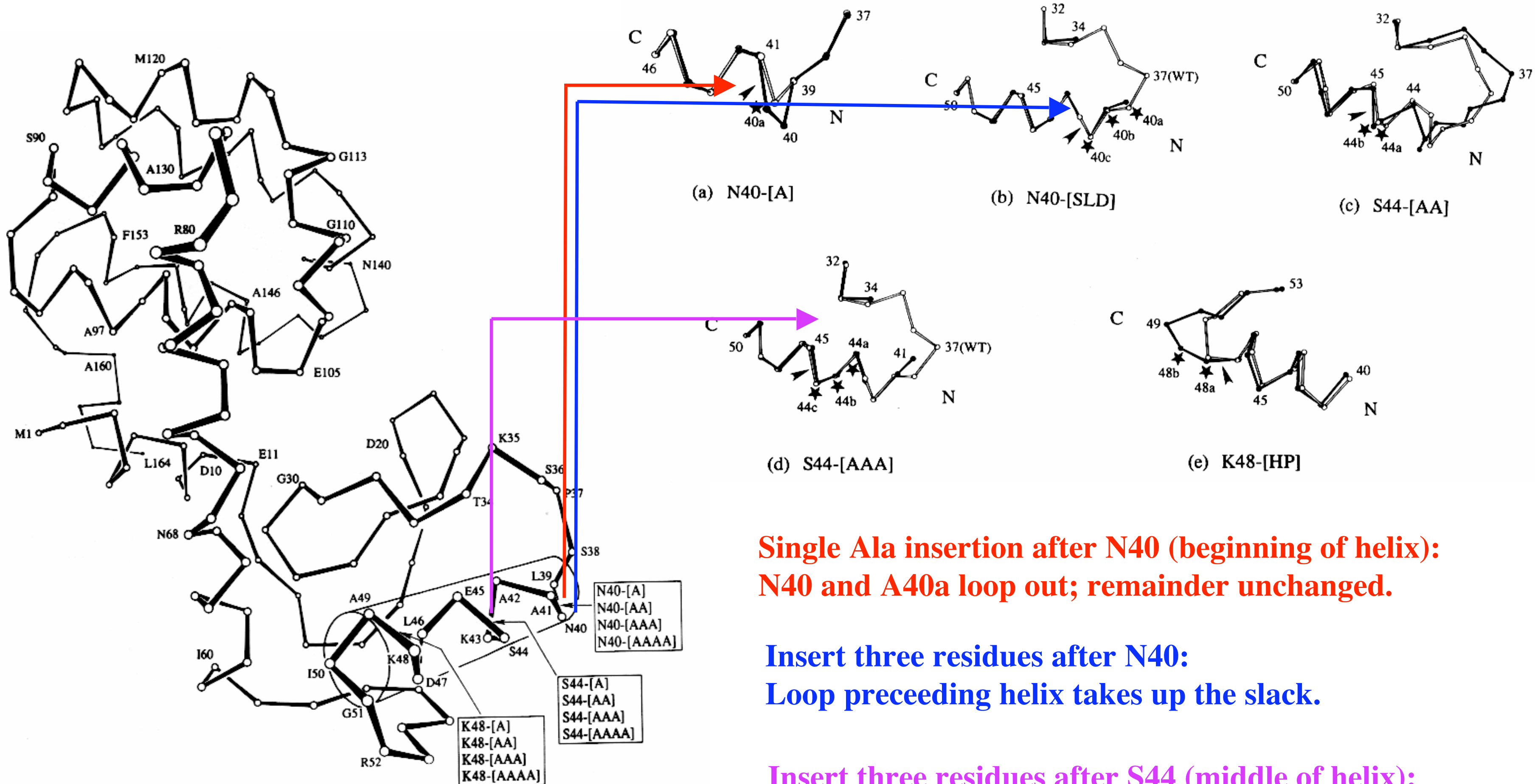
=> the most well-ordered part of the structure is where mutations have the most destabilizing effect. (the more flexible regions of the protein have the ability to adjust to mitigate effect of mutations).

Introduced disulfides (when oxidized) stabilize T4 lysozyme structure





Insertions in a helix



Single Ala insertion after N40 (beginning of helix):
N40 and A40a loop out; remainder unchanged.

Insert three residues after N40:
Loop preceeding helix takes up the slack.

Insert three residues after S44 (middle of helix):
Loop preceeding helix takes up the slack; shift of register in helix.

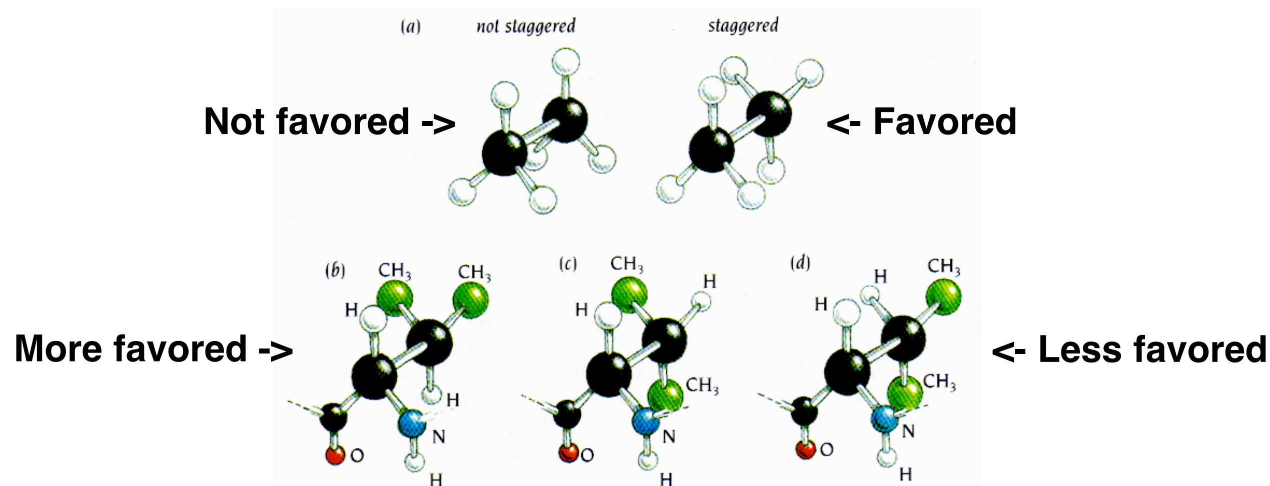
=> Interface between the helix and the rest of the structure is maintained in response to insertions.

Summary

B.W. Matthews, Adv. Prot. Chem. 46, 249-278 (1995)

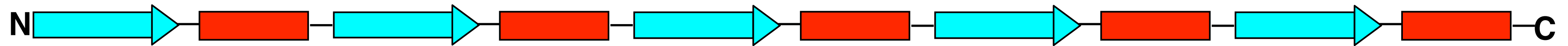
"One striking result has been the ability of the protein to accommodate changes at many sites, yet still fold and retain activity. This shows that many amino acids in the sequence … are nonessential for protein folding and stability. Such amino acids appear, predominantly, to be those that are mobile and/or largely exposed to solvent. Even though buried residues seem to be more important, it is still possible to change these as well. … it has been found that a combination of substitutions permits repacking of the core. Such repacking is associated with adjustments of both the main chain and side chains. Only rarely do side chains rotate into radically new orientations*".

"Substitutions of core residues confirm the overall importance of the hydrophobic effect as the dominant factor in stabilizing the folded structures of proteins. "Cavity-creating" substitutions of the form Leu->Ala show that the burial of the bulky leucine side chain within the core confers greater hydrophobic stabilization than … the smaller alanine side chain. Leu->Ala substitutions that create large cavities are especially destabilizing because they result in a loss of both hydrophobic and van der Waals interactions. In cases where the protein relaxes to reduce the size of the putative cavity, alternative van der Waals interactions are generated and the overall destabilization of the protein may be less severe than in cases where a large cavity is formed".

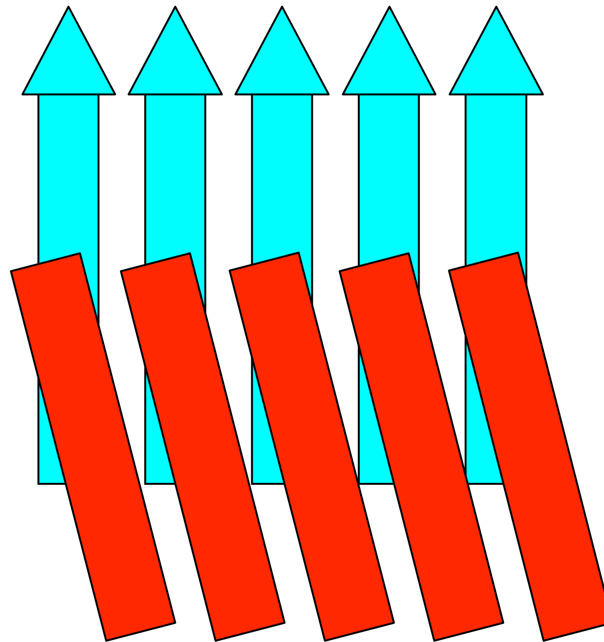


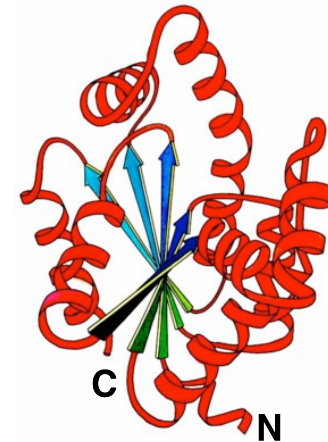
***side chains have preferred rotamer conformations; "flipping" to alternative rotamers is rare.**

Sequence ->(?) Secondary structure ->(?) Tertiary structure:

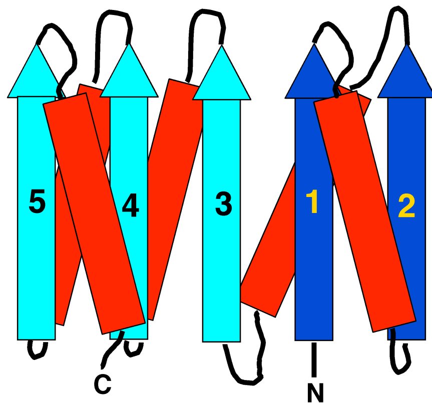


Flavodoxin





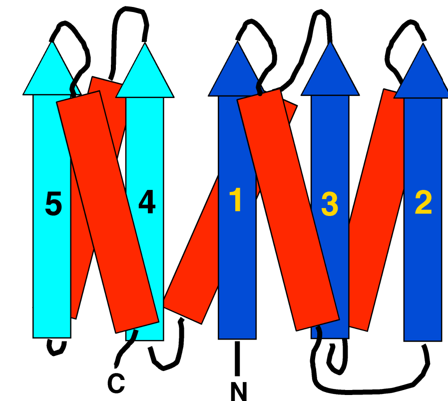
Adenylate Kinase



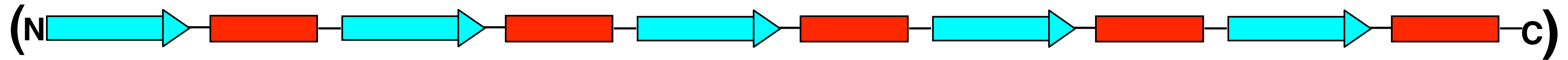
#ways to order strands = 5!/2

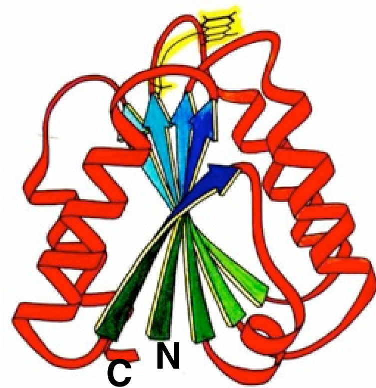
Two ways (right versus left turn) for helix on each strand

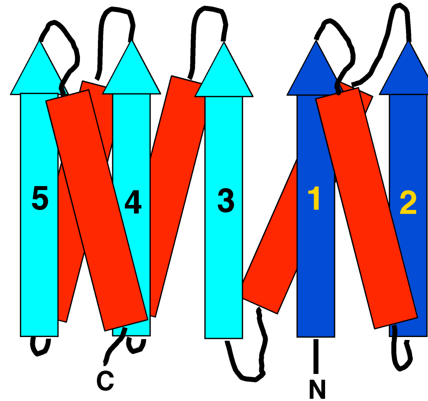
=> number of theoretical folds:
$5!/2 \times 2^5 = 960$



Suppose you have been able to get as far as:
Sequence -> (Secondary structure) -> Tertiary template







i.e. you know the tertiary template for a sequence, but not how the amino acids pack together in the structure.

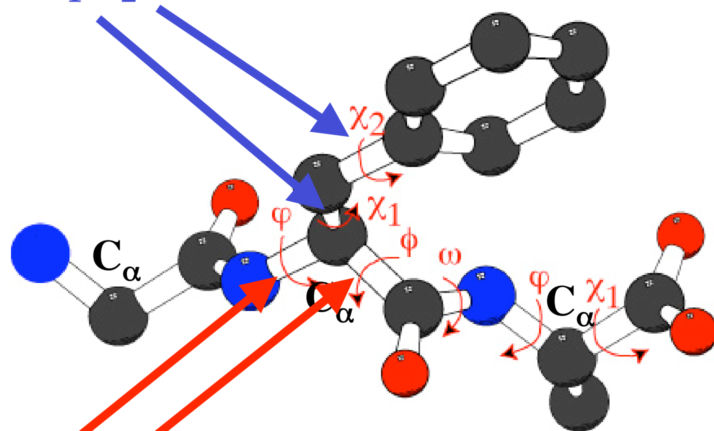
[Useful perspective: Baker, D., and Sali, A., Protein structure prediction and structural genomics, *Science*, *294*, 93-96. (2001).]

[Useful resource: http://csb.stanford.edu/koehl/BioEbook/]

<u>Outline</u> of <u>one</u> approach to modeling/packing amino acid side chains

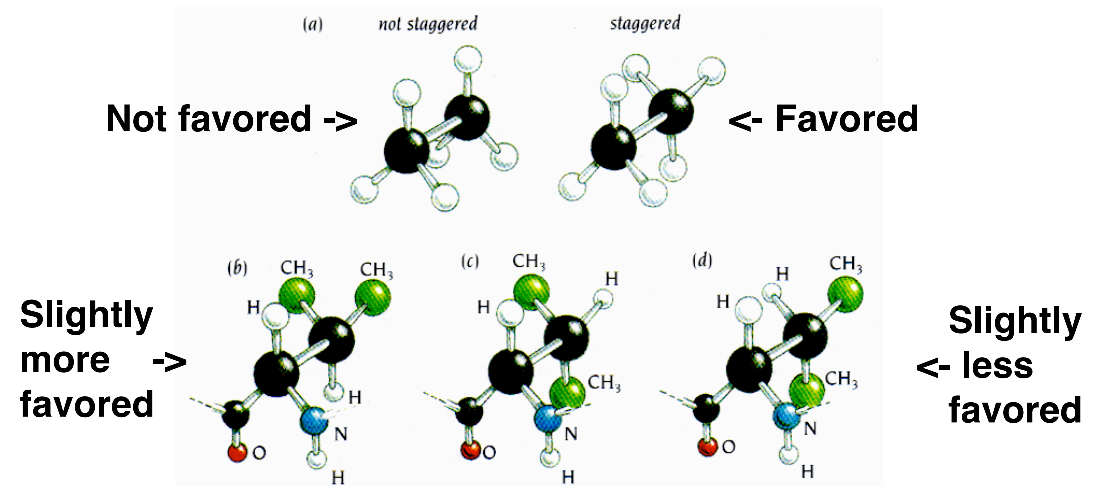
Each side chain >Gly has, in principle, a continuum of possible conformations for each "rotamer".

(χ_1,χ_2,...) define side chain conformation

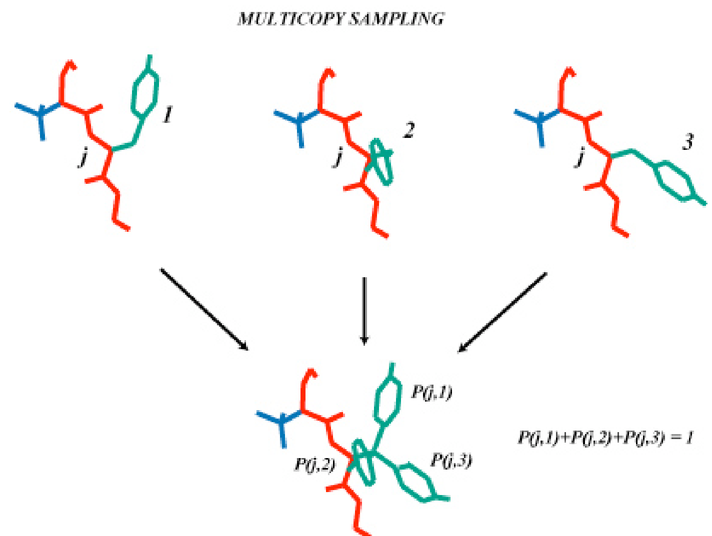


(ϕ,ψ) define backbone conformation

In practice, side chains have preferred rotamer conformations; e.g. three favored χ_1 angles for valine:



To simplify the problem, consider a discrete set of conformations for each side chain, (this may be the preferred rotamers, or it may be sampling at discrete angular intervals).



Then, consider the statistical problem: what does the i-th conformation of the j-th residue see?

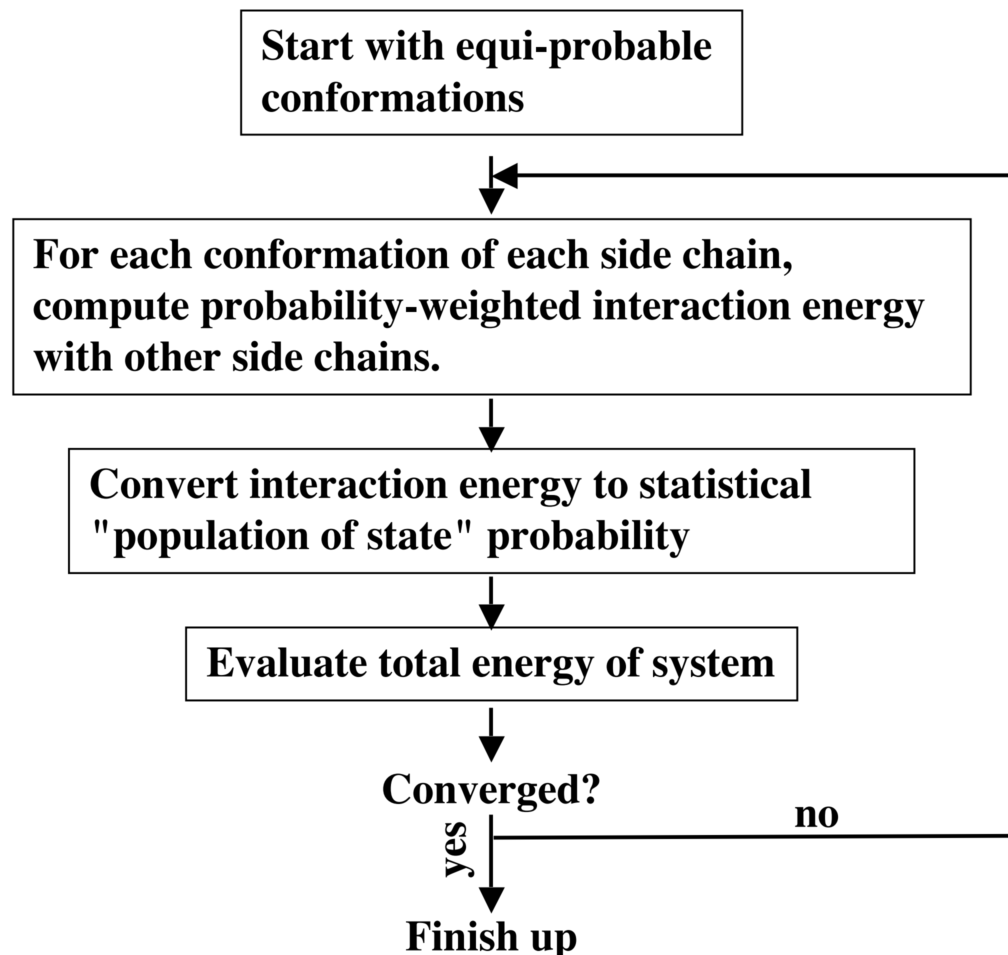
In terms of an ensemble of allowed conformations,

$$E(i,j) = \sum_{\text{all other residues}} \; \sum_{\text{each conformation of residue}} \text{Probability}(\text{k-th conf., l-th res.}) \times \text{Energy of interaction}(i,j,k,l)$$

For example, the major contributor to the interaction energy for internal packing of hydrophobic residues would be the van der Waals energy.

We can then define the probability of a specific conformation of a side chain in terms of (a) the energy of interaction with a statistical average of all other conformations of all other residues (the "mean field" of other residues), and (b) an effective temperature T:

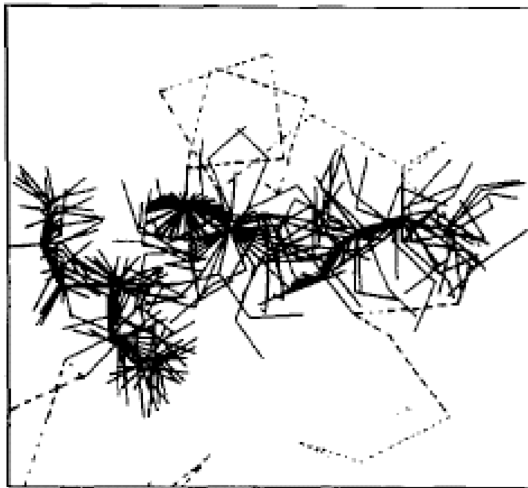
P(i-th residue, j-th conformation) = Exp (E(i,j)/kT)



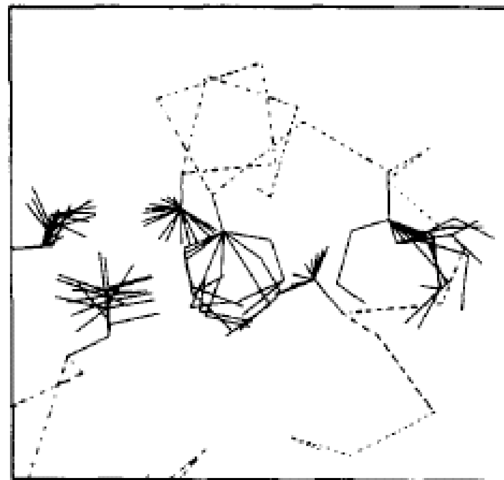
Example: computing conformations of residues in hydrophobic core of λ repressor protein

Lee, C., Predicting protein mutant energetics by self-consistent ensemble optimization, *J Mol Biol, 236*, 918-39. (1994).

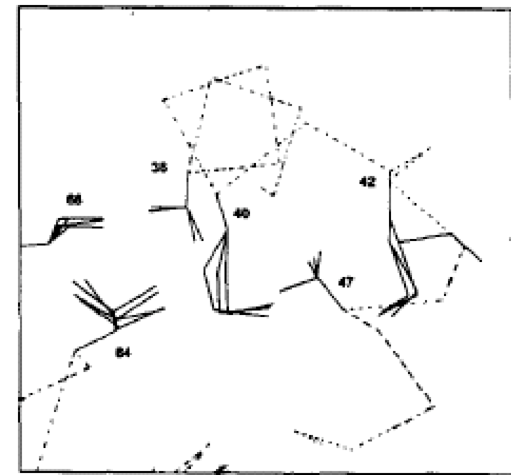
Six residues in interior:



Starting conformations

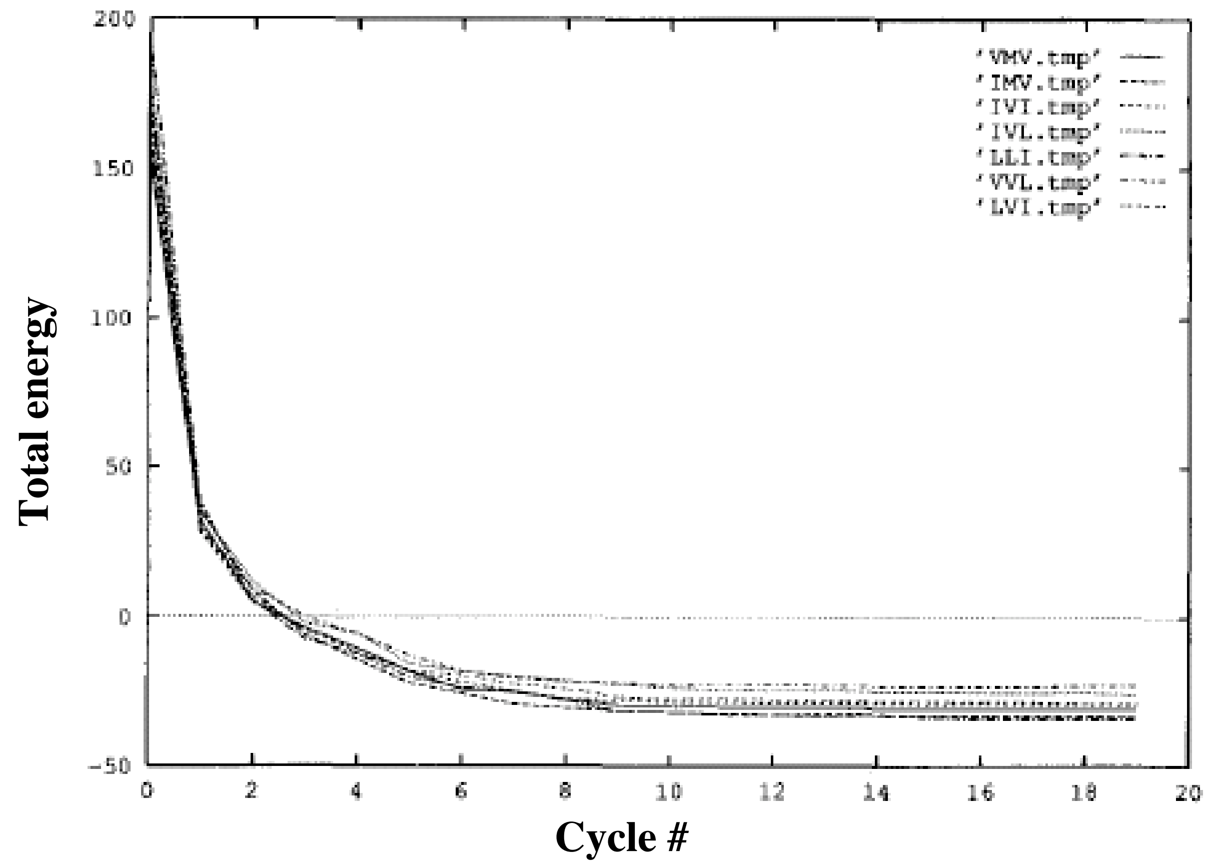


Intermediate conformations

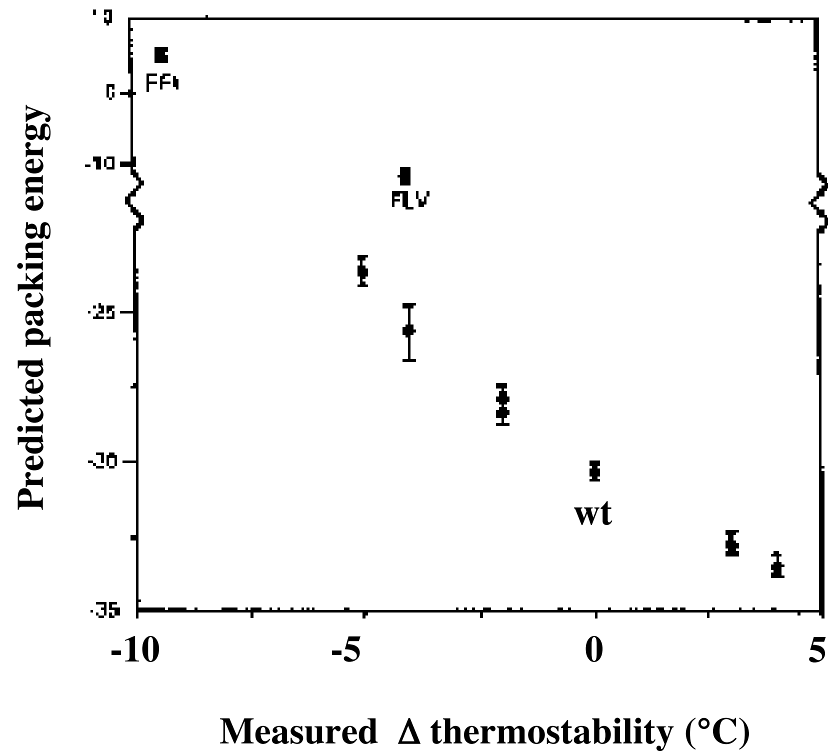


Final conformations

Convergence of method for a series of wild type and mutant models



Correlation with experimental data on the effects of mutations on the thermostability of λ repressor protein:



=> method gives a correlation between predicted and measured effects of mutations on stability of the protein

Some web sites for molecular modeling:

MODELLER	**http://www.salilab.org/**
3D-Jigsaw	**http://www.bmm.icnet.uk/servers/3djigsaw/**
SDSC1	**http://cl.sdsc.edu/hm.html**
SCWRL	**http://dunbrack.fccc.edu/SCWRL3.php**
SWISS-MODEL	**http://www.expasy.ch/swissmod/SWISS-MODEL.html**