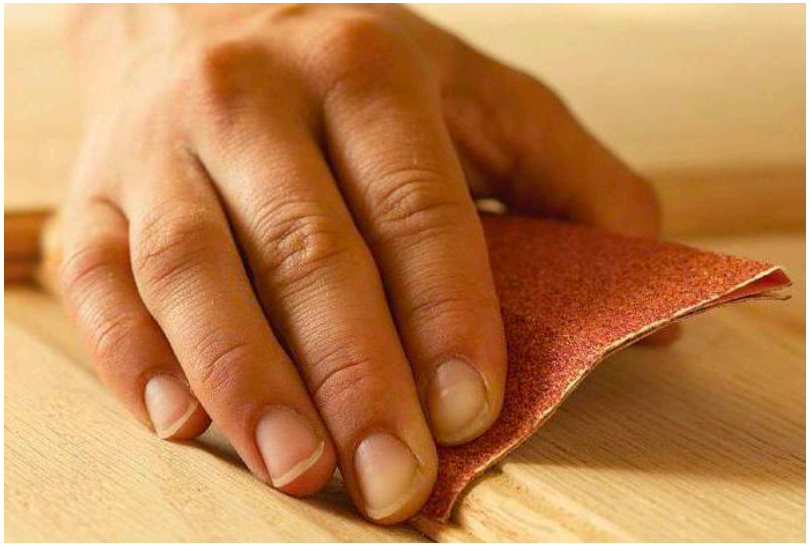
Molecular modeling 2020 -- Lecture 23

Loop modeling using evolution
Model validation

Side chain modeling



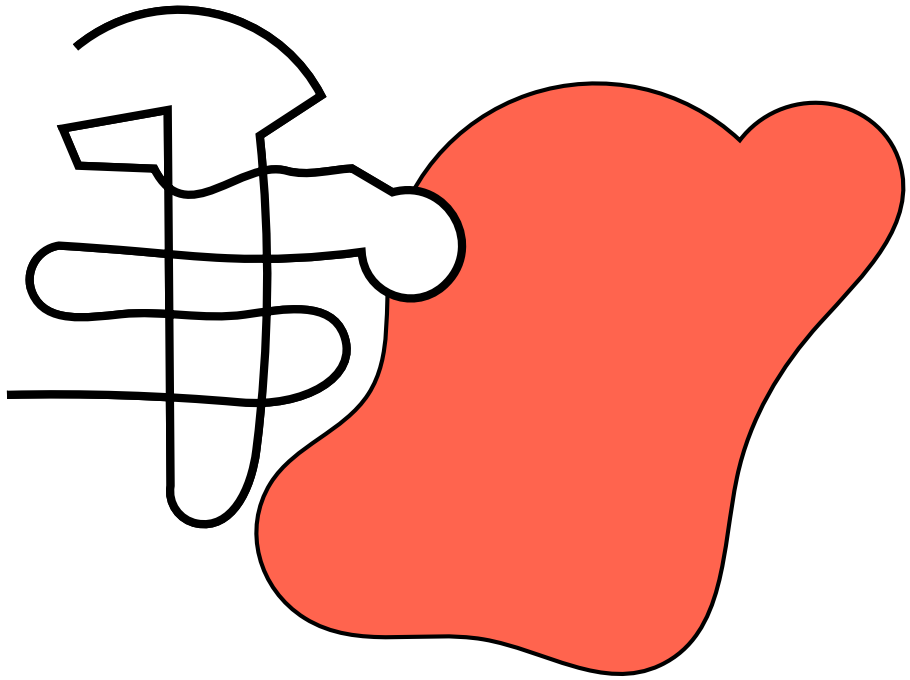
sand paper

Loop modeling

chisel

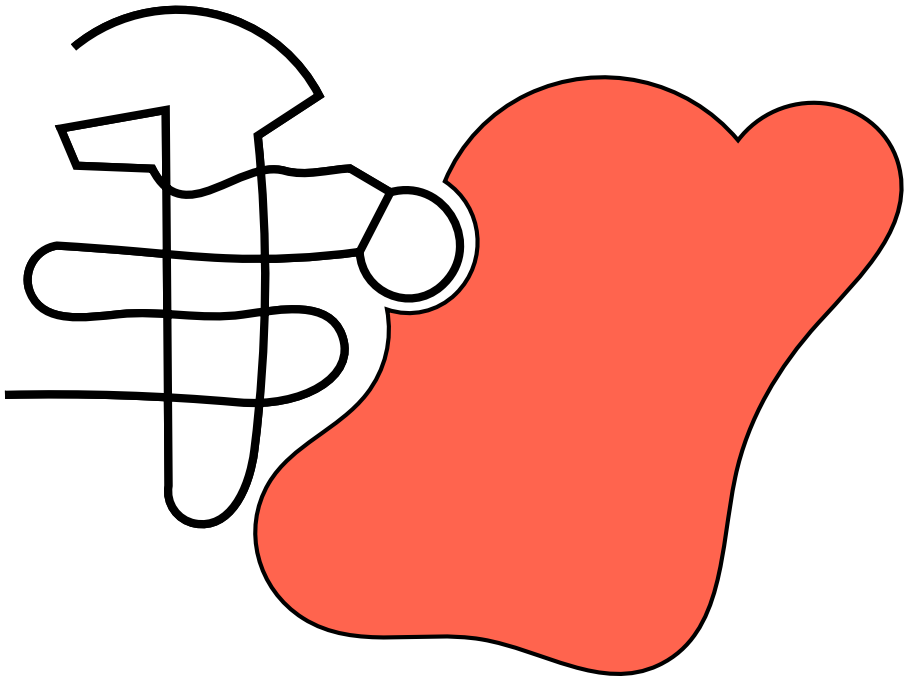
Loop modeling in the context of protein design

Docked ligand molecule fits the receptor very well except at one point. Can the loop be shortened (lengthened) to fit the receptor?

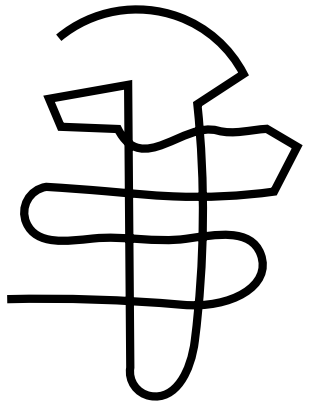


Loop modeling in the context of protein design

Docked ligand molecule fits the receptor very well except at one point. Can the loop be shortened (lengthened) to fit the receptor?



Considerations...

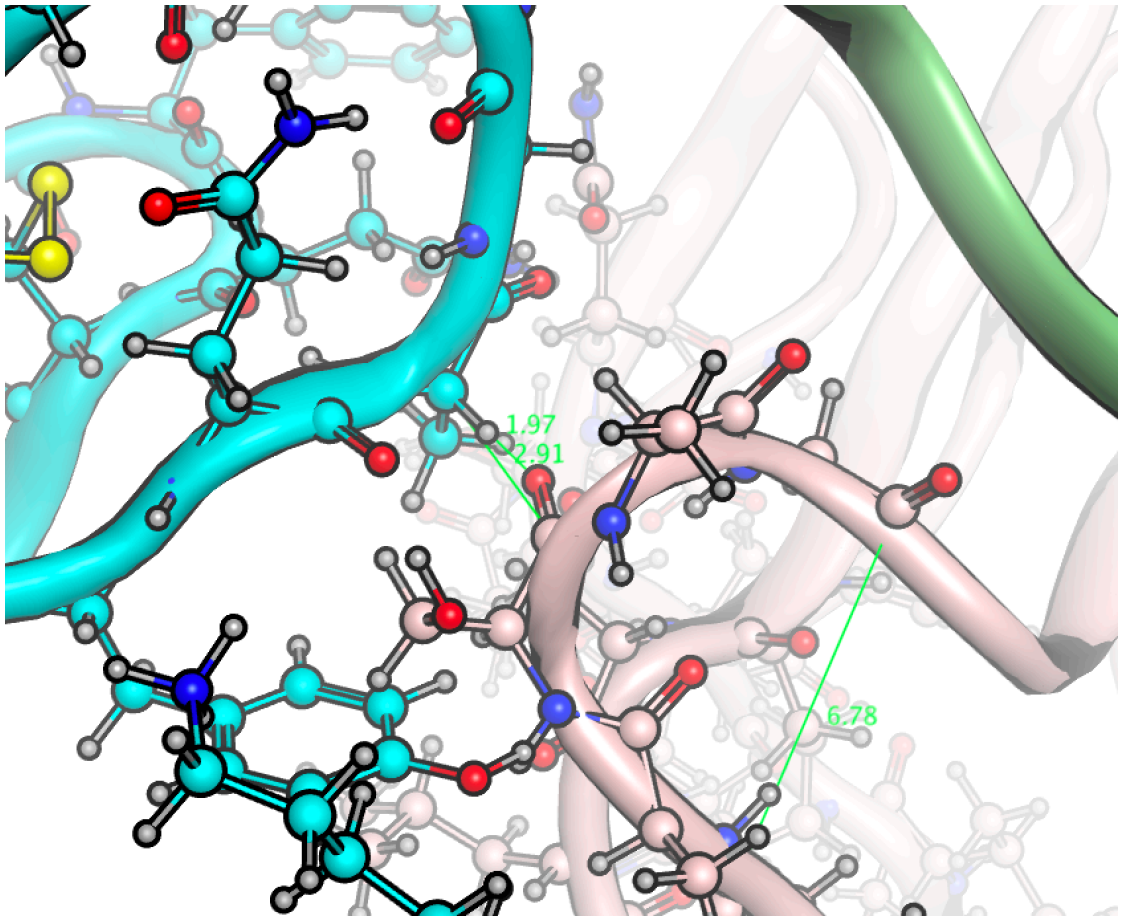


A. The loop must "look good".
B. The protein must fold.
C. The protein must assemble.

How do we know a given protein will fold and trimerize after loop design?

Backbone atoms too close after docking

spike RBD



4oul

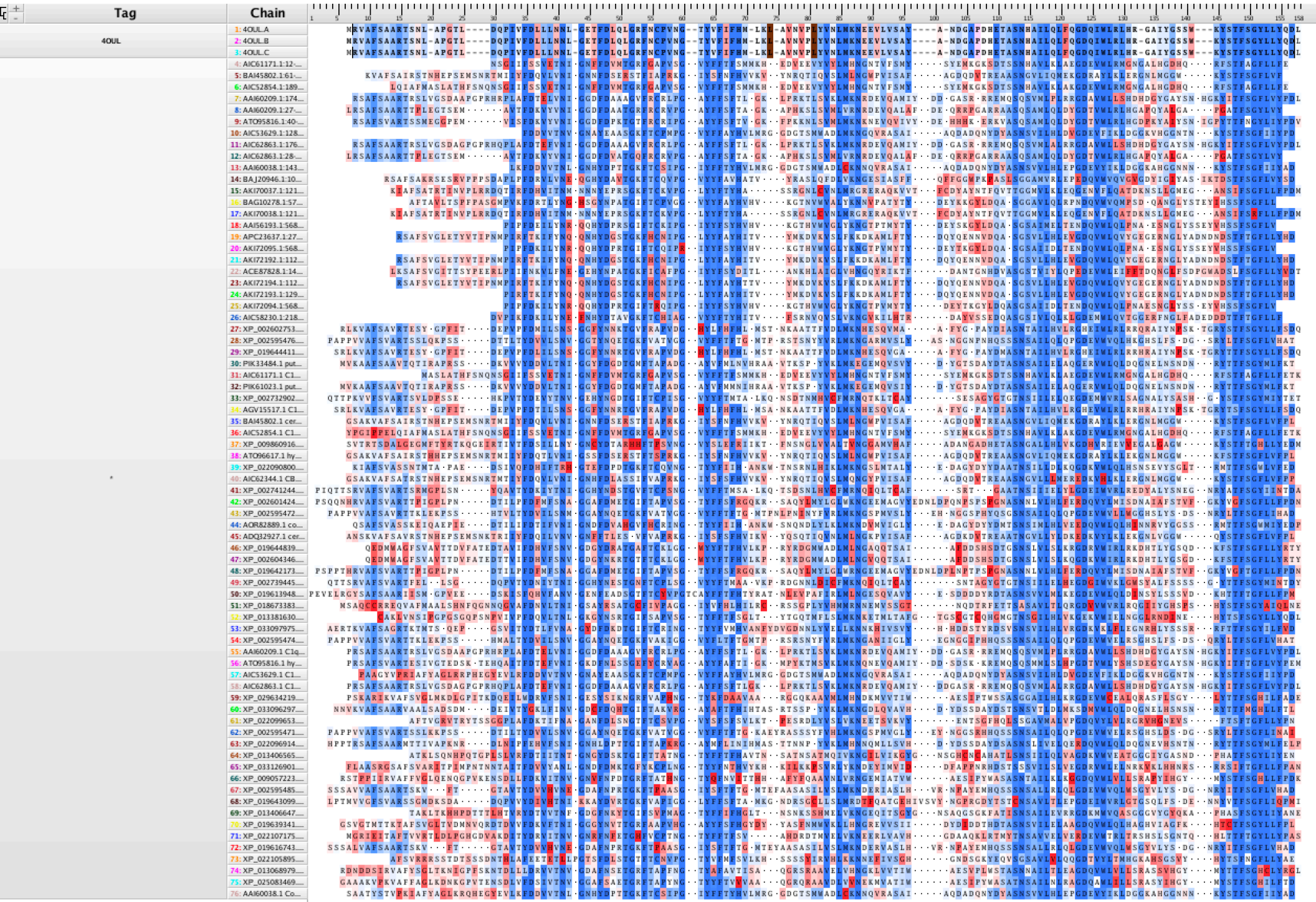
Backbone atoms collisions can't be fixed by mutation. Either...

1. Energy minimize, or
2. Loop model.

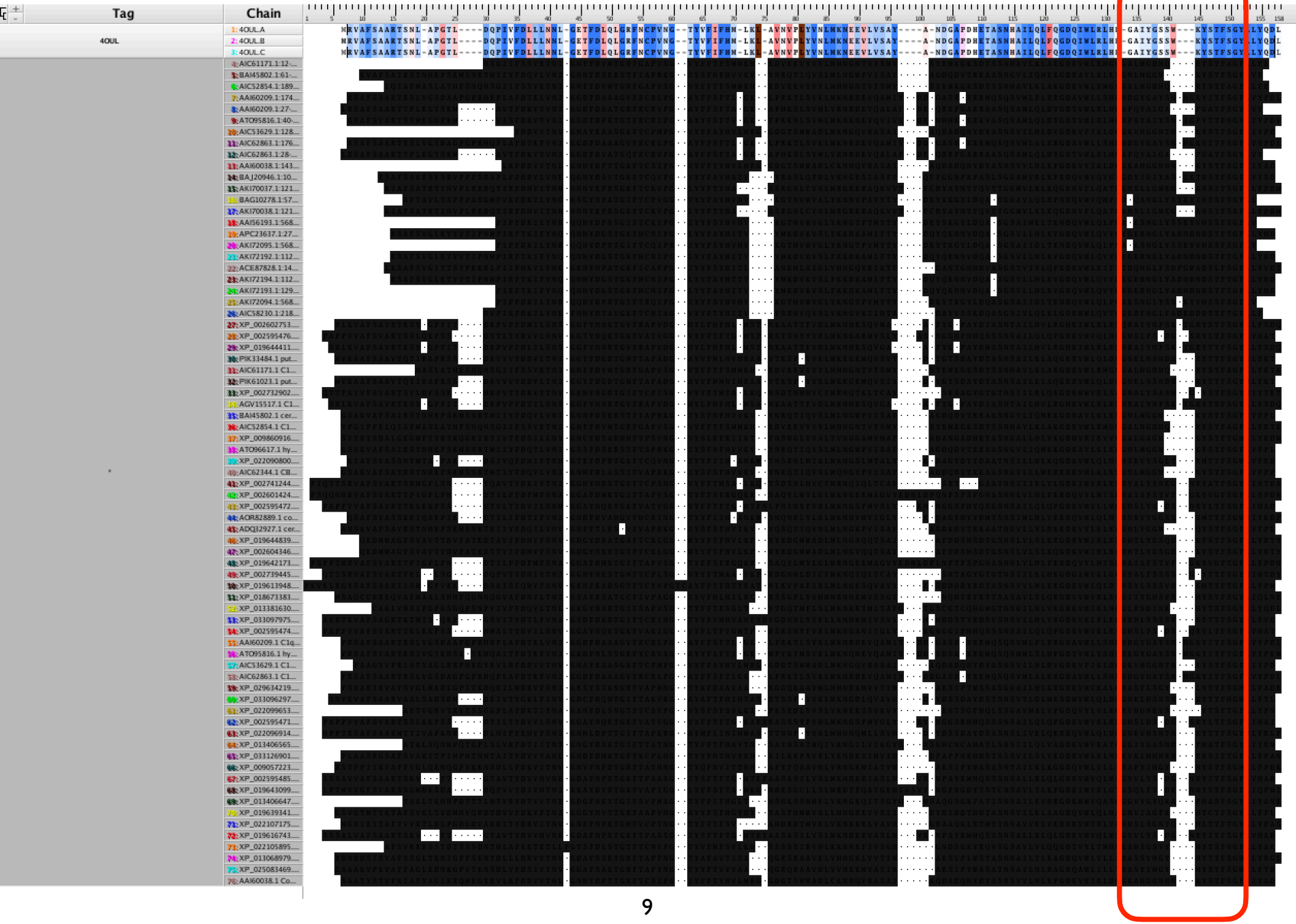
But, the crystal structure is the lowest energy state. Any shift from this state is higher energy. Moving the loop will cost energy, binding energy. Binding affinity will suffer.

Choosing a loop length that is compatible with evolutionary history.

1. Search BLAST using your template.
2. Restrict the search to remote species (i.e. not vertbrate, or even not eukaryote.)
3. Look at the indels in the region of interest. Those are the loop lengths that Nature has *allowed*.
4. Choose conserved positions as anchors.
5. Do a loop search with one of the allowed loop lengths.
6. Select the best loop based on E_{vdw} and E_{hb}
7. Mutate side chains.



Non-vertebrate sequences for 4OUL homologs

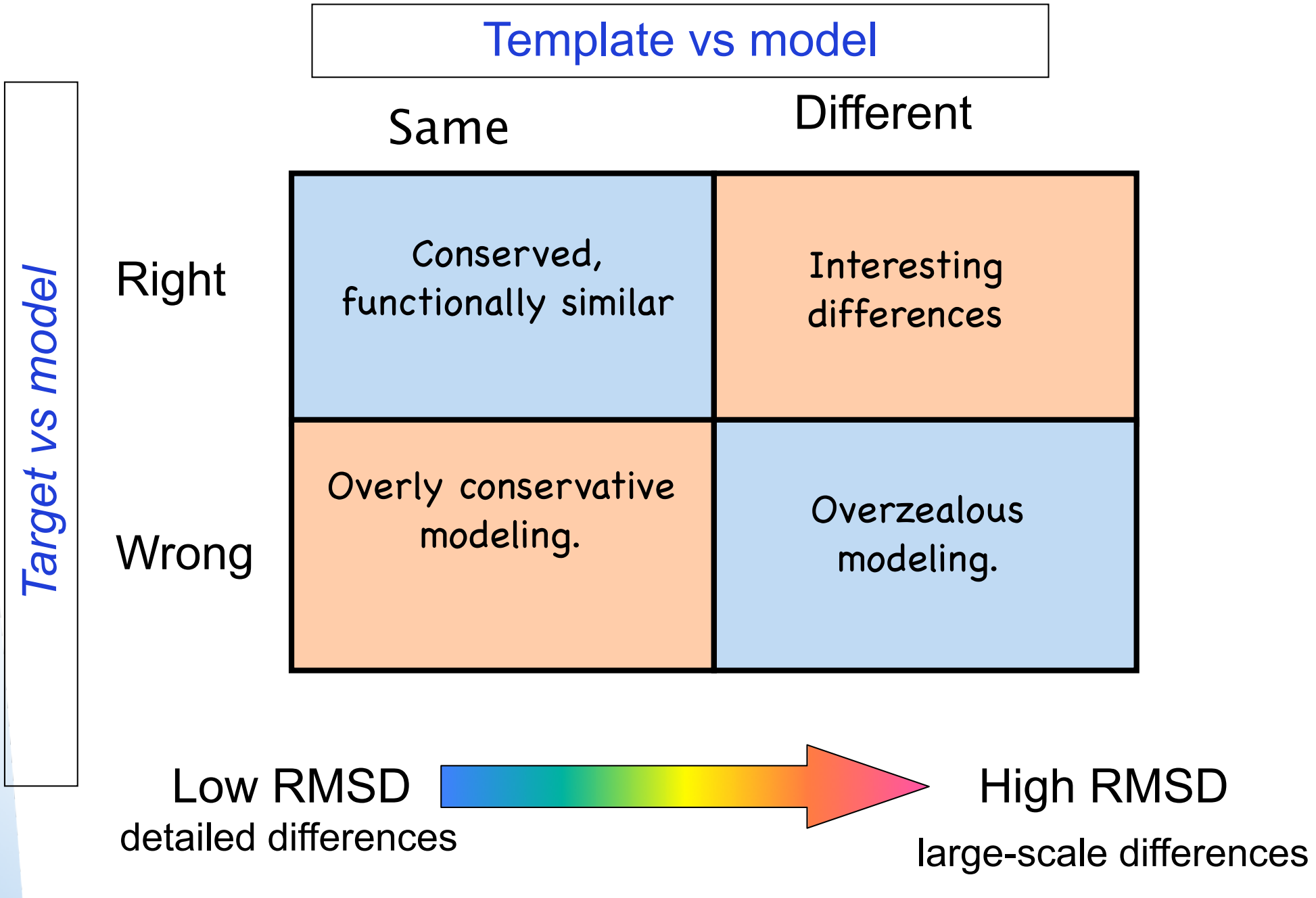

Tag
Chain
4OUL
1: 4OUL.A
2: 4OUL.B
3: 4OUL.C

Validation of your model

- You can never ***know*** if the model is right.
- You can only ***know*** when the model is wrong.
- When you are "done" with a model, check:
  - Bond distances, bond angles, D-amino acids, cis-peptides, clashes
  - H-bonding, especially buried unsatisfied donors/acceptors.
  - Buried charges without counter-ions.
  - Excessive exposed hydrophobics
  - Ramachandran outliers. Positive φ angle not in a glycine.
  - Buried cavities or deep pockets.

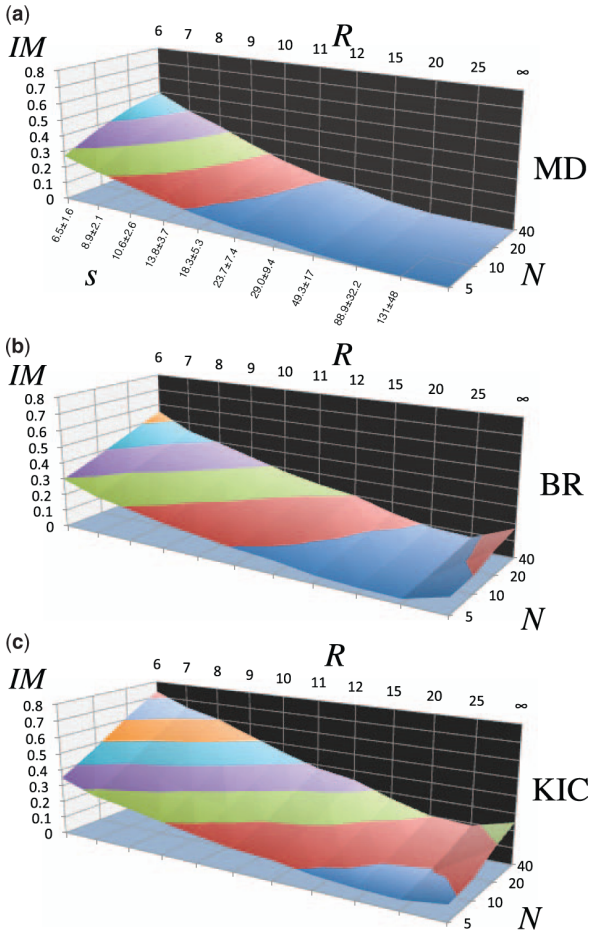
ELIMINATE ALL REASONS TO DISBELIEVE THE MODEL.

Comparing model to template



Stay close to the template!

Studies show loop searches (KIC method) , short molecular dynamics (MD) and monte carlo (backrub motions, BR) fail to sample the true backbone structure, more often make things *worse*.



Criterea for improvement (IM)
Template 1e40A (magenta tubes) and homolog target 1bglB (orange tubes) superimposed with a minimized, diversified *de novo* structure (thin gray string) based on the template. If the gray string more closely resembles target, then we say the method locally improved the model.

Fig. 5 Improvement (IM) of local substructures starting with template, using three methods. Small regions (R) can improve with many tries (N)

Bioinformatics, Volume 32, Issue 10, 15 May 2016, Pages 1454–1461, https://doi.org/10.1093/bioinformatics/btw001

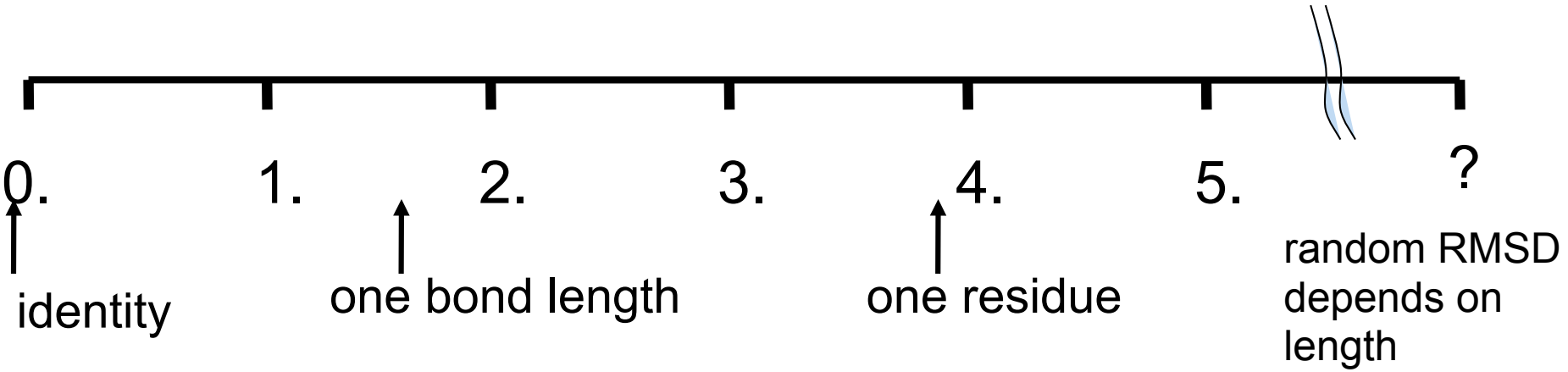
OXFORD UNIVERSITY PRESS

Cartesian coordinate differences: RMSD

- RMSD = root mean square deviation|

$$\sqrt{\frac{\sum_{i=1,N}\left(\vec{x}_i - \vec{y}_i\right)^2}{N}}$$

By far, the most widely used and accepted metric for structural difference.



Confidence

Confidence= the estimated probability of being right.

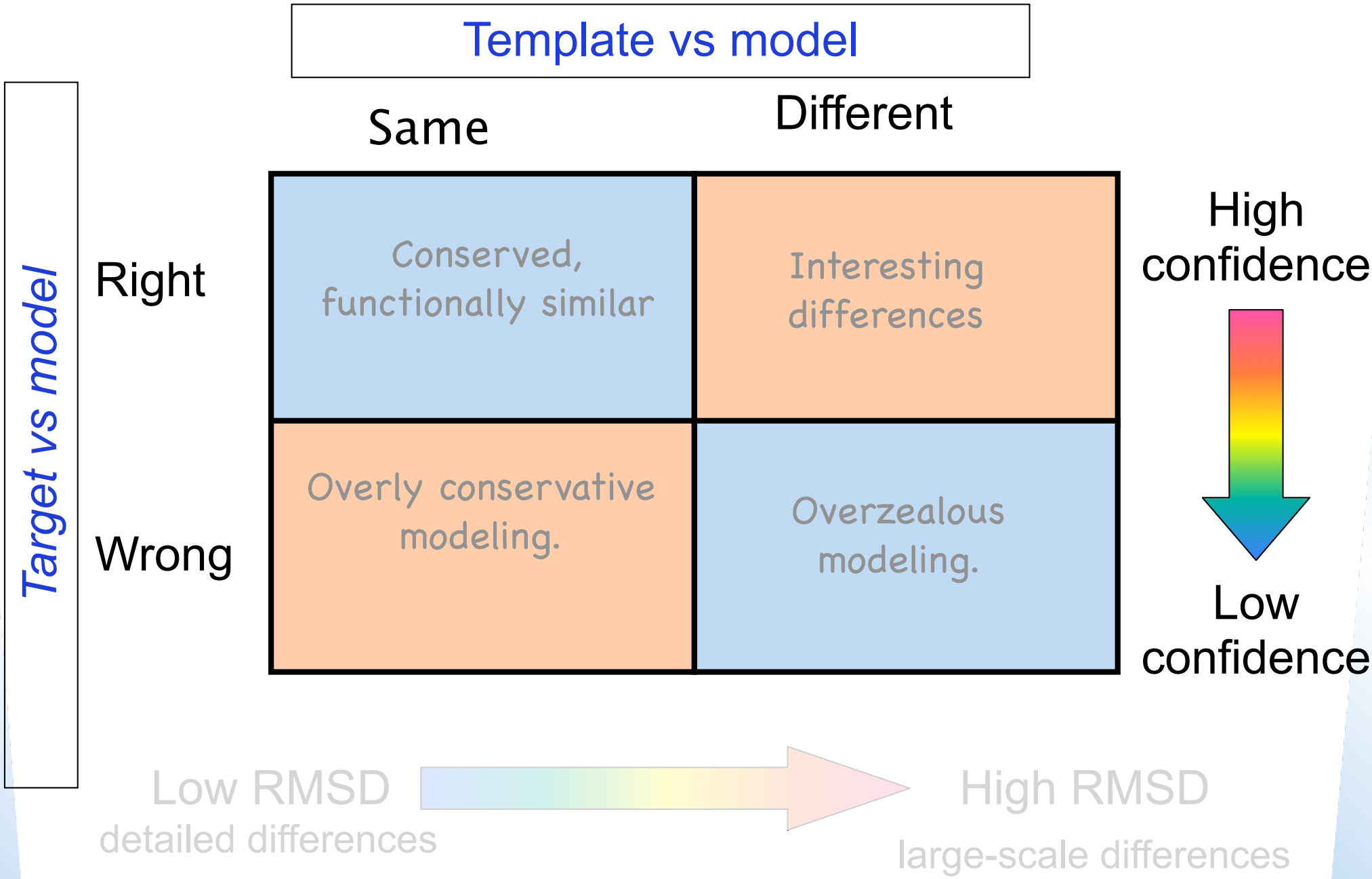
Physics-based confidence estimate:

Based on **modeling experience**, knowledge of **stereochemistry**, **function**, other factors, not statistics. Case specific.

Knowledge-based confidence estimate:

Based on **statistics** of known structures and repeated modeling experiments. **Empirical**, not theoretical. Not specific to one case.

Confidence *should* measure correctness



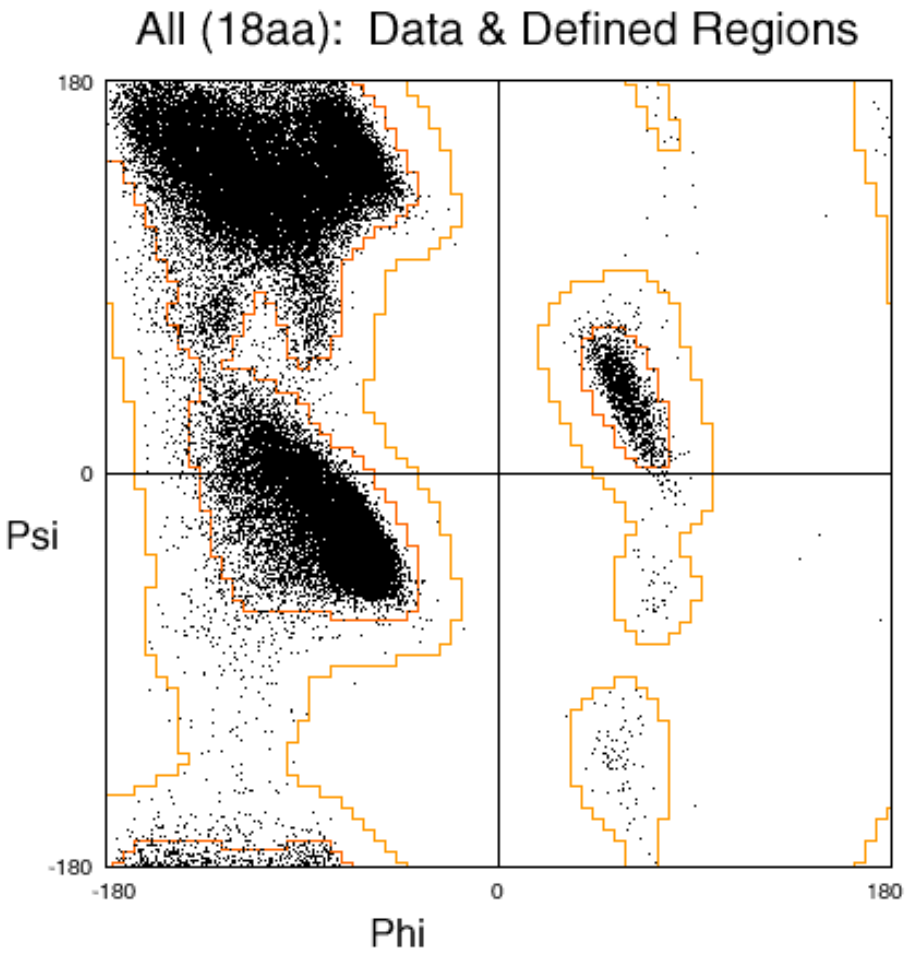
Knowledge-based statistics: Ramachandran allowed regions

- Check for other amino acids outside the allowed regions.
- If it is an outlier, is it conserved? Then it's real.

Remedies for <u>suspicious</u> outliers:

(1) energy minimize with restraint

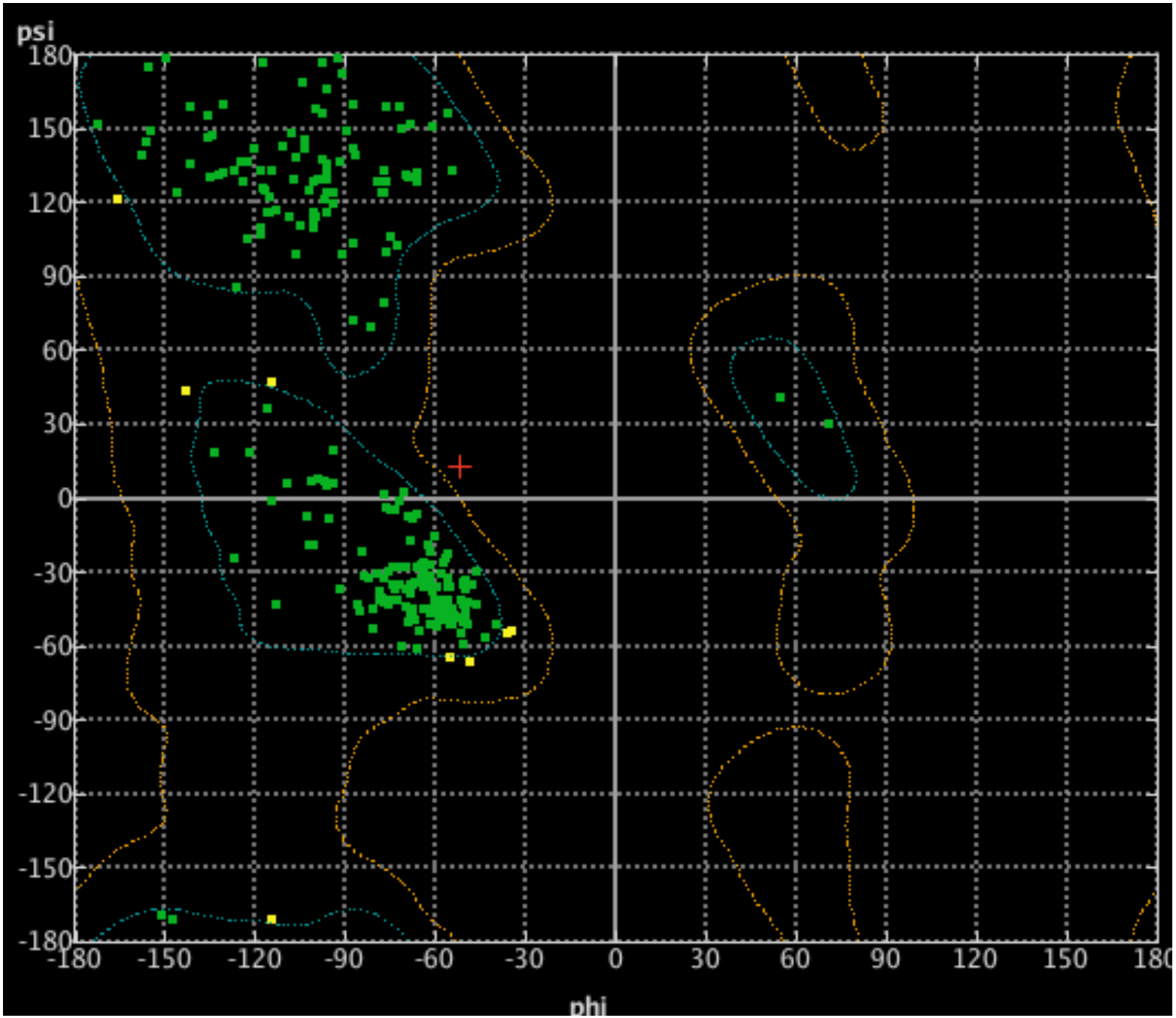
(2) Ignore it. Outliers happen.
But watch out. Too many outliers makes the whole model suspect...



Courtesy of Jane & David Richardson

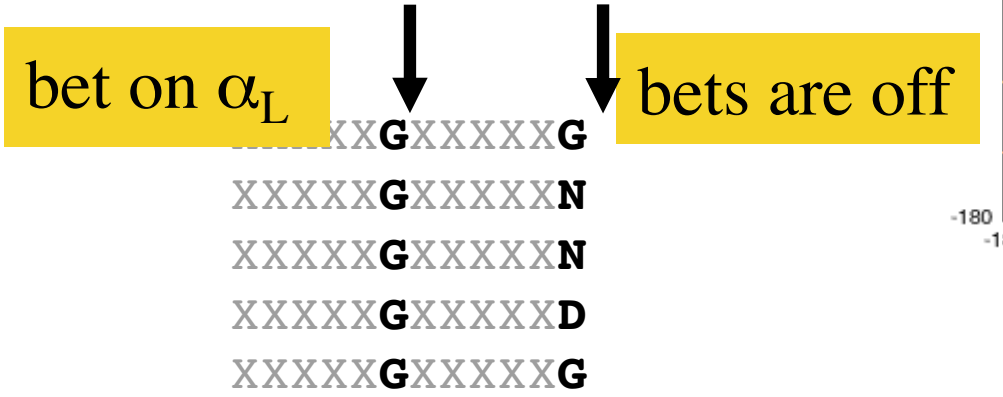
kinemage.biochem.duke.edu

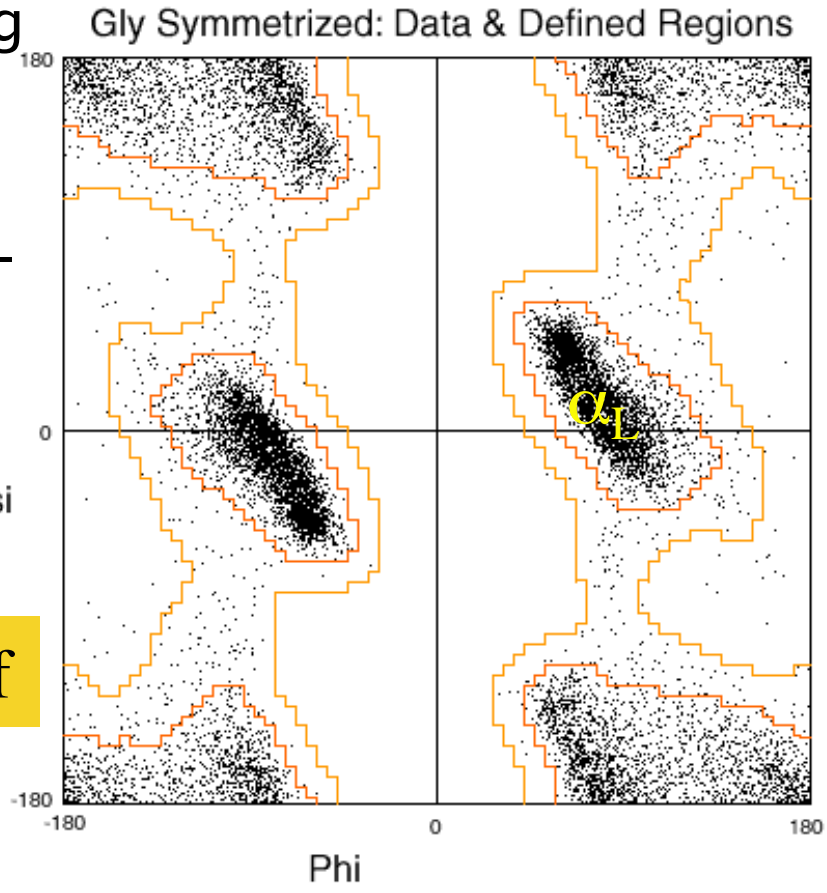
Ramachandran plot: outliers should be rare



Knowledge-based confidence: conserved glycines are probably

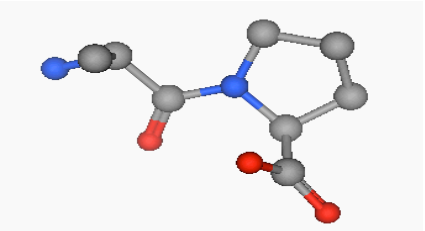
- Glycines are allowed in a wider Ramachandran region, including the "α_L" region.
- If glycine is <u>conserved</u>, you can bet it is is in one of the glycine-only zones. If not conserved, then it must remain in the standard Ramachandran zones.





Courtesy of Jane & David Richardson

Knowledge-based confidence: Proline phi angle always ≈-60°



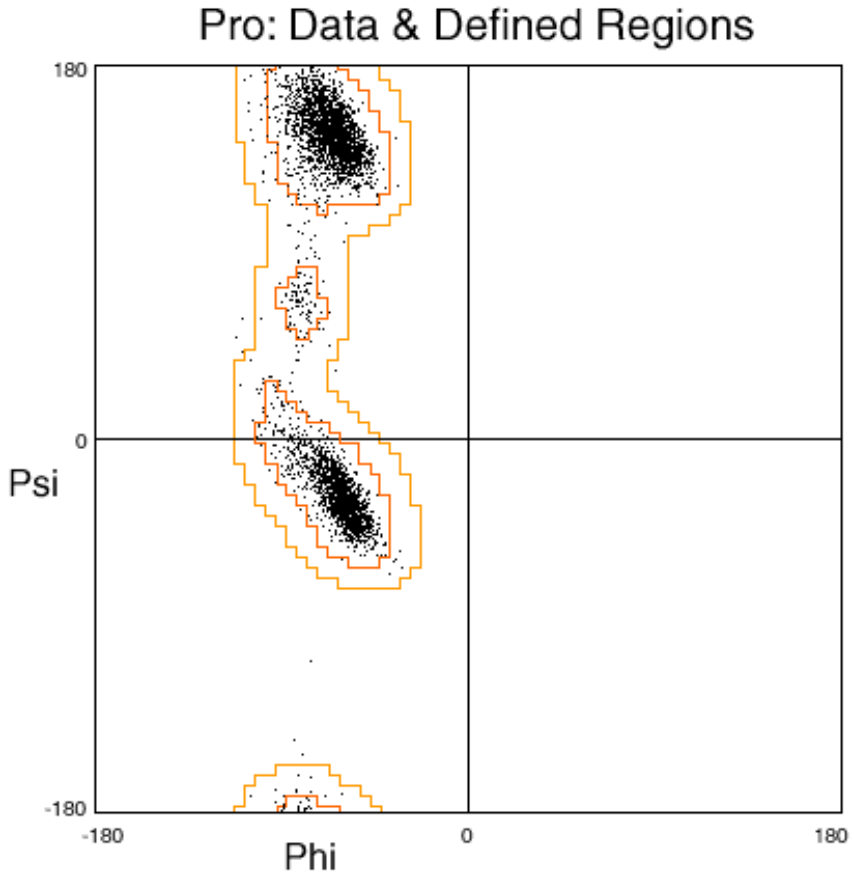
- Check for impossible phi angles at Proline positions.

If you find one, there are two possible remedies

(1) energy minimize it away

(2) re-align the Proline.

never leave it like that.



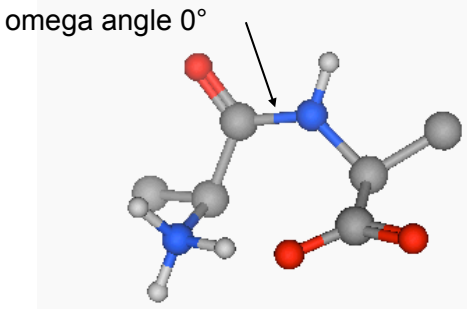
Courtesy of Jane & David Richardson

Knowledge-based confidence: cis peptide bond at X-Pro

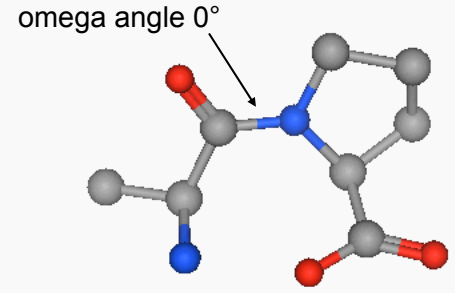
- "cis peptides" : ω (omega) torsion angle may only be 180° or 0° (because of double-bond character), but 0° is highly disfavored (and therefore rare!) unless the residue following the peptide bond is a Proline. Why is this true?
- X = the residue before Pro. X = big (F,Y,W) favors the *trans* state.

cis without Pro is relatively rare!

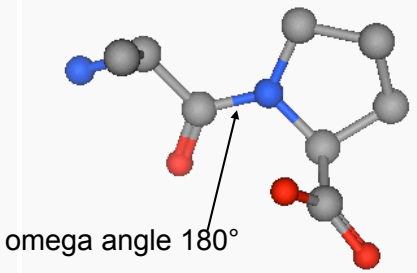
with Pro...
------no big energy difference-----



cis X-Ala



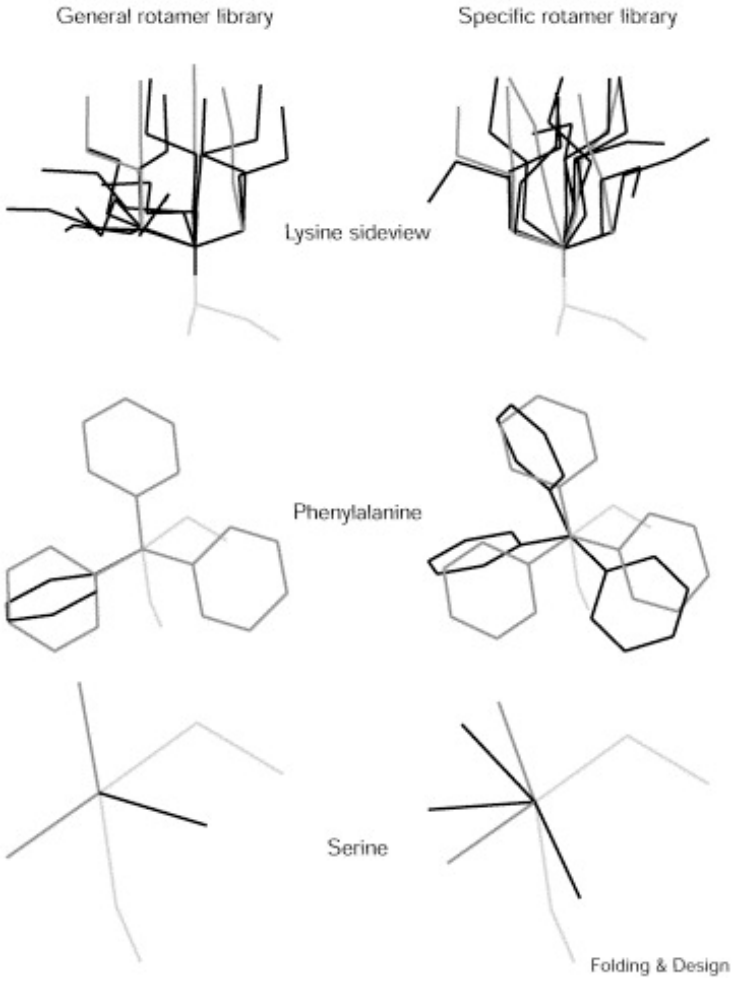
cis X-Pro



trans X-Pro

Knowledge-based statistics: Preferred rotamers

•**Rotamers** are preferred sidechain conformations, found by clustering database sidechains. •**Rotamer** sets (libraries) may be coarse grained or fine grained (pulldown menu in Rotamer explorer). •**Rotamers** have intrinsic energies, due to local interactions.



Compute | Biopolymer | Rotamer explorer

Allows modeler to test rotamer swaps.

Compute | Biopolymer | Protein geometry, rotamer

Finds side chains that need help.

Physics-based confidence: void regions

- Nature abhors a void.

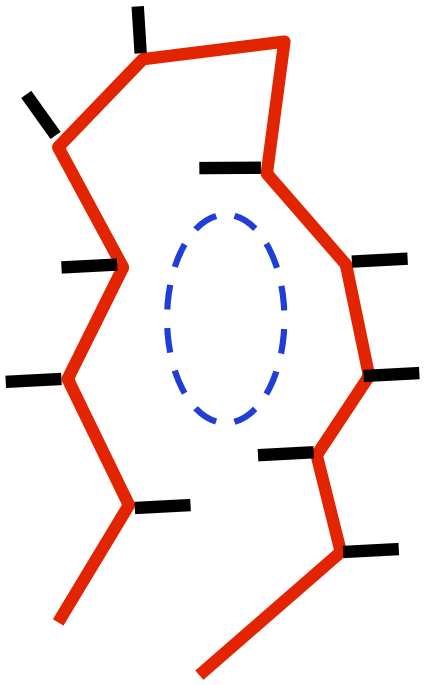
Remedies:

(1) re-pack sidechains with rotamer explorer.

(2) add waters.

(3) energy minimize with distance restraints

(4) Leave it alone. Voids may be functionally important. See (Paredes et al, BMC Bioinformatics 2011)



Physics-based confidence: buried charges are rare, always paired

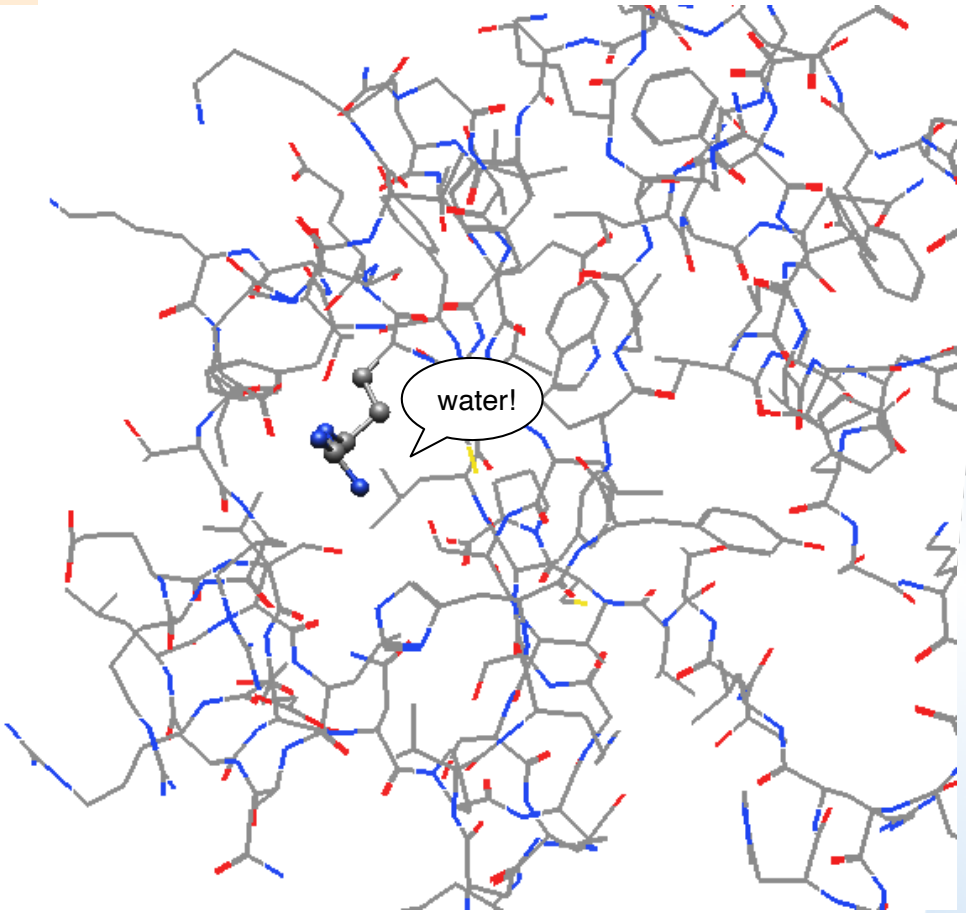
- Charges hate to be de-solvated.

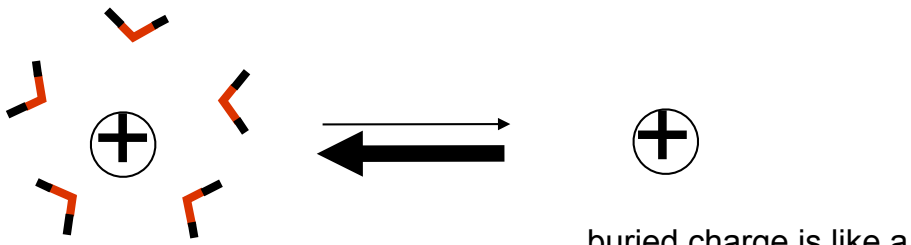
Remedies:

(1) re-pack sidechains. Find a sa
bridge.

(2) re-align. Put it on the outsid

(3) Leave it alone.





water dipoles
delocalize the charge

buried charge is like a
charge in a vacuum.

Summary

A model is as "correct" as it can be if....

- It stays close to the template
- It breaks the fewest possible "rules." (buried H-bonds, voids, phi/psi outliers, etc.)
- Template/model differences are confidently predicted.

Review: validation

- How do you know your model is right?
- How to you know your model is wrong?
- What does confidence mean?
- What is "physics-based" confidence?
- What is "informatics-based" confidence?
- How is a multiple sequence alignment used in protein design?
- How do you generate a "deep" MSA?