

Bioinformatics 2 -- lecture 11

SWISS-MODEL journal club

Evaluating Models

same/different versus right/wrong

confidence -- 2 types

MolProbity

11.1 SWISS-MODEL

AUTOMATED HOMOLOGY
MODELING

SWISS-MODEL

First approach mode

needs to find at least 25% identity

Alignment mode

You provide an alignment of target to template.

Project mode

You provide a DeepView project file containing multiple templates and a model. Useful for second pass modeling.

Algorithm

1) Template selection

Template batches for target domains.

2) target/template alignment

≤5 templates/batch.

High RMSD templates removed.

Placement of indels optimized.

Islands moved to flanks.

3) Model building

Core coordinates averaged.

weighted by seq similarity to target

Loop ensemble created using Constraint Space Programming.

Flanking residues added in if no good loop found.

If no good loop found or length > 10, then a database search is done.

Side chains built iso-sterically.

Uses backbone dependent rotamer library for side chains.

Scores H-bonds, SS-bonds

Energy minimization used only to regularize structure.

4) Evaluation

WhatCheck -- atomic mean force potential -- used to identify problem areas.

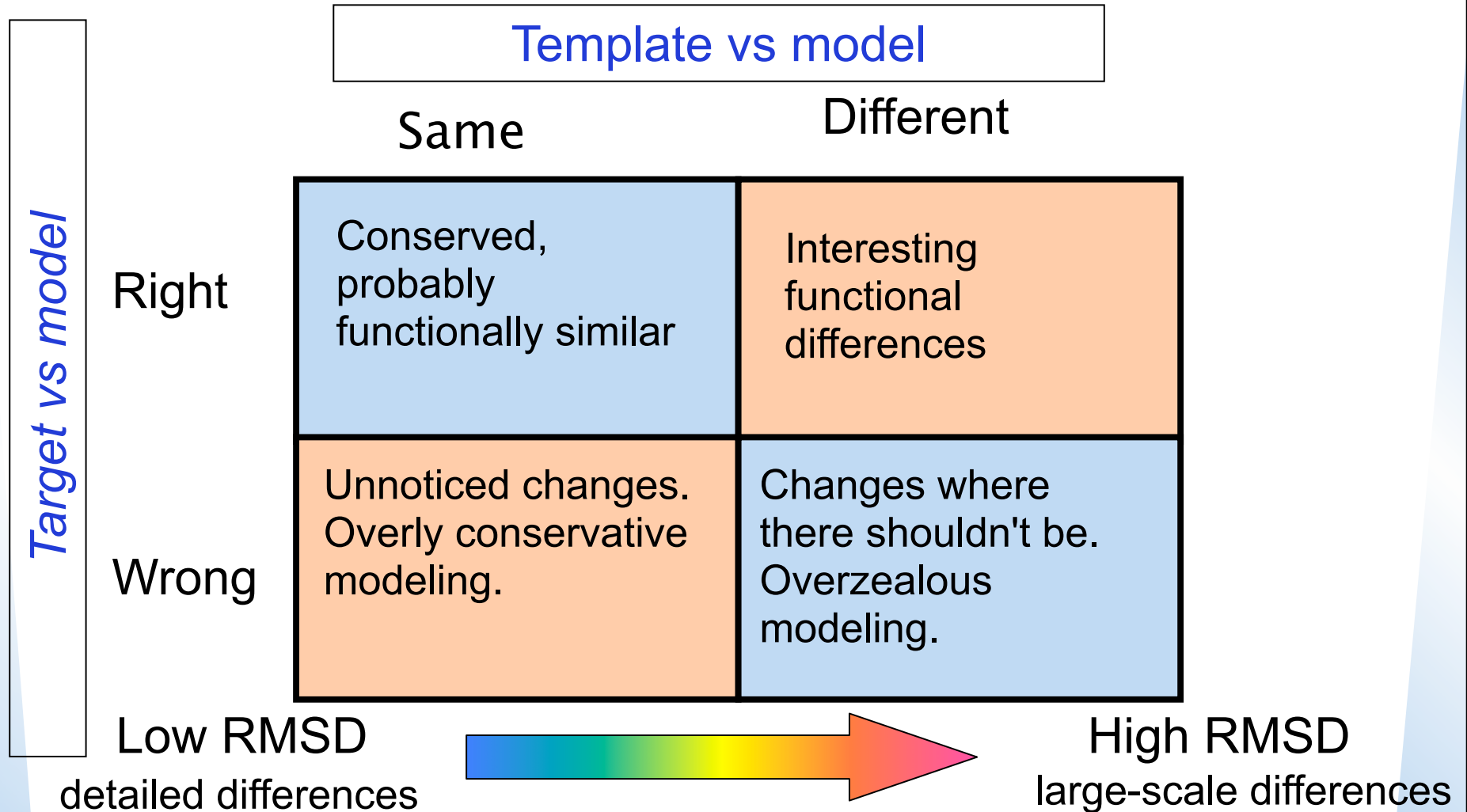
Alignment is seen as the main source of modeling errors.

11.2 EVALUATING MODELS

SAME / DIFFERENT

"Same/different" versus "right/wrong."

There are 2 dimensions to models: model vs template is something we **can** see. Model vs target is something we can't see, but can only **infer**.

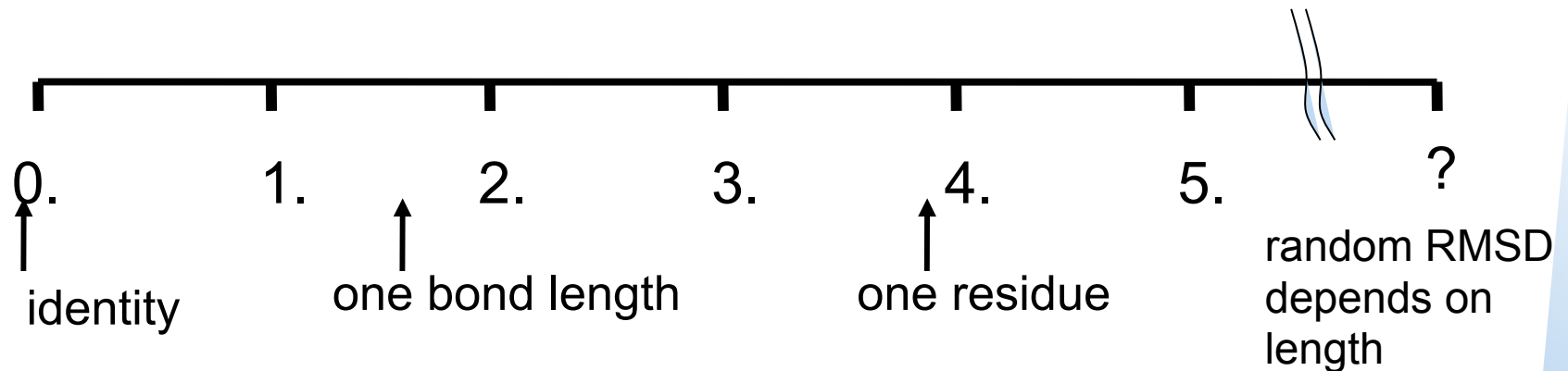


Cartesian coordinate differences: RMSD

- RMSD = root mean square deviation|

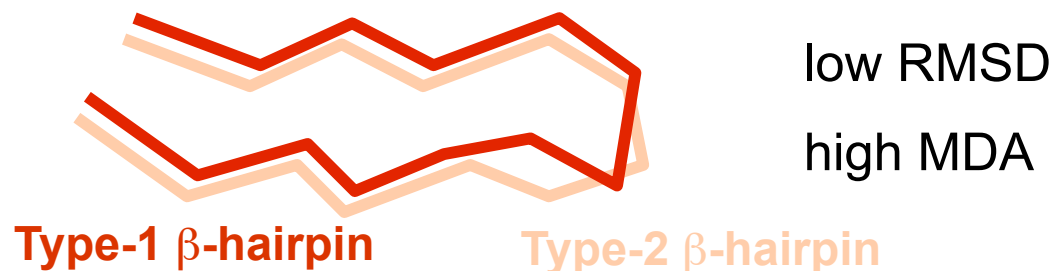
$$\sqrt{\frac{\sum_{i=1,N} (\vec{x}_i - \vec{y}_i)^2}{N}}$$

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



Internal coordinate differences complement Cartesian ones

- Internal coordinates = bond distances, bond angles, torsion angles
- Deviations indicate **local** functional differences.
- MDA = maximum deviation in backbone angles
- Protein segments with $\text{mda} < 120^\circ$ almost always have superimposable structures.
- Superimposable structures do not always have $\text{mda} < 120^\circ$.



Internal coordinate differences: Distance Matrix Error

- **DME = distance matrix error** (average or RMS)
Distance matrix D^x_{ij} = distance from i to j in structure x

$$\sum \left| \begin{array}{c} \triangle \\ D^x_{ij} \end{array} - \begin{array}{c} \triangle \\ D^y_{ij} \end{array} \right| \Bigg/ \begin{array}{c} \diagup \\ \text{N choose 2} \end{array} \qquad \frac{\sum_{i < j=1, N} |D^x_{ij} - D^y_{ij}|}{N(N-1)/2}$$

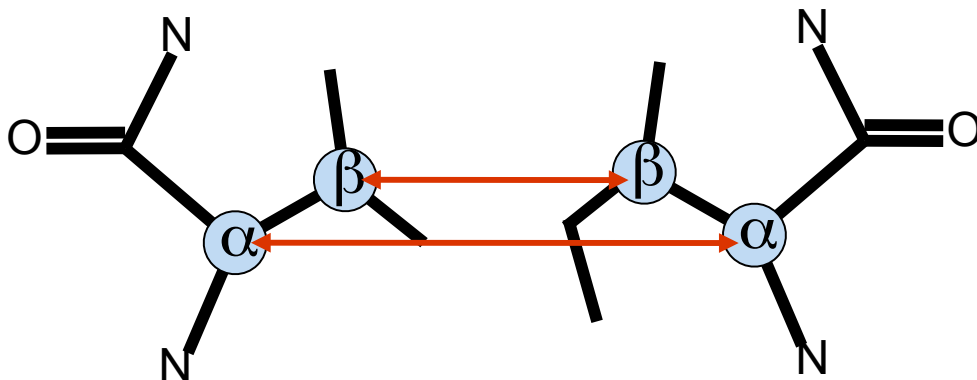
----- or -----

$$\sqrt{\sum \left(\begin{array}{c} \triangle \\ D^x_{ij} \end{array} - \begin{array}{c} \triangle \\ D^y_{ij} \end{array} \right)^2} \Bigg/ \begin{array}{c} \diagup \\ \text{N choose 2} \end{array} \qquad \sqrt{\frac{\sum_{i < j=1, N} (D^x_{ij} - D^y_{ij})^2}{N(N-1)/2}}$$

“N choose 2” = the number of pairs possible with N items = $N(N-1)/2$

DME, continued

- As for any difference metric, we must have an alignment first. The alignment associates D_{ij}^y with D_{ij}^x .
- D_{ij} may be measured from C_α to C_α , or from C_β to C_β . (In the latter case, if the residue is a Gly, then C_α is used instead.)



11.3 EVALUATING MODELS

CONFIDENCE

Confidence should measure correctness

Template vs model

Same

Different

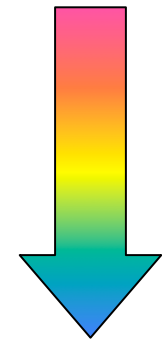
Target vs model

Right

Wrong

Right	Same	Different
Wrong	Different	Same

High confidence



Low confidence

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:

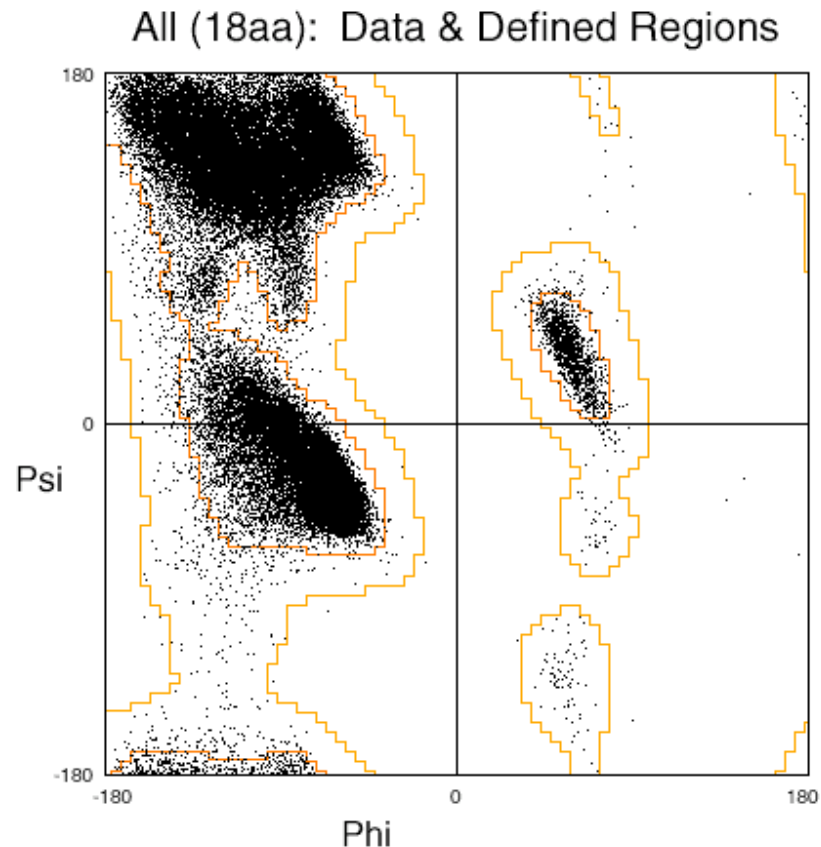
Based on **statistics** of known structures and repeated modeling experiments. **Empirical**, not theoretical. Not specific to one case.

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 suspicious outliers:

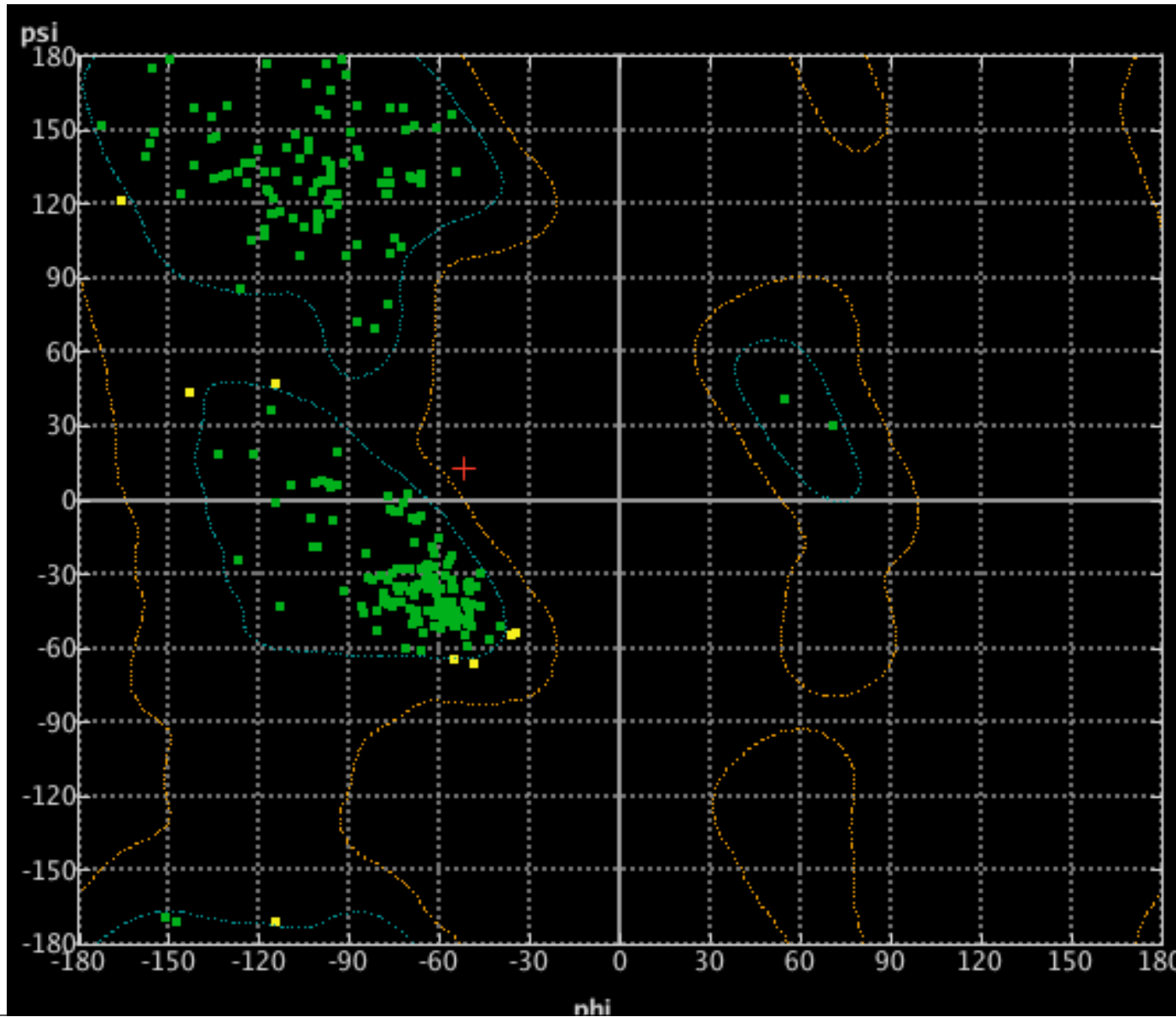
- (1) energy minimize with restraints
- (2) Ignore it. Outliers happen. But watch out. Too many outliers makes the whole model suspect...



Courtesy of Jane & David Richardson

kinemage.biochem.duke.edu

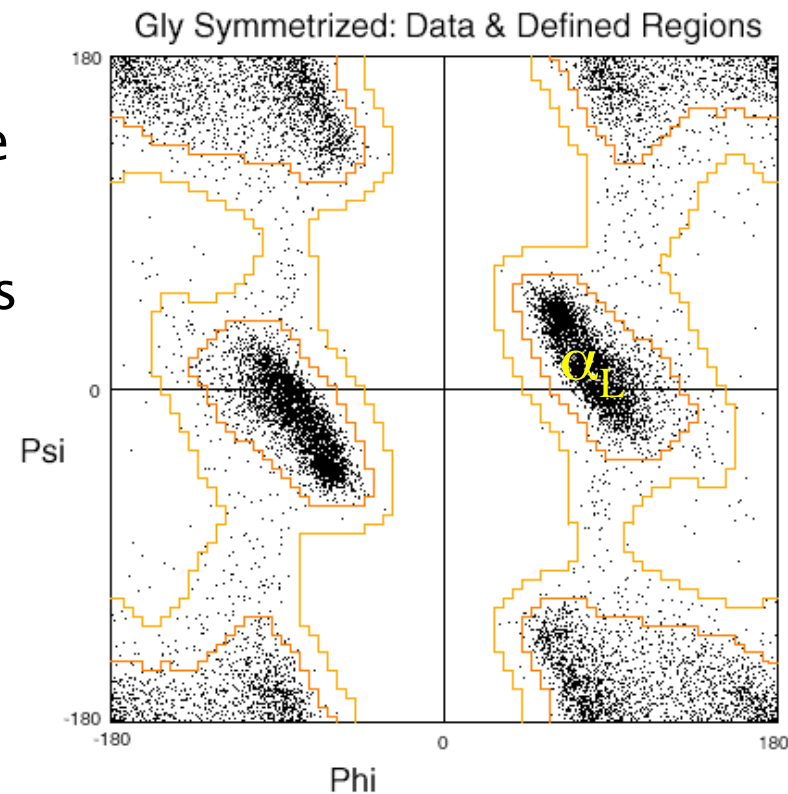
Ramachandran plot: an outlier



Knowledge-based confidence: positive phi angle at Glycine

- Glycines, lacking a C-beta, have a greater allowed Ramachandran region, including the " α_L ", or positive phi, region.
- 2-fold symmetrized statistics for Glycine $\phi\psi$ angles show a more realistic picture of the energy landscape.

XXXXXXGXXXXXG
 bet on α_L XXXXXXGXXXXXN
 XXXXXXGXXXXXN
 XXXXXXGXXXXXD
 bets are off XXXXXXGXXXXXG



Courtesy of Jane & David Richardson

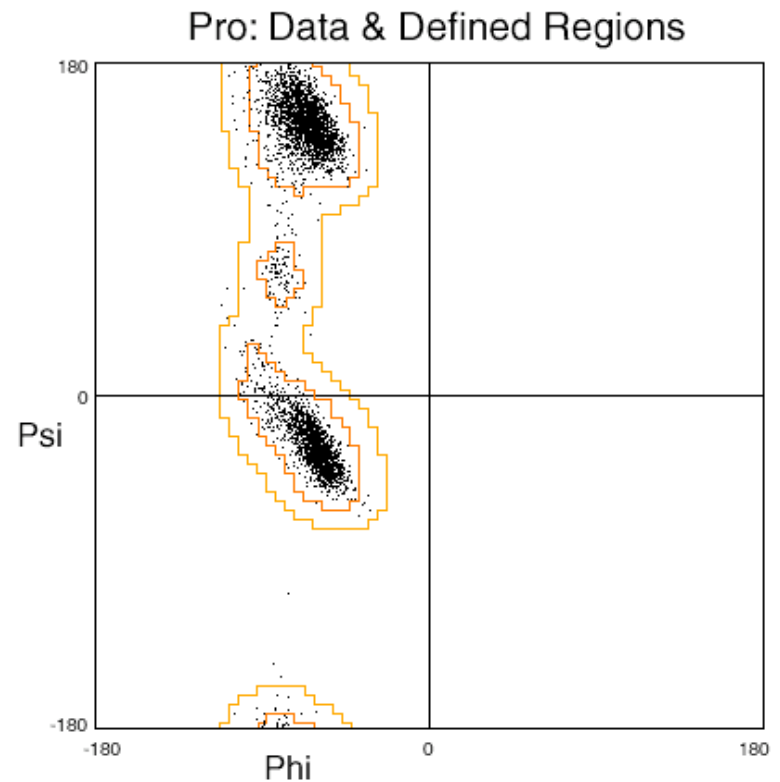
Knowledge-based confidence: Proline phi angle

- 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 Pro.

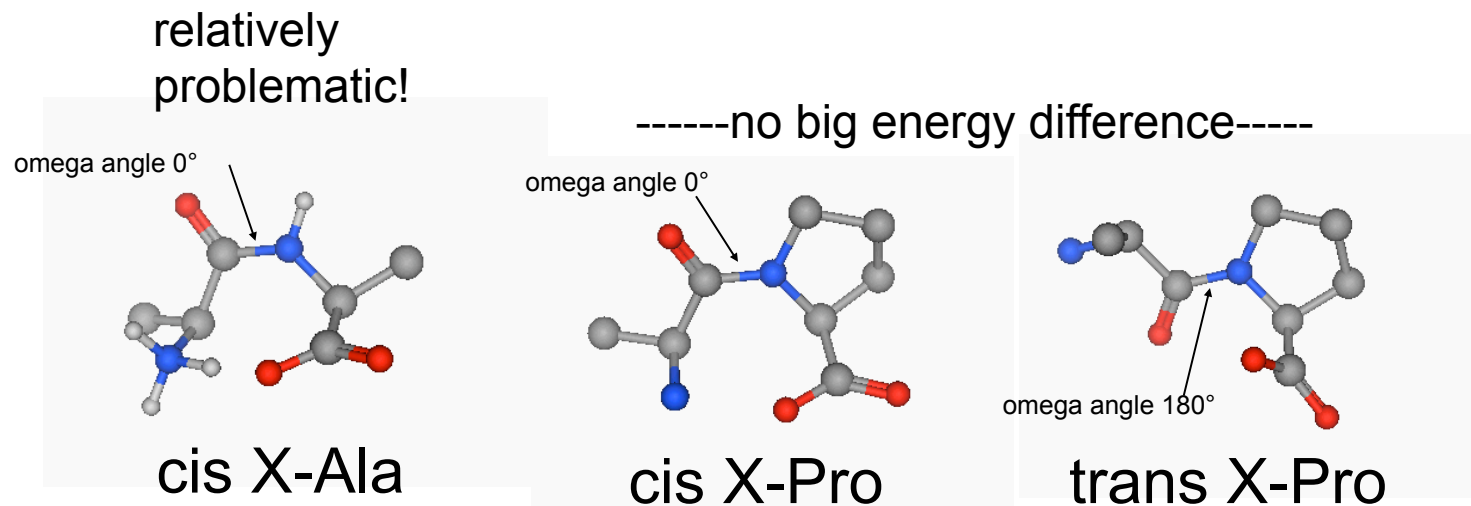
never leave it like that.



Courtesy of Jane & David Richardson

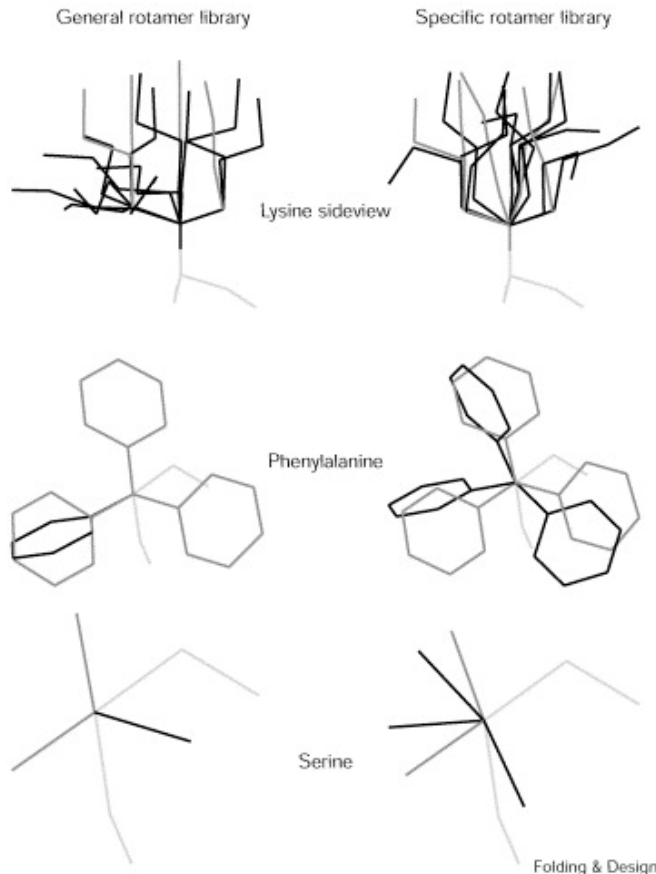
Knowledge-based confidence: cis peptide bond at X-Pro

- “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.



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

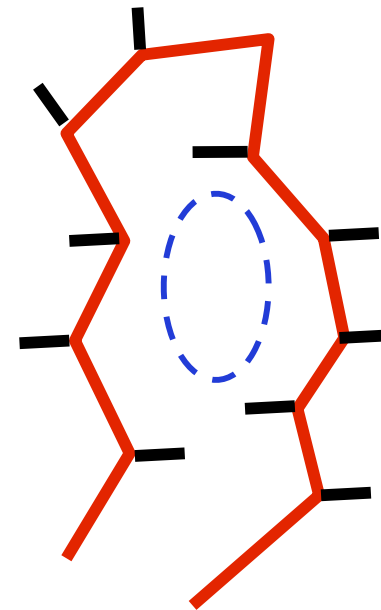
Finds side chains that need help.

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)

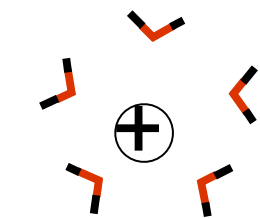
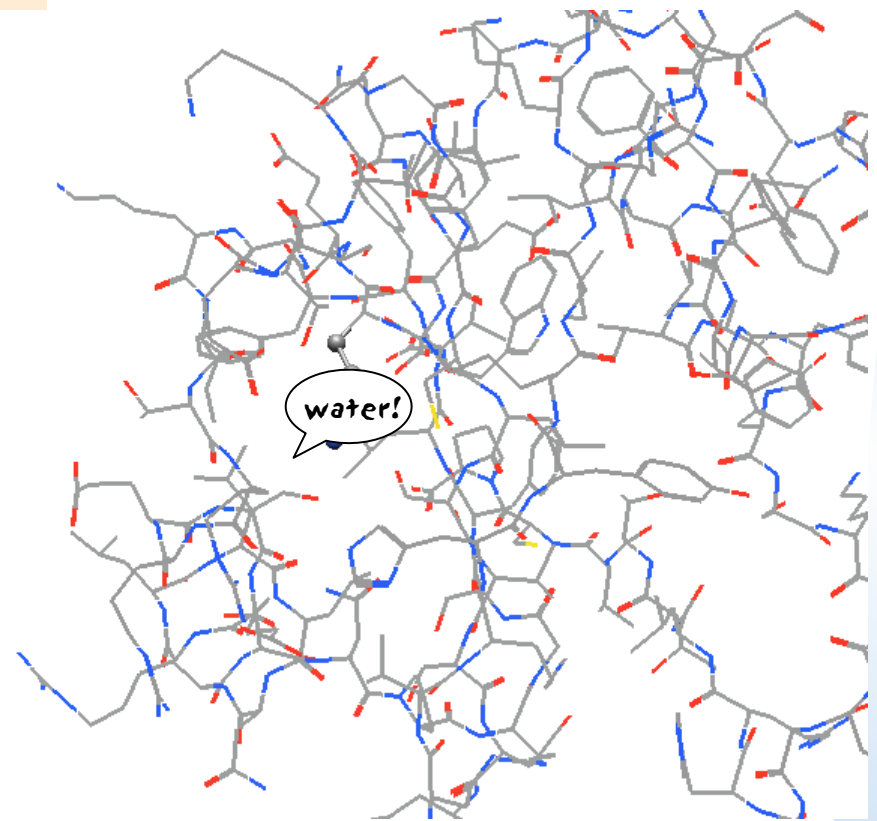


Physics-based confidence: buried charges

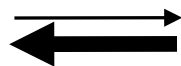
- Charges hate to be de-solvated.

Remedies:

- re-pack sidechains. Find a salt bridge.
- re-align. Put it on the outside.
- Leave it alone.



water dipoles
delocalize the
charge



buried charge is like
a charge in a
vacuum.

11.4 MOLPROBITY

GUIDED TOUR

`molprobability.biochem.duke.edu`

- Automated checker for correctness of a model.