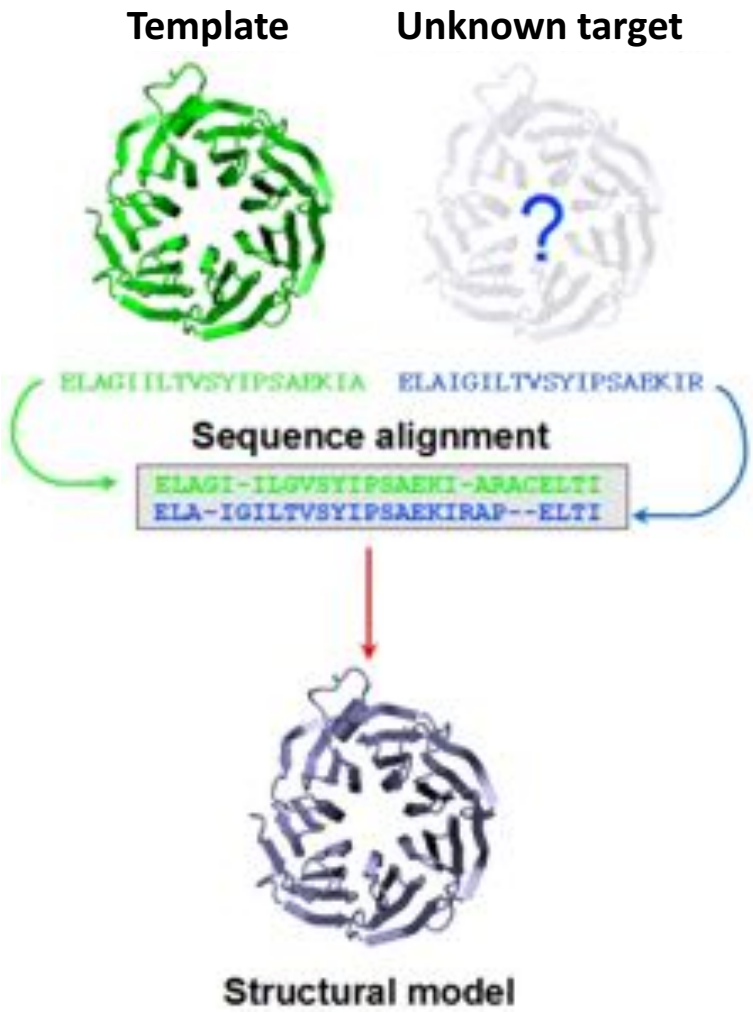
# Homology modeling

Kamilla Kjærgaard Jensen
kamjen@bioinformatics.dtu.dk

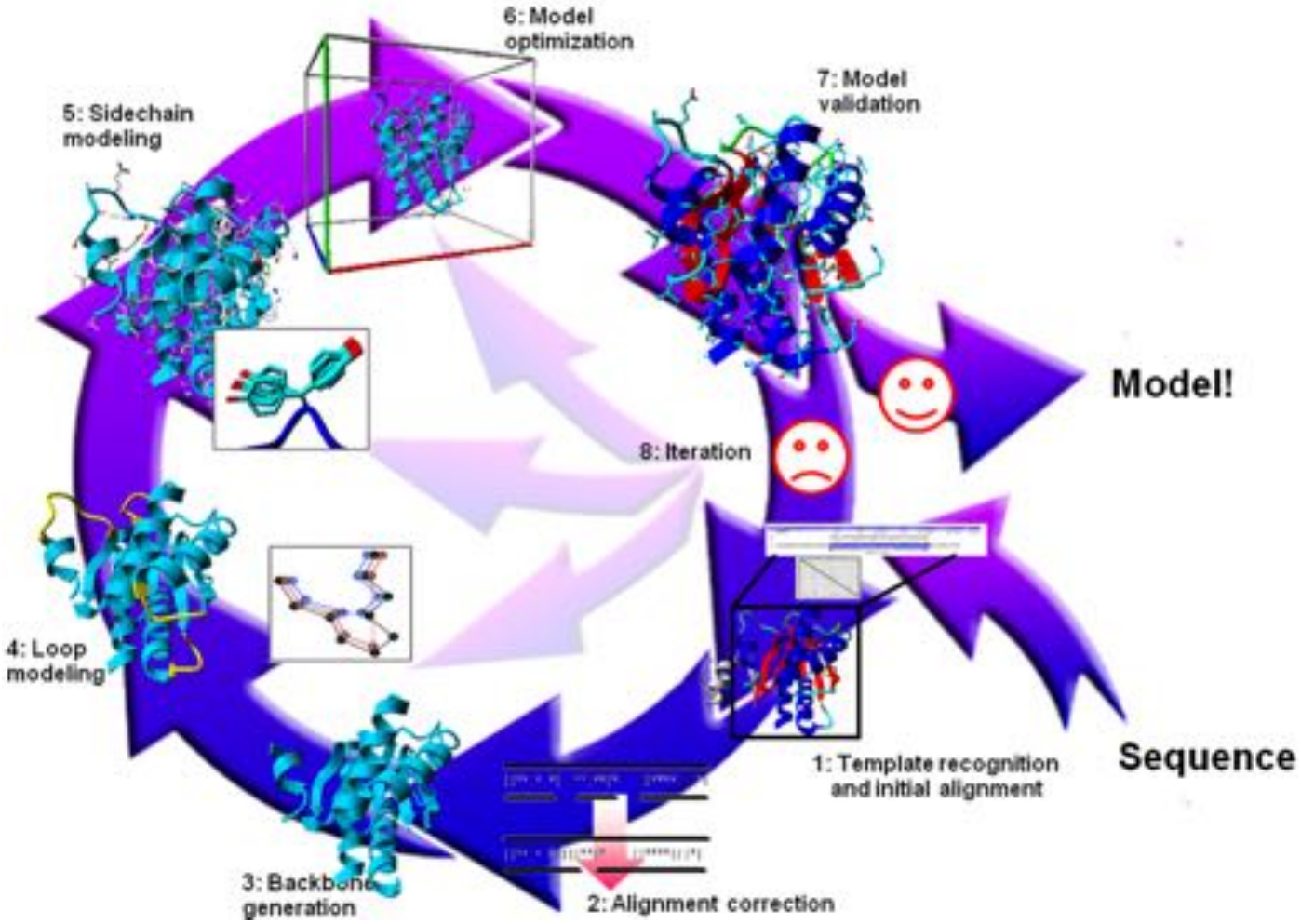
# Homology modeling

- Find **template** structures which are homologous to your **target** sequence
- You need good template and alignment
- Validate the model.
- Adjust expectations and use accordingly!



# Homology modeling

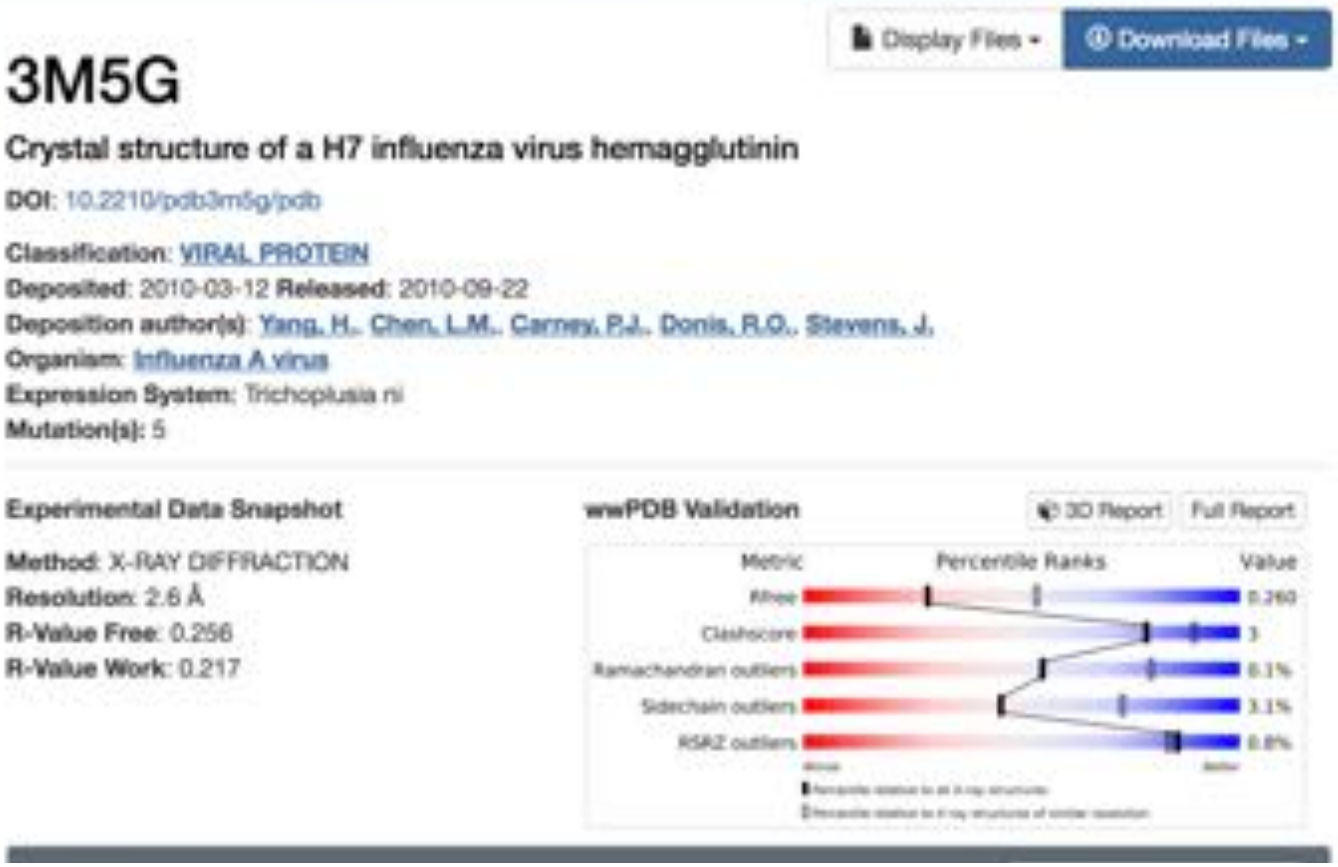
1. Identify template(s)
   - Initial alignment
2. Improve alignment
3. Copy backbone conformation
4. Loop modelling
5. Side chains
6. Model optimization
7. Validation ←



# Template identification

- Structural templates are found in Protein Data Bank (PDB)

- Structure quality
  - Resolution
  - R-free
  - R-work

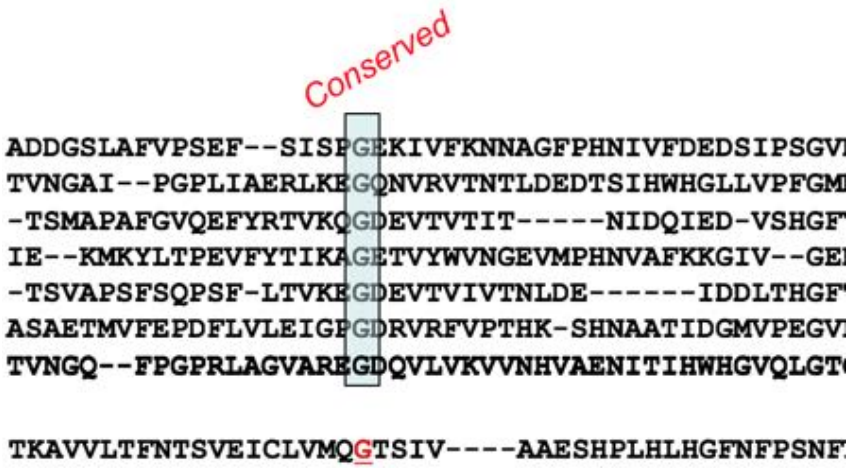


# Template Identification

- Search with sequence
  - Blast (>60%)
  - Psi-Blast (>30%)
  - Fold recognition methods (>10%)

# Improve alignment

- Use biological information
  - Enzyme / transmembrane



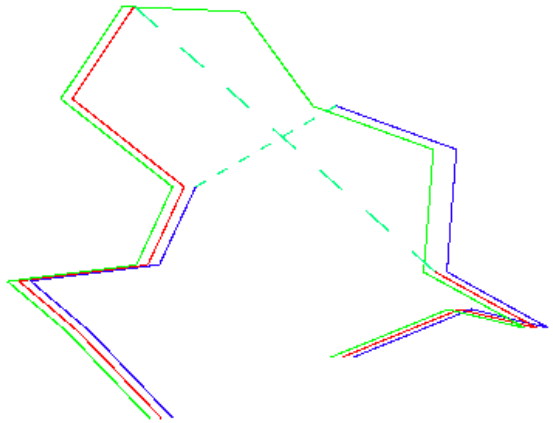

- Active site/motifs
  - Identify residues that are important and use these to make the aligned

# When to improve the alignment?

1) no deletion where N- and C-terminal is far apart in the template structure

```
 1   2   3   4   5   6   7   8   9   10  11  12  13  14
PHE ASP ILE CYS ARG LEU PRO GLY SER ALA GLU ALA VAL CYS   Template
PHE ASN VAL CYS ARG THR PRO --- --- --- GLU ALA ILE CYS   1st alignment
PHE ASN VAL CYS ARG --- --- --- THR PRO GLU ALA ILE CYS   2nd alignment
```

| | Alignment score |
|---|---|
| 1st alignment | Small value |
| 2nd alignment | Big value |



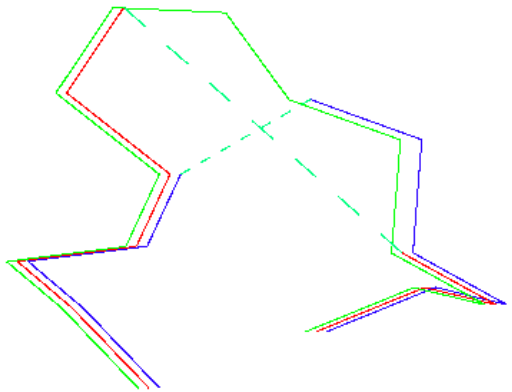
From ”Professional Gambling” by Gert Vriend
http://www.cmbi.kun.nl/gv/articles/text/gambling.html

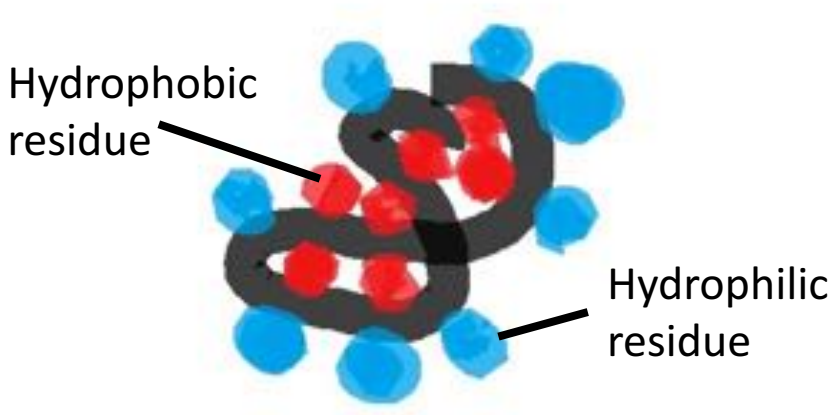
# When to improve the alignment?

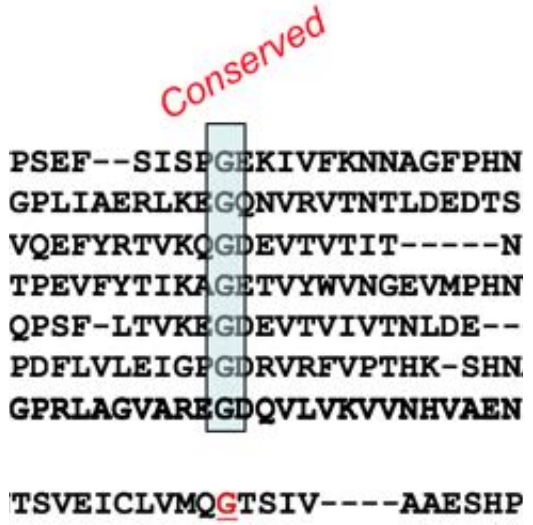
1) no deletion where N- and C-terminal is far apart in the template structure

2) no insertion in the protein core

3) check if conserved residues are aligned

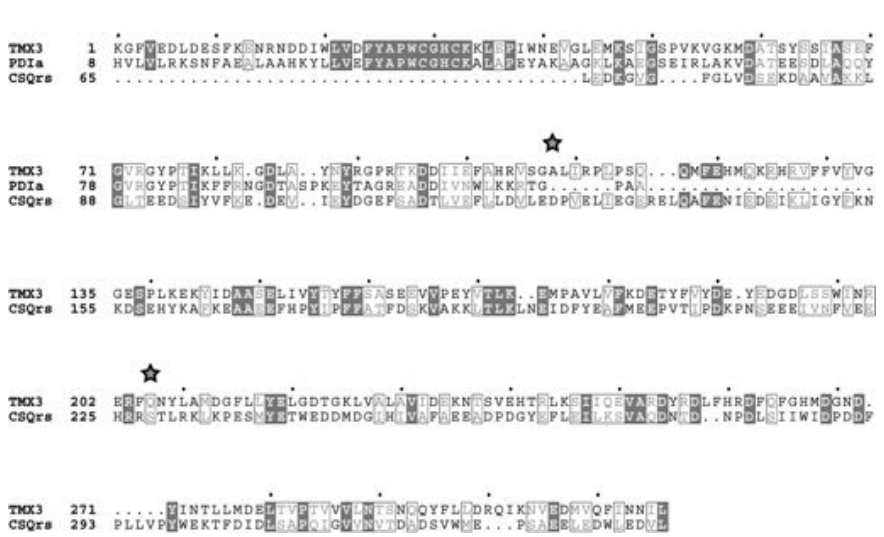


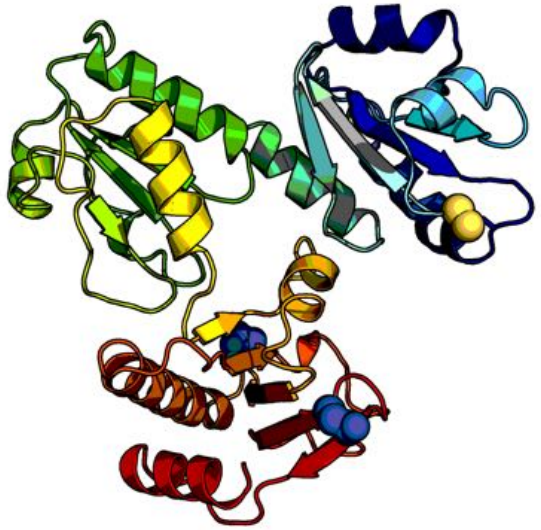


# Backbone Generation

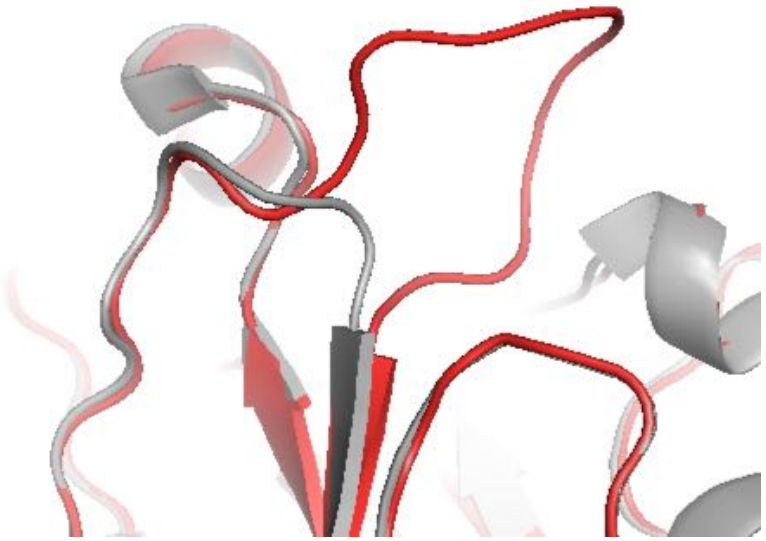
- Generate the backbone coordinates from the template for the aligned regions.
- Several programs can do this, most used is Modeller: http://www.salilab.org/modeller/



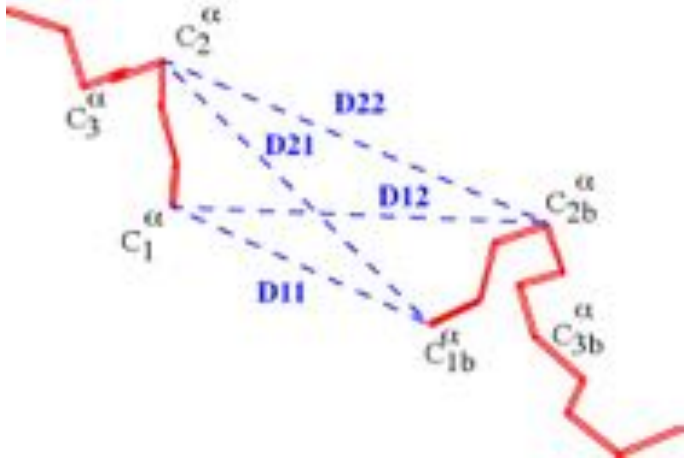



# Loop Modelling

- Knowledge based:
  - PDB search for fragments matching the target sequence (Levitt, Holm, Baker etc.).



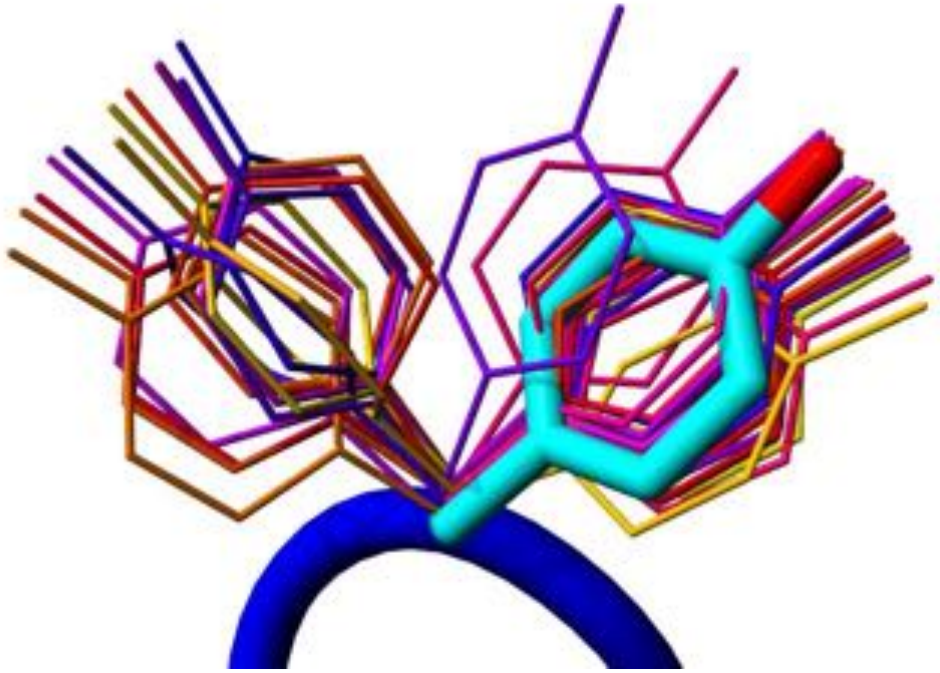
- Energy based:
  - Uses an energy function to evaluate the quality of the loop and minimizes this function by Monte Carlo (sampling) or molecular dynamics (MD) techniques.




http://nook.cs.ucdavis.edu/~koehl/BioEbook/images/Loop_database.jpg

# Side Chains

- Most successful homology methods uses SCWRL by Dunbrack *et al.*:
  - Side chain rotamers are dependent on backbone conformation.
  - SCWRL http://dunbrack.fccc.edu/scwrl4/index.php

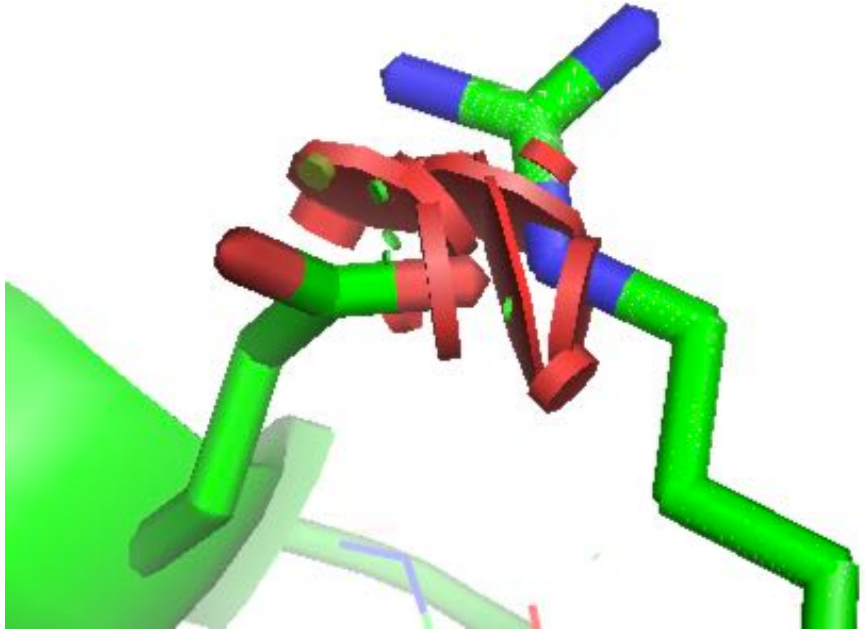


# Model optimization

- To remove atomic clashes across the protein

- Energy minimization
  - *Removes big errors*
    - *Atom clashes*
  - *Introduces small errors*

- *Keep minimization to a minimum.*

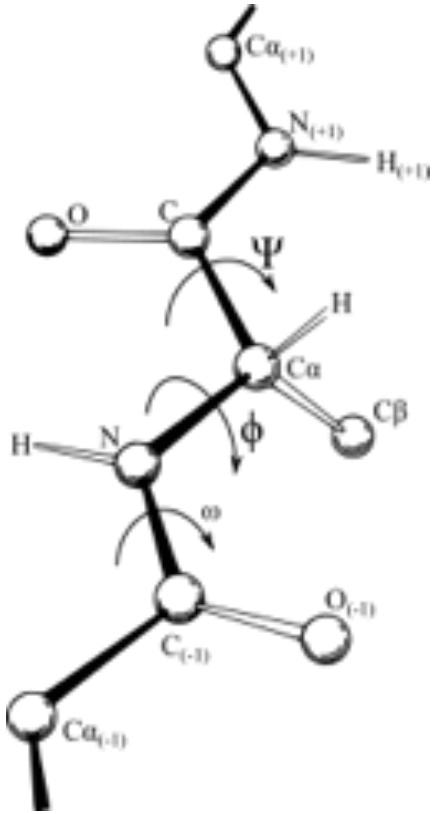


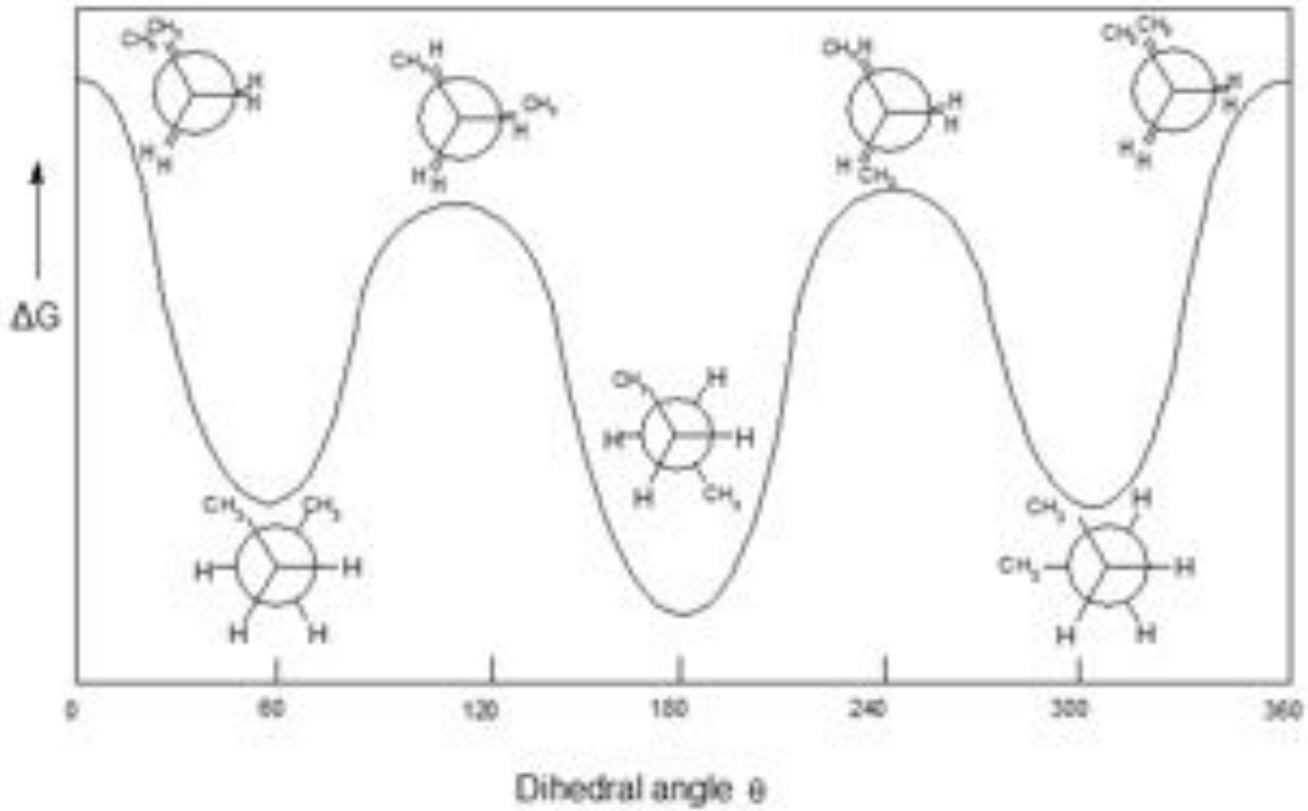
# Model Validation

- ProQ is a neural network-based predictor
  - Structural features → quality of a protein structure.
- Ramachandran plot
  - visualize backbone conformations of the protein
  - Show allowed regions for backbone conformations

# Diheral angels

- In protein the dihedral angles are ψ (psi) and ϕ (phi)
  - they describe the backbone conformation of the protein chain



# Ramachandran plot

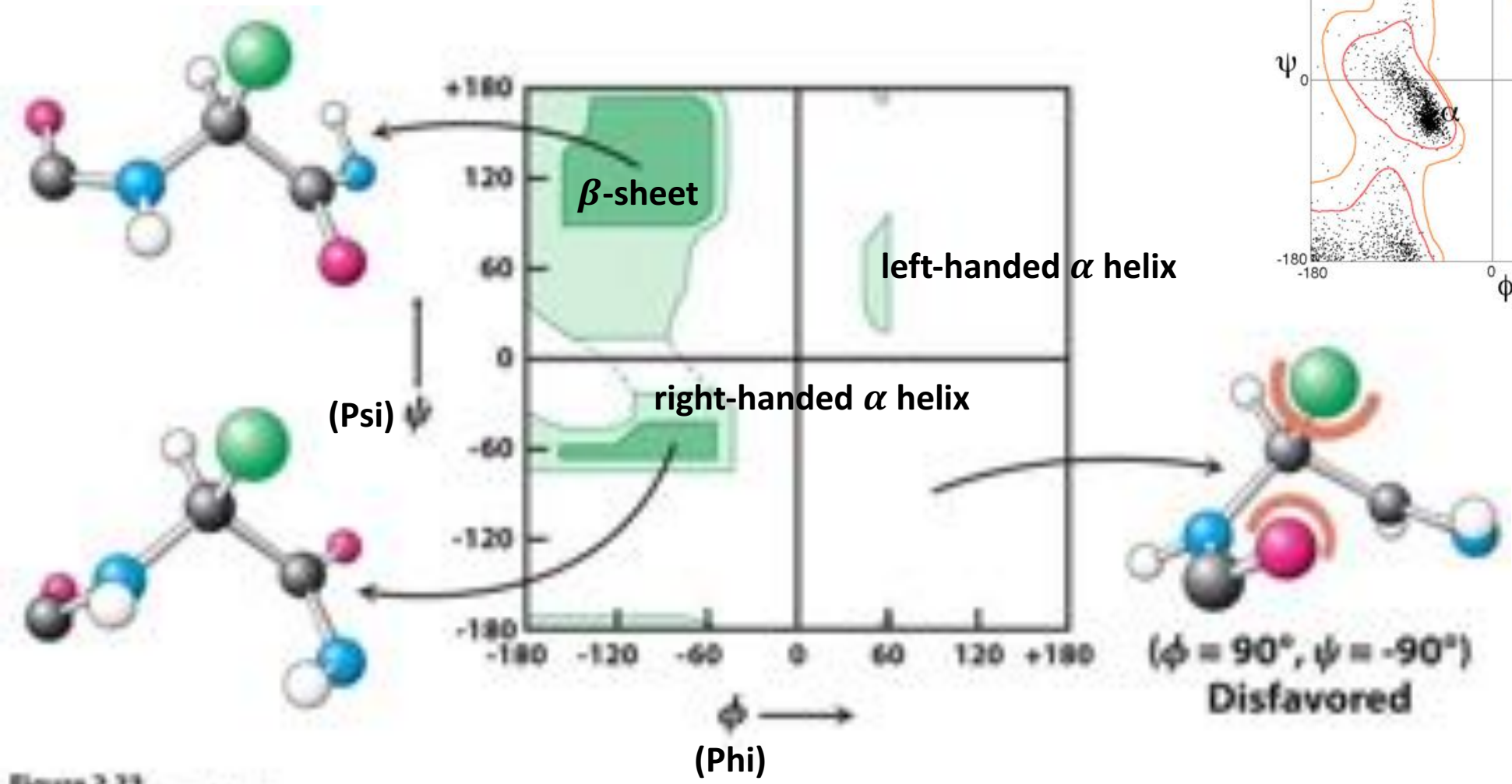
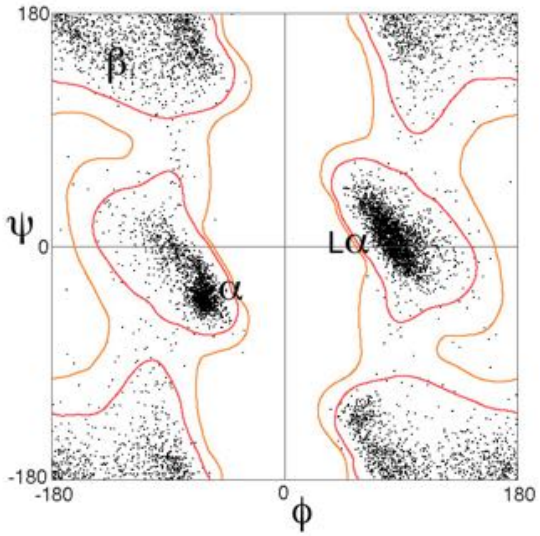
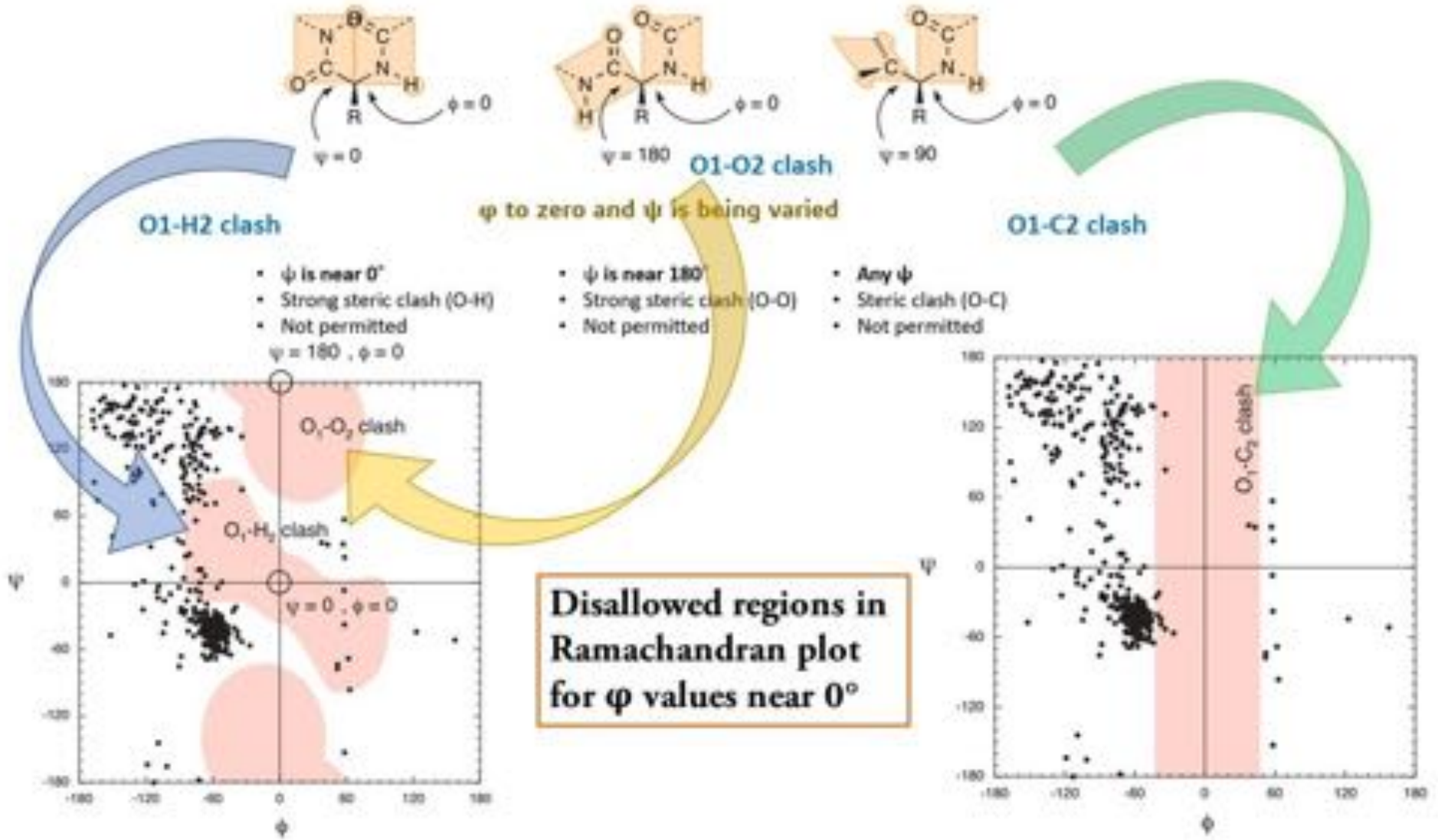



Figure 2.23
Biochemistry, Seventh Edition
© 2012 W. H. Freeman and Company

Glycine

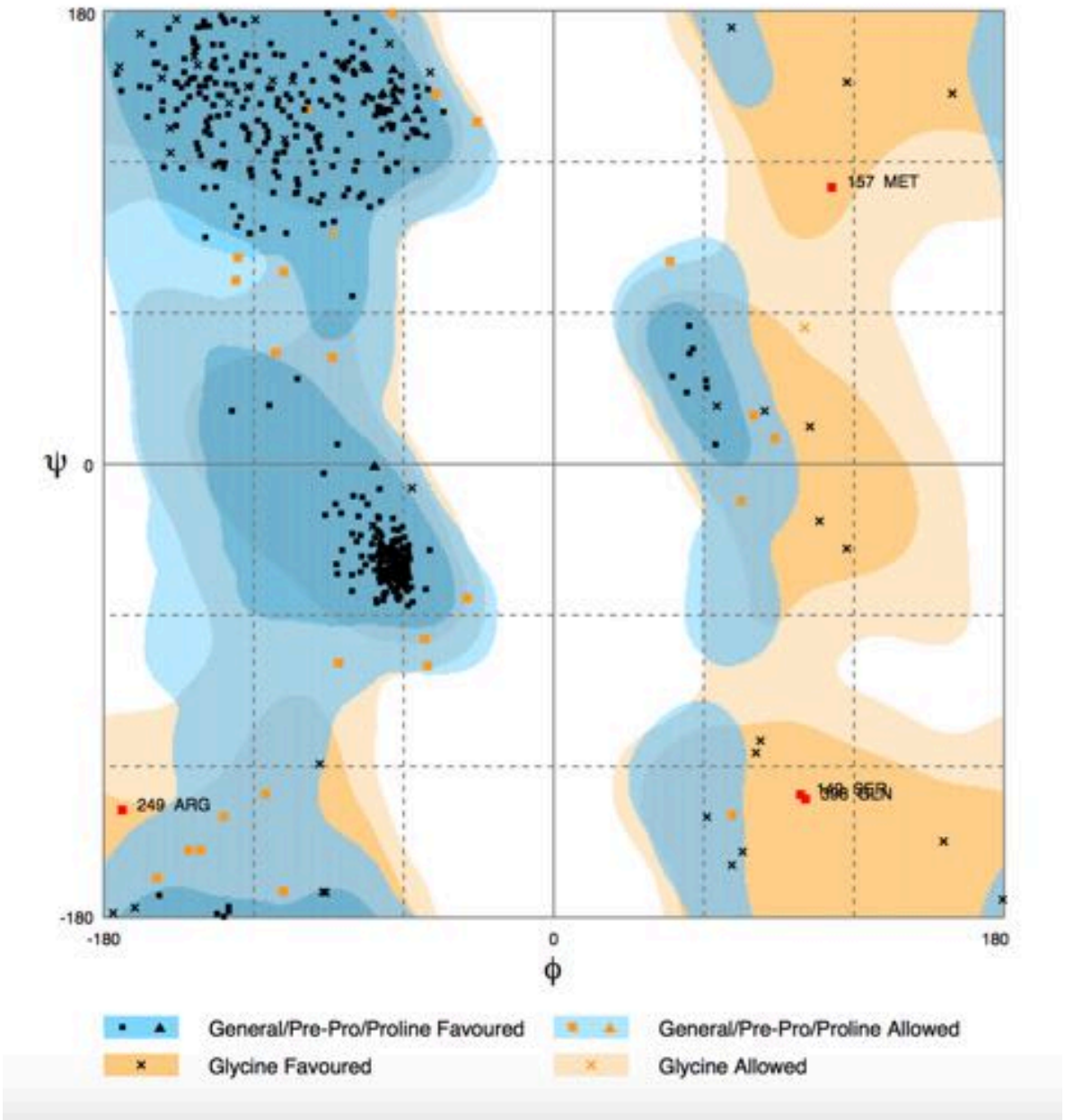


# Ramachandran plot



# Ramachandran plot

- Make Ramachandran plots
  - Using RAMPAGE
  - http://mordred.bioc.cam.ac.uk/~rapper/rampage.php
- Outliers in the Ramachandran plot



# Summary

- Successful homology modelling depends on the following:
  - Template quality
    - Good resolution
  - Alignment
    - add biological information
  - Modelling program/procedure
    - use more than one

- Always validate your final model!

# Homology modeling exercise

- CPH-models
  - Simple to use and very fast
  - Only models the part it is most confident about
  - No manually edit the sequence alignment
- HHpred
  - Fairly simple to use and slower
  - Many options to manually edit the template and sequence alignment
- PyMOL
  - Program to visualize protein structures