

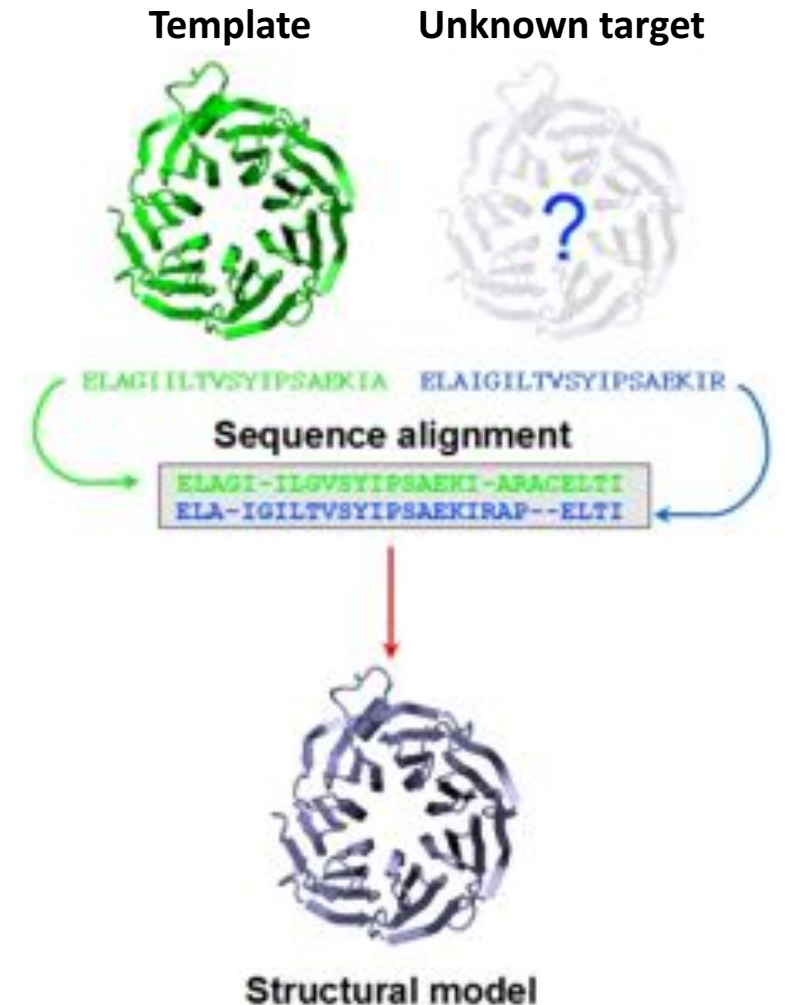
# Homology modeling

Kamilla Kjærgaard Jensen

[kamjen@bioinformatics.dtu.dk](mailto:kamjen@bioinformatics.dtu.dk)

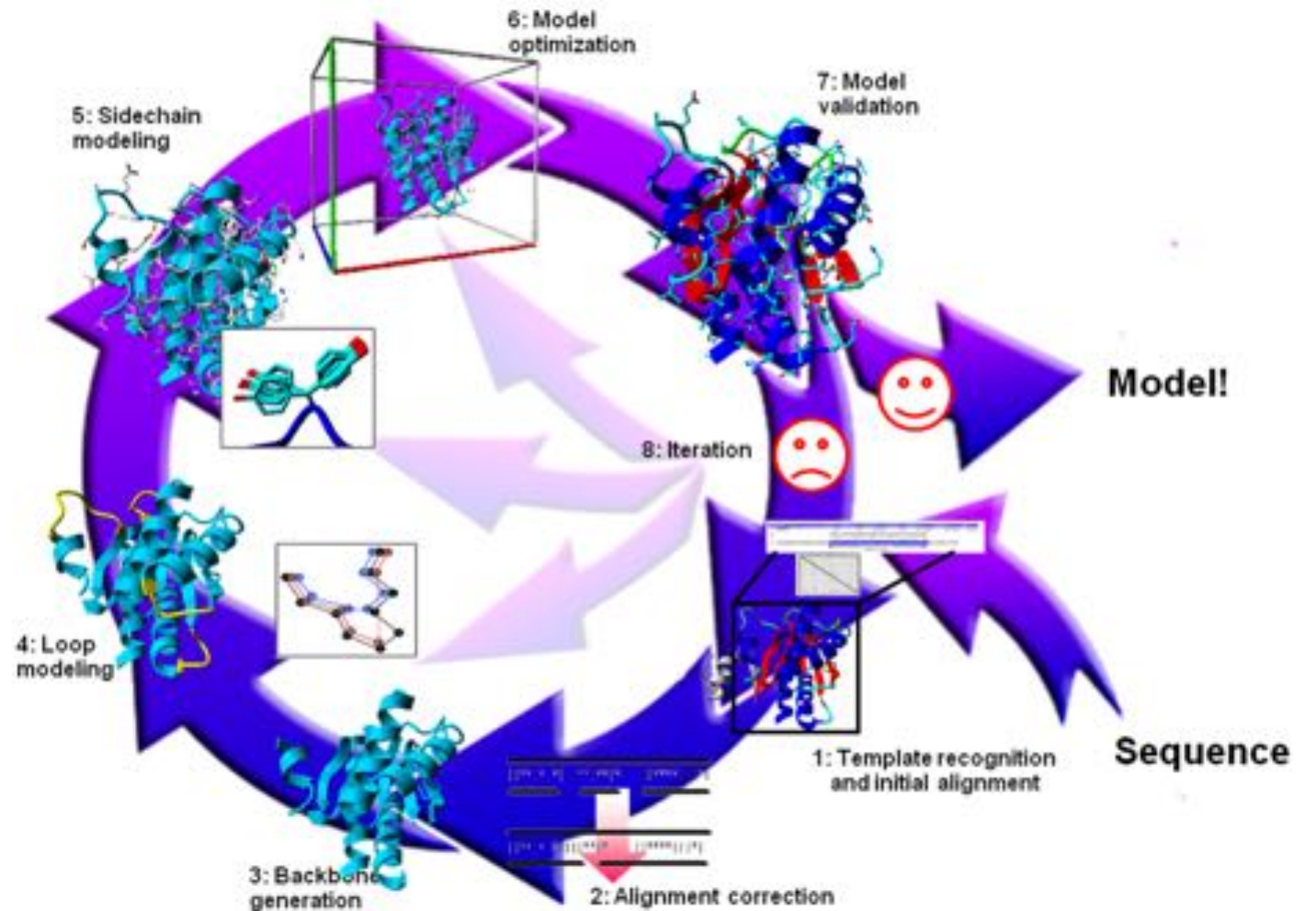
# Homology modeling

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



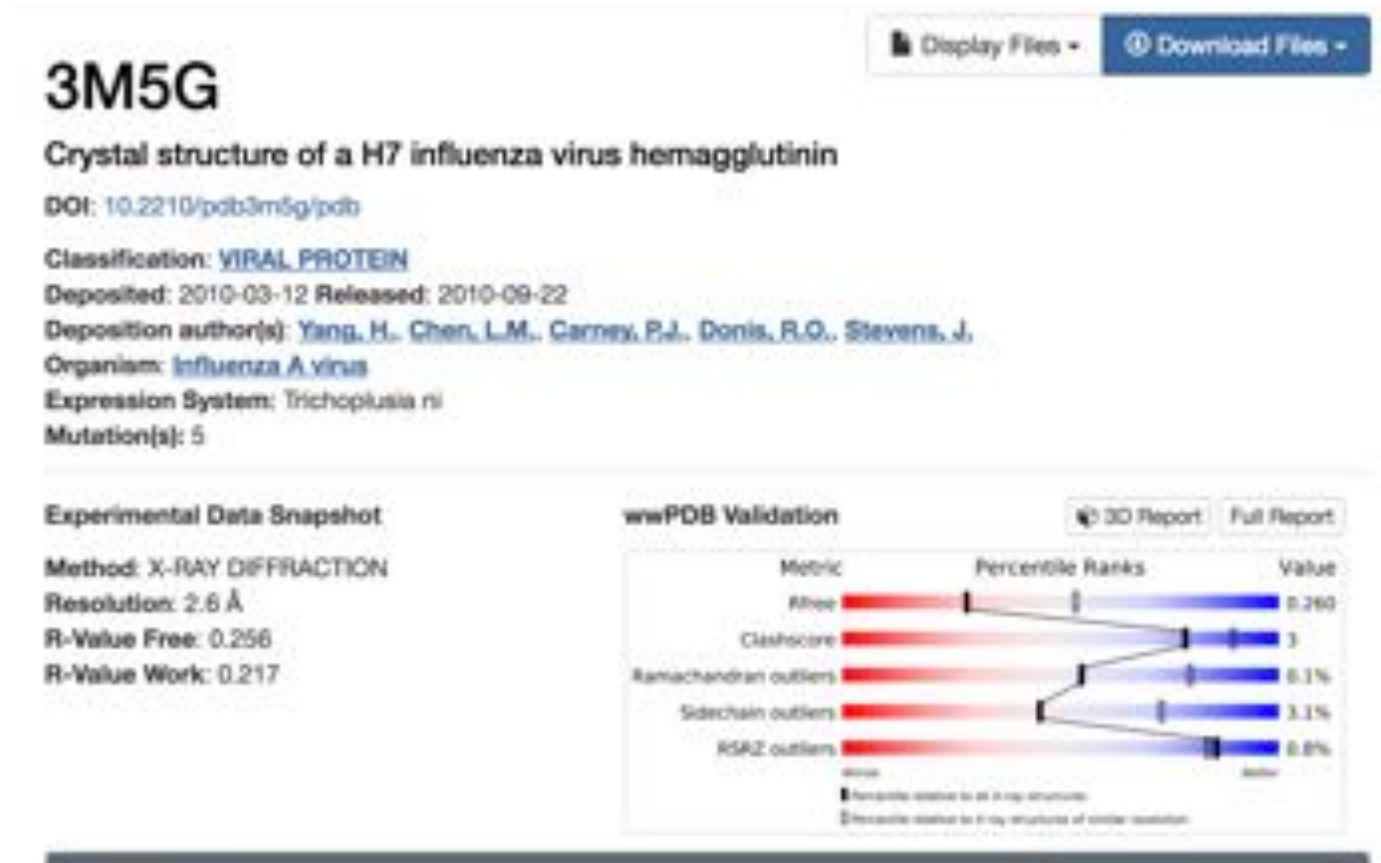
# Homology modeling

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

Conserved

```
ADDGSLAFVPSEF--SISPGEKIVFKNNAGFPHNIVFDEDSIPSGVI
TVNGAI--PGPLIAERLKEGONVRVTNTLDEDTSIHWHGLLVPFGM
-TSMAPAFGVQEFYRTVKQGDEVTVTIT-----NIDQIED-VSHGF
IE--KMKYLTPEVFYTIKAGETVYWVNGEVMPHNVAFKKGIV--GEI
-TSVAPSFSPSF-LTVKEGDEVTVIVTNLDE-----IDDLTHGF
ASAETMVFEPDFLVLEIGFGDRVRFVPTHK-SHNAATIDGMVPEGV
TVNGQ--FPGPRLAGVAREGDQVLVKVNHVAENITIHWHGVQLGT

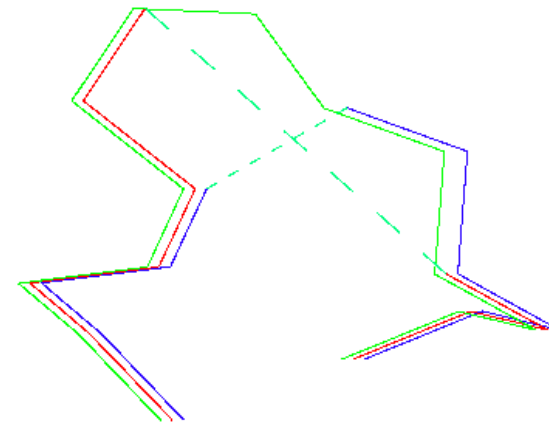
TKAVVLTFTNTSVEICLVMQGTSIV-----AAESHPLHLHGFFPSNF
```

# When to improve the alignment?

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

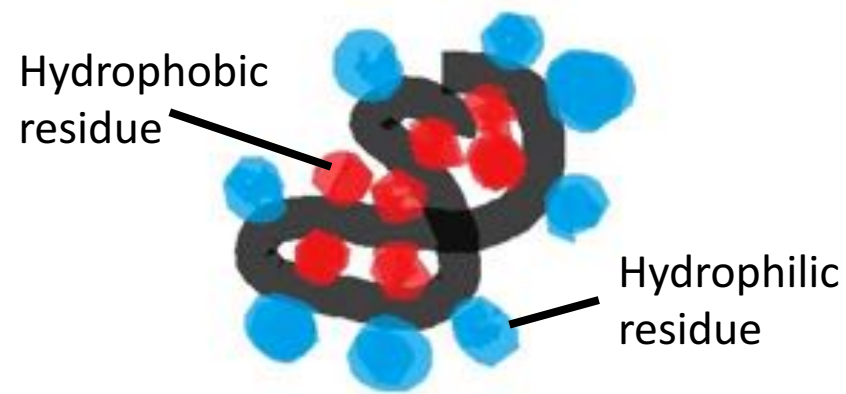
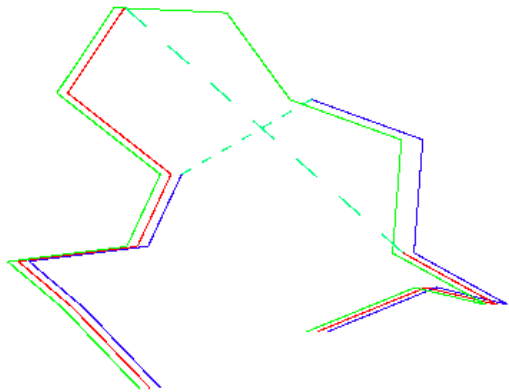
1	2	3	4	5	6	7	8	9	10	11	12	13	14	
PHE	ASP	ILE	CYS	ARG	LEU	PRO	GLY	SER	ALA	GLU	ALA	VAL	CYS	Template
PHE	ASN	VAL	CYS	ARG	THR	PRO	---	---	---	GLU	ALA	ILE	CYS	1 <sup>st</sup> alignment
PHE	ASN	VAL	CYS	ARG	---	---	---	THR	PRO	GLU	ALA	ILE	CYS	2 <sup>nd</sup> alignment

	Alignment score
1 <sup>st</sup> alignment	Small value
2 <sup>nd</sup> alignment	Big value



# 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



Conserved

```
PSEF--SISPGEKIVFKNNAGFPHN  
GPLIAERLKEGQNV RVNTLDEDTS  
VQEFYRTVKQGDEVTVTIT-----N  
TPEVFYTIKAGETVYWVNGEVMPHN  
QPSF-LTVKEGDEVTVIVTNLDE--  
PDFLVLEIGFGDRVRFVPTHK-SHN  
GPRLAGVAREGDQVLVKVNVHVAEN  
  
TSVEICLVMQGTSIV----AAESHP
```



# 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/>

TNx3	1	KGFVEDLDESFKNRNDDIWLDPYAPWCOHC	KI	IIWNEVGLM	SI	SPVKVGKM	ATSY	STAS	EF
PD1a	8	HVLVLRKSNFALAAHKYLVEFTAPWCOHC	AA	IEYAK	AG	KLAA	SEIRLAKV	ATER	DLAQY
CSQr#	65	.....	LD	EGV	.....	FGLV	SKD	A	AKL

★

TNx3	71	GVRGYP	TKLL	GLLA	..	YNYRCPR	YDDI	IR	FAHR	VSGAL	R	P	PS	..	ME	HM	K	R	H	V	F	V	V	G						
PD1a	78	GVRGYP	TKLL	GLLA	..	YNYRCPR	YDDI	IR	FAHR	VSGAL	R	P	PS	..	ME	HM	K	R	H	V	F	V	V	G						
CSQr#	88	GI	TEED	YVP	E	E	ERV	..	IEYDGE	P	AD	LV	P	LD	V	LED	P	EL	EG	REL	AE	NI	D	D	K	T	I	G	Y	K

★

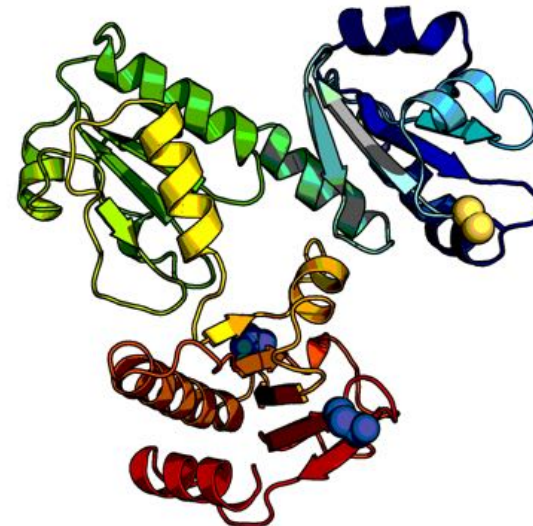
TNx3	135	GE	PLKEK	II	AAS	ELIV	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y							
PD1a	155	KD	EHYKA	K	EA	E	PH	P	PA	TPD	K	V	A	K	K	FL	LN	ID	P	Y	E	L	M	E	P	V	T	P	K	P	N	E	E	I	V	P	E
CSQr#	155	KD	EHYKA	K	EA	E	PH	P	PA	TPD	K	V	A	K	K	FL	LN	ID	P	Y	E	L	M	E	P	V	T	P	K	P	N	E	E	I	V	P	E

★

TNx3	202	EE	P	NYL	A	D	G	F	L	E	L	G	D	T	G	L	V	A	L	A	T	D	E	K	N	S	V	E	H	T	L	K	E	H	I	O	V	A	D	Y	D	L	F	H	R	F	P	F	G	H	M	G	N	
PD1a	225	H	R	S	TL	R	K	K	P	E	S	E	T	W	E	D	D	M	D	G	H	W	A	F	E	E	D	P	D	G	Y	F	L	E	L	K	S	V	A	D	N	T	D	..	N	P	D	L	S	I	I	W	I	D
CSQr#	225	H	R	S	TL	R	K	K	P	E	S	E	T	W	E	D	D	M	D	G	H	W	A	F	E	E	D	P	D	G	Y	F	L	E	L	K	S	V	A	D	N	T	D	..	N	P	D	L	S	I	I	W	I	D

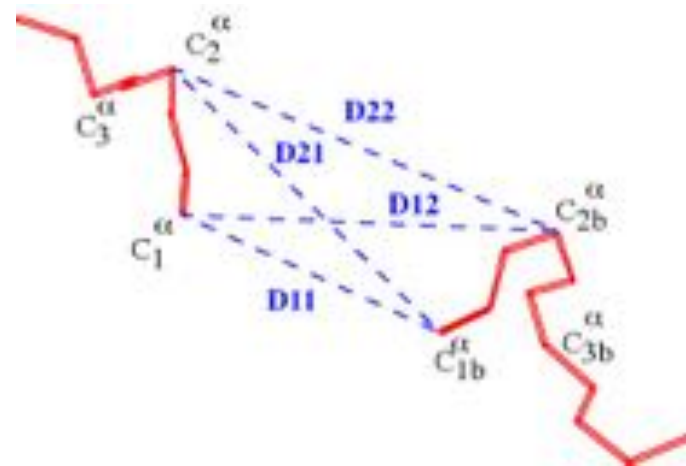
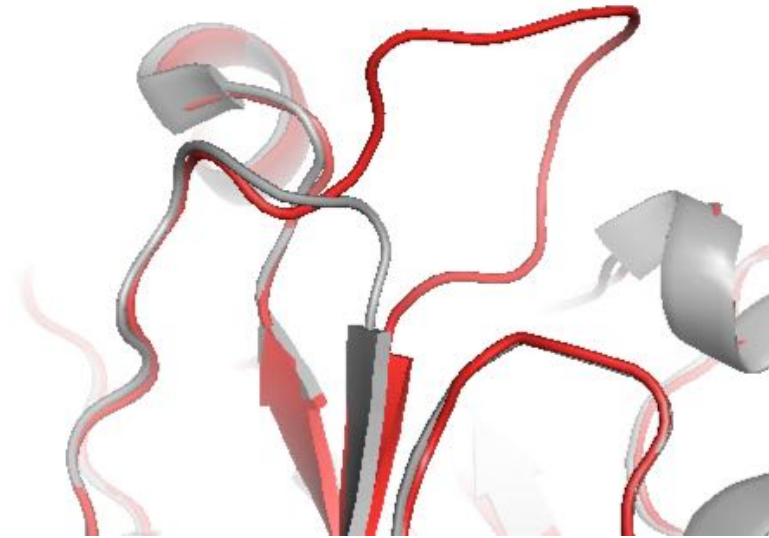
★

TNx3	271	...	I	N	T	L	L	M	D	E	F	V	T	V	V	L	I	T	S	N	O	Y	F	L	D	R	Q	I	K	N	V	D	M	V	O	F	I	N	N	E	
PD1a	293	P	L	L	V	P	W	E	K	T	F	D	I	D	E	R	A	D	I	G	V	V	T	D	D	S	V	M	E	..	P	S	A	E	L	E	D	W	E	D	V
CSQr#	293	P	L	L	V	P	W	E	K	T	F	D	I	D	E	R	A	D	I	G	V	V	T	D	D	S	V	M	E	..	P	S	A	E	L	E	D	W	E	D	V



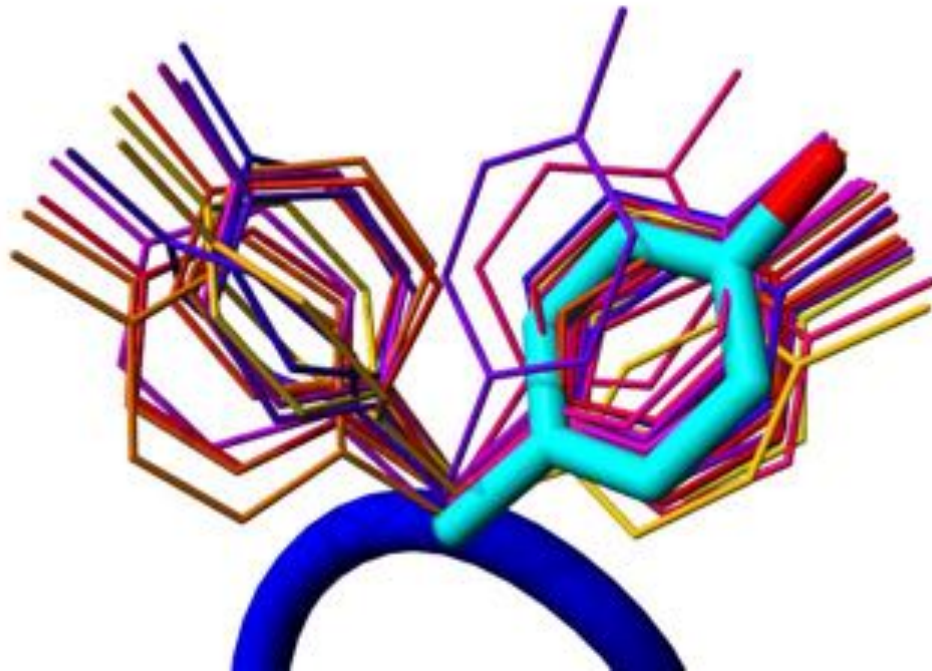
# Loop Modelling

- 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.



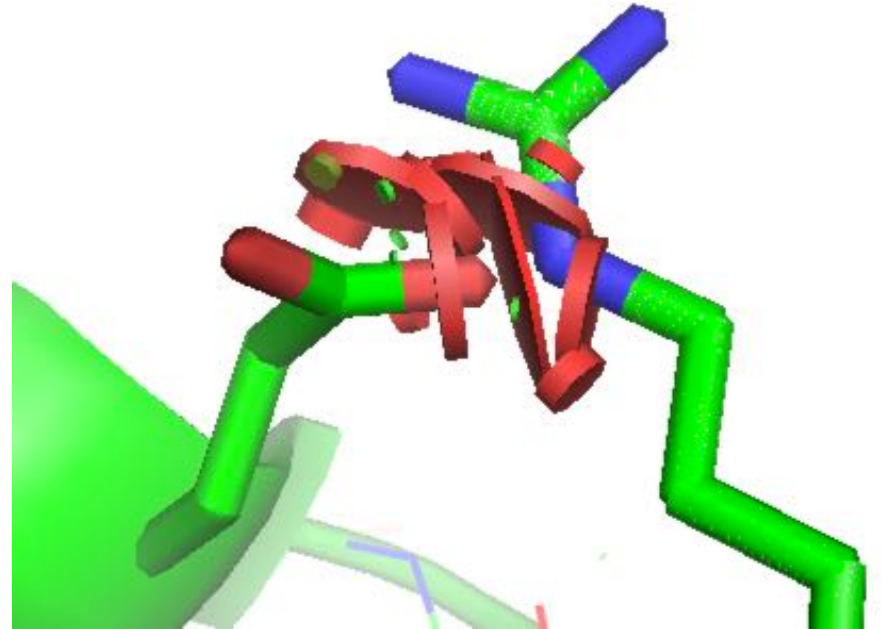
# Side Chains

- 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.*

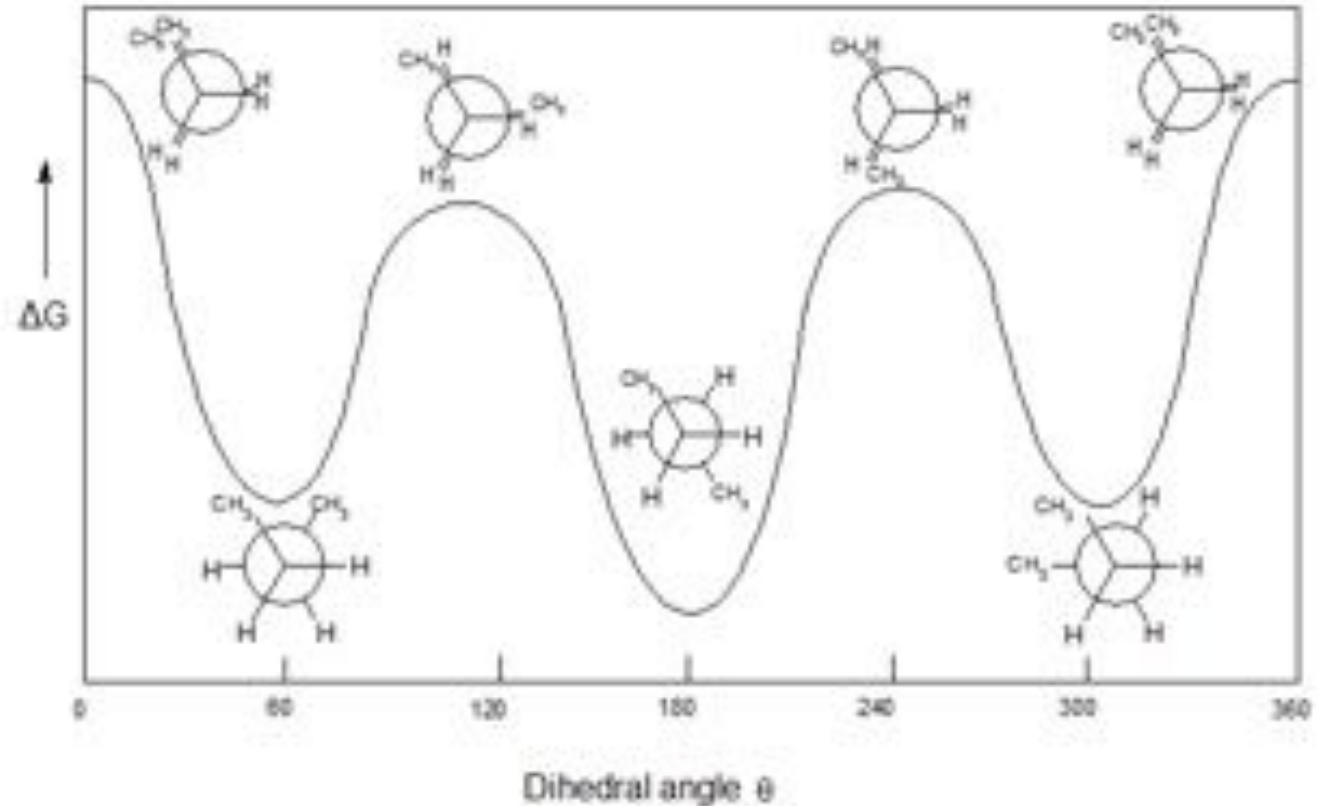
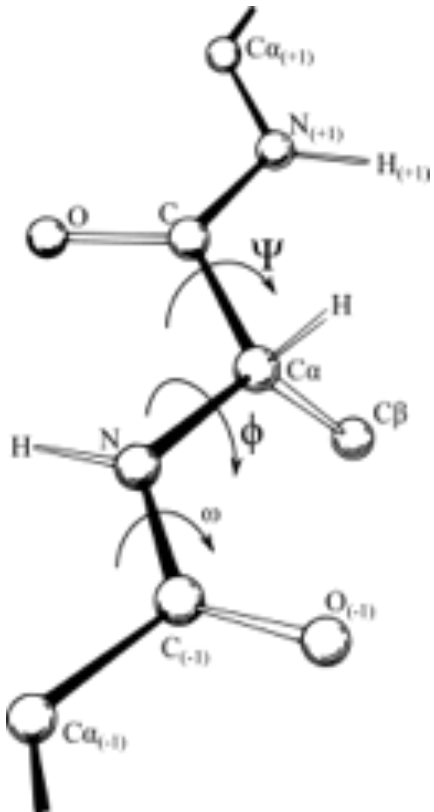


# Model Validation

- 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

# Dihedral angles

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





# Ramachandran plot

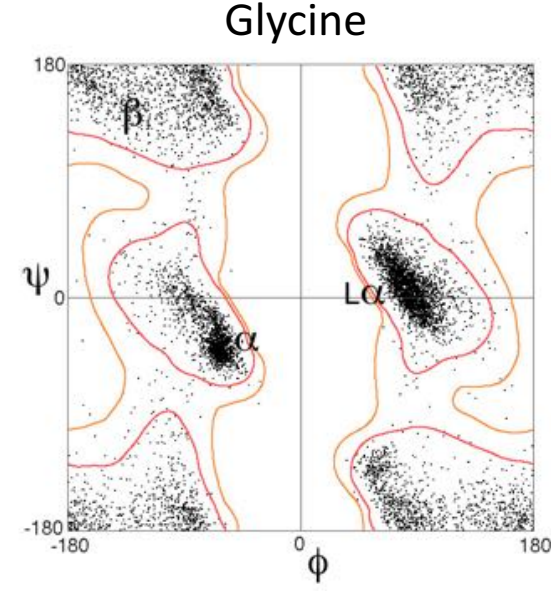
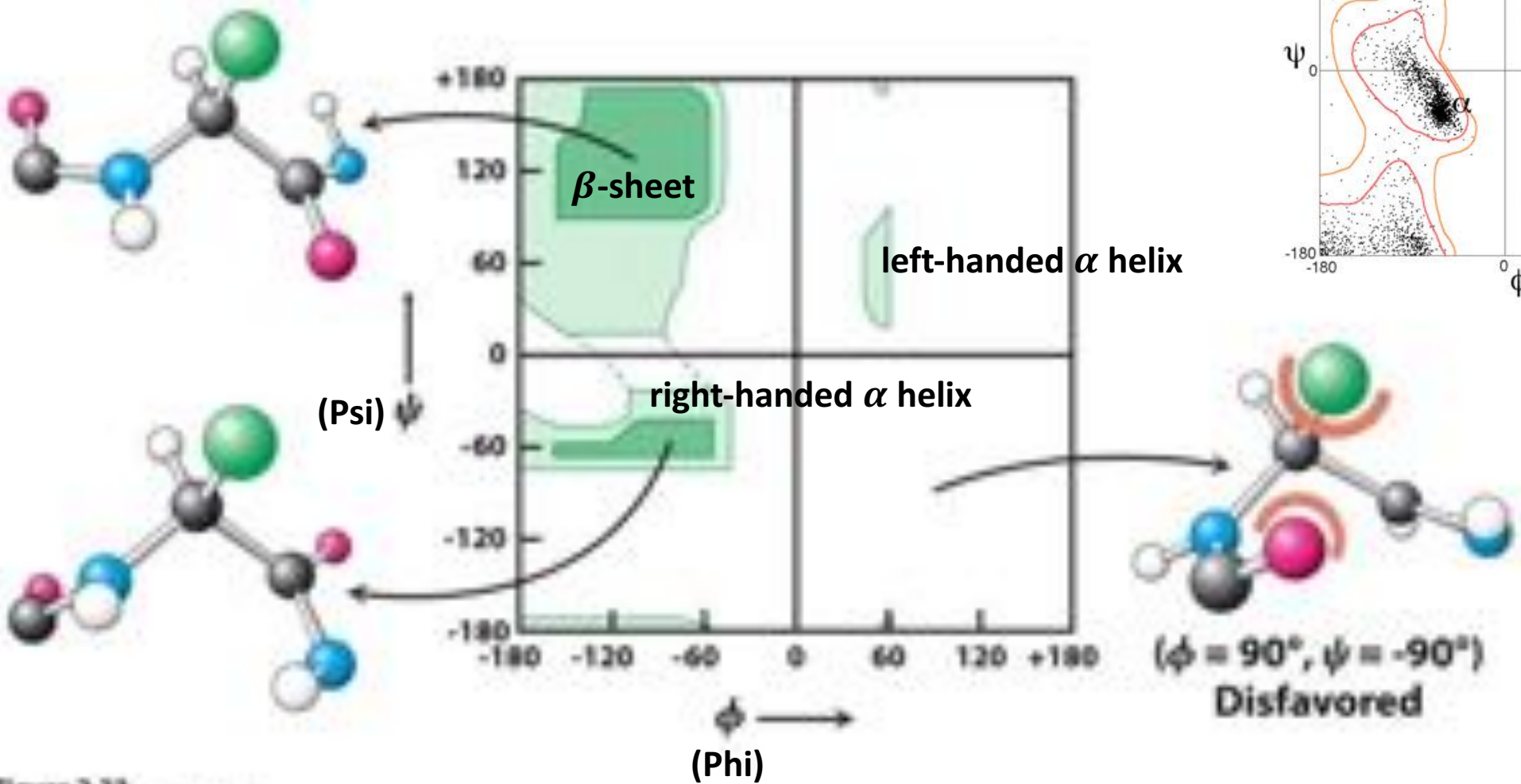
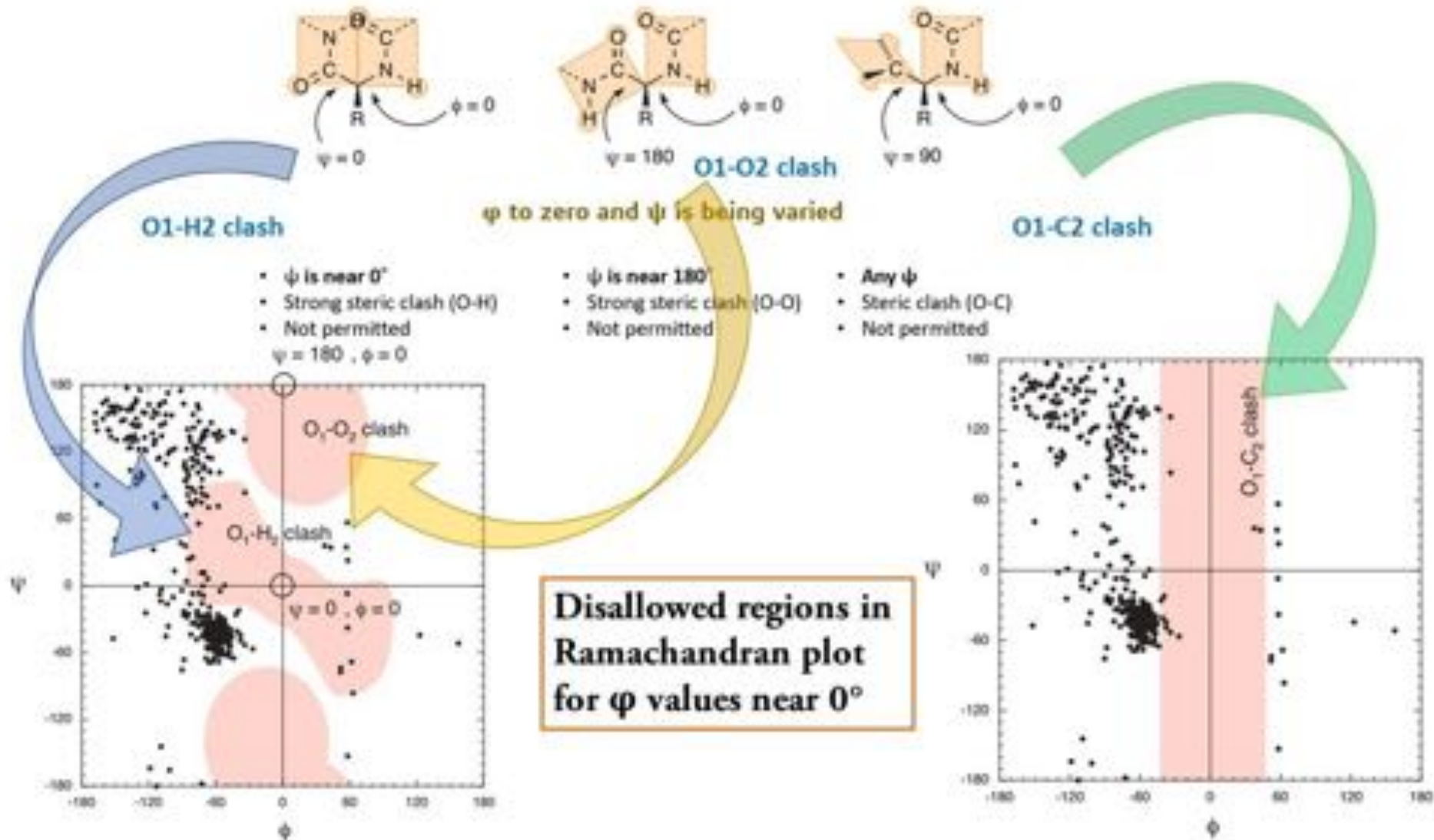


Figure 2.33  
Biochemistry, Seventh Edition  
© 2012 W. H. Freeman and Company

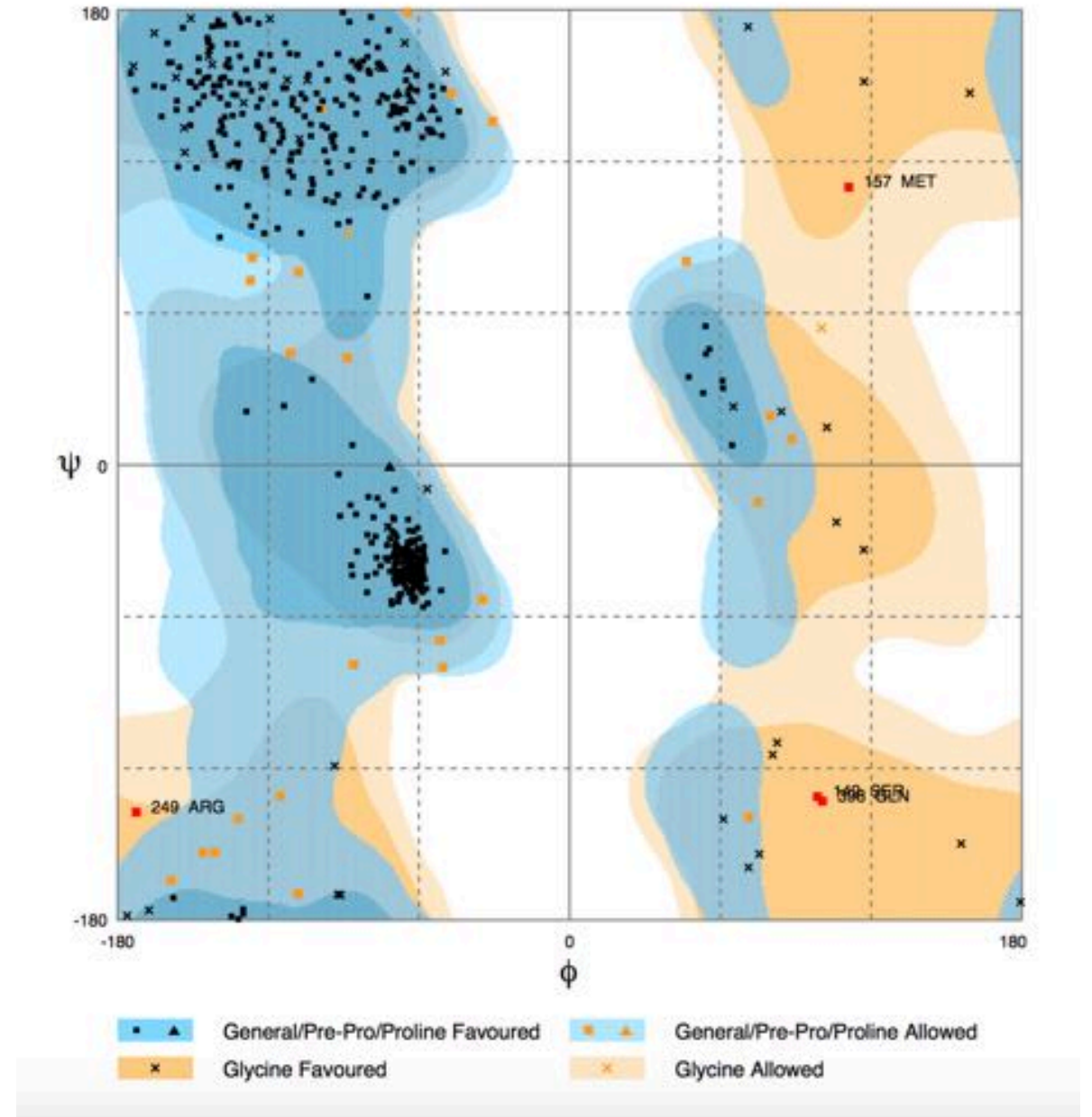
# Ramachandran plot





# 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