Protein Modelling:

Modelling a protein structure using non-experimental methods i.e, using computational methods which give utmost similar results to experimental structure is known as protein Modelling.

There are 3 main methods of modeling a protein structure:

- Homology modeling
- Threading/fold recognition
- Ab initio modeling.

1.Homology Modelling:

We only use this method when there is a protein with an experimental structure available that you suspect resembles your target protein (say over 30% sequence identity).

It relies on programs such as BLAST to search for similar proteins in various databases, structural or otherwise, such as the Protein Data Bank (PDB). Your homology model is then constructed from the suite of search results.

Homology modeling is also called “comparative modeling,” because we are comparing the model structure with known template structures as we build it. Swiss model works based on this principle.

2. Threading/Fold Recognition:

With this method, we predict the structure of our target protein using known protein folds for similar proteins found in various different databases. We can easily generate models using the online server I-Tasser.

3. Ab initio Methods:

This method predicts protein structures when experimental structural information of similar proteins is not available. The model is built from scratch by calculating the most favourable energy conformations of the amino acids. This method should only be used as a last resort since results from it are less reliable than template-based methods.

The Steps of Homology Modeling:

There are web-based servers for automated homology modelling. Protein Homology/analogy Recognition Engine V 2.0 (PHYRE2) and SWISS-MODEL are examples for this.

For this we just need to,

1.Upload a primary protein sequence.

2. search for templates, and let it load the templates.

3.When the templates are ready, select a template and build the model.

4. Our structure prediction will be out after quite a time.

These homology modelling servers are very sophisticated and generally produce excellent results.If we are going to perform homology modelling manually, then consider the process as a ladder with rungs on it. Each rung, or step, is important in its own way and we cannot skip a step and jump up.

If we do then, our final structure may turn out to be grossly incorrect and lead to erroneous interpretations.We will need some software for manual homology modeling. A good example is MODELLER.

Steps for manual Homology modelling:

1. Target Sequence Selection:

This step depends upon our needs. The protein sequence we wish to model is termed the "target sequence." After picking the appropriate length of target protein that allows us to just focus on the bit we are interested in. That is , in some cases, we don't need to model an entire protein. And in such cases, sticking with the essential protein sequence/domains will save us work and speed things up.

2. Template Protein Recognition:

The template proteins are the reference protein structures. In homology modeling, we match our target protein with all of the related protein structures that are present in the various databases using simple sequence alignment software. Select the proteins that are most closely related to our target to use as template structures.

3. Preparation of Template Protein:

Now, we have to trim back the template proteins because the experimental structures will contain extraneous matter. For example, symmetry equivalent protein chains, water molecules, ligands, and solvent of crystallization.

4. Sequence Alignment:

Next, we need to align our target and template protein sequences using a sequence alignment tool such as CustalW. This is a very important step in homology modeling because using an appropriate alignment algorithm is necessary for bagging the most valid template structures.

The alignment compares all of the proteins, target, and template(s), and can tell us which parts have completely conserved amino acid sequences.

5. Prediction of Secondary Structure:

The secondary structure of the model is built using tools present in ExPASy Portal. It compares the proposed secondary structures of our target to those of the template proteins and ranks them to iteratively build up the model.

6. Tertiary Structure and the Tentative Model:

With the secondary structure elements of the model built, we can start to piece these together and predict the tertiary structure of our target. We then need to visualize the model and check its overall structure to make sure it's sensible.

7. Loop Modeling:

If any loops are present in the model, we can optimize them further if they have unrealistic conformations. Use loop modeling software such as Modloop to do this and improve the quality of our final model.

8. Model Optimization and Validation:

Finally, once we have an almost complete model, we'll need to improve it to attain a nearnative confirmation via energy minimization. We can do this using protein model validation tools and verification servers. Such validation tests show whether our protein model is energetically satisfactory based on the spread of conformations observed in experimental structures for any given fold or feature.

Template used and it's Secondary Structure:

Template used: 4xi2.1.A

Secondary structure:

Steps :

Step 1:

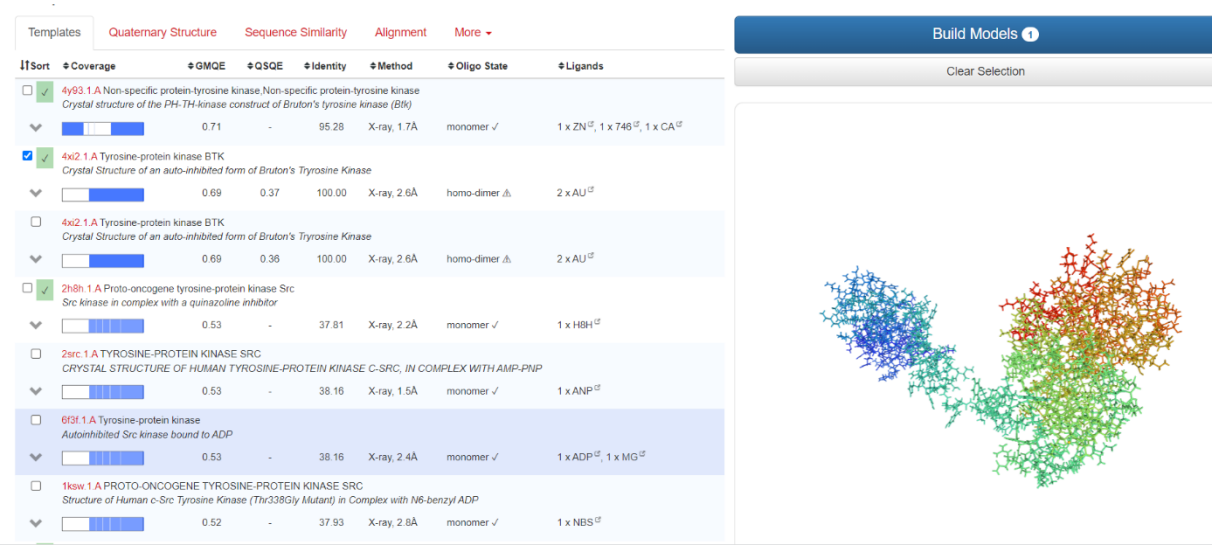
Extracting the protein sequence of UNIPROT IDENTIFIER : P35991

The sequence is:

```
>sp|P35991|BTK_MOUSE Tyrosine-protein kinase BTK OS=Mus
musculus OX=10090 GN=Btk PE=1 SV=4
MAAVILESIFLKRSQQKKKTSPLNFKKRLFLLTVHKLSYYEYDFERGRRGSKKGSIDVEK
ITCVETVIPEKNPPPERQIPRRGEESSEMEQISIIERFPYPFQVVYDEGPLYVFSPTEEL
RKRWIHQLKNVIRYNSDLVQKYHPCFWIDGQYLCCSQTAKNAMGCQILENRNGSLKPGSS
HRKTKKPLPPTPEEDQILKKPLPPEPTAAPISTTELKKVVALYDYMPMNANDLQLRKGEE
YFILEESNLPWWRARDKNGQEGYIPSNYITEAEDSIEMYEWYSKHMTRSQAEQLLKQEGK
EGGFIVRDSSKAGKYTVSVFAKSTGEPQGVIRHYVVCSTPQSQYYLAEKHLFSTIPELIN
YHQHNSAGLISRLKYPVSKQNKNAPSTAGLGYGSWEIDPKDLTFLKELGTGQFGVVKYGK
WRGQYDVAIKMIREGSMSEDEFIEEAKVMMNLSHEKLVQLYGVCTKQRPIFIITEYMANG
CLLNYLREMRHRFQTQQLLEMCKDVCEAMEYLESKQFLHRDLAARNCLVNDQGVVKVSDF
GLSRYVLDDEYTSSVGSKFPVRWSPPEVLMYSKFSSKSDIWAFGVLMWEIYSLGKMPYER
FTNSETAEHIAQGLRLYRPHLASERVYTIMYSCWHEKADERPSFKILLSNILDVMDEES
```

Step2:

Selecting a templating to build model.

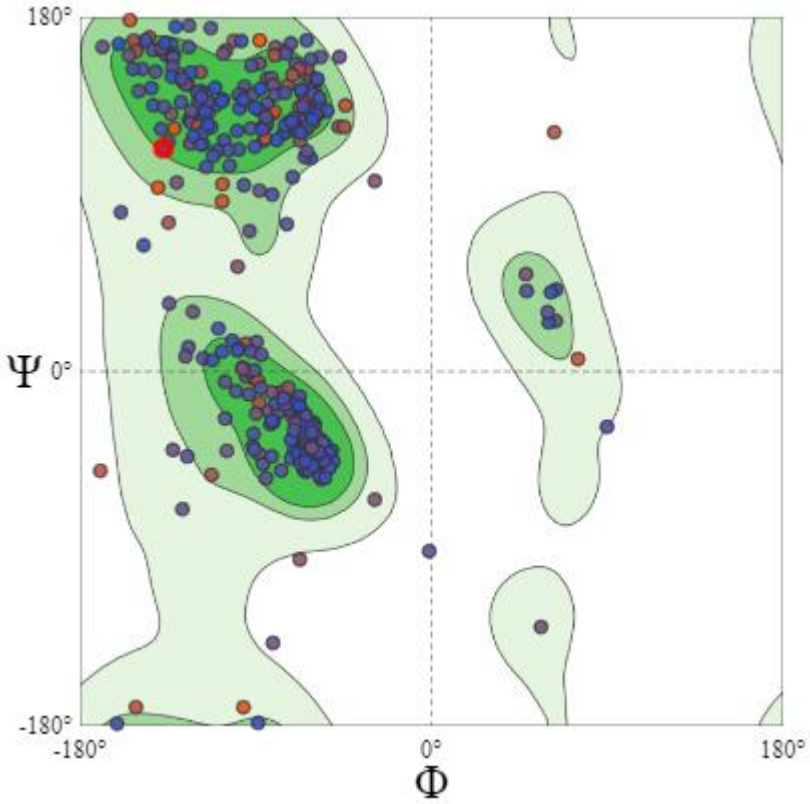


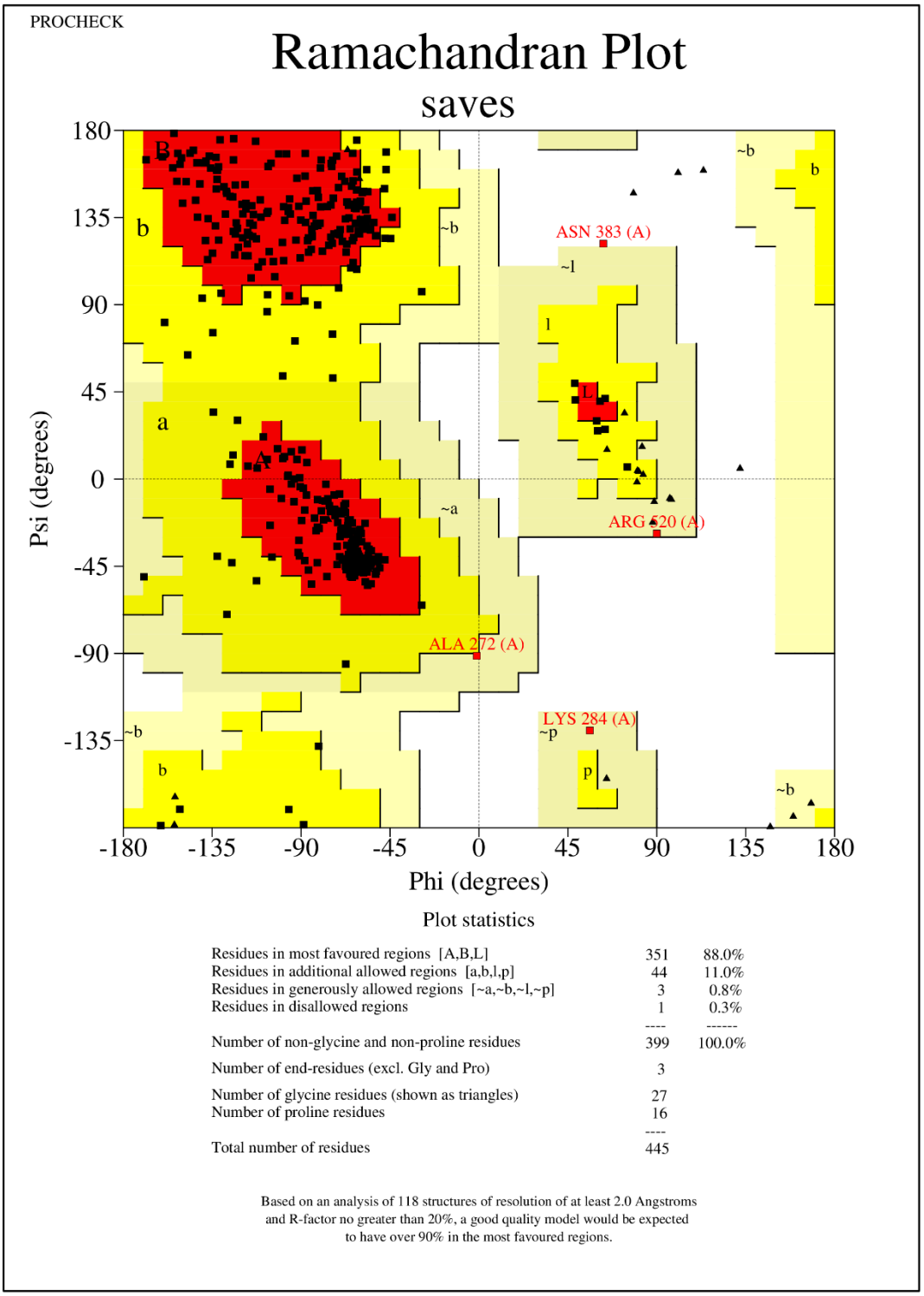
Step 3:

The built model for specified template is:

Ramachandran Plots:

The Ramachandran plots of the protein sequence shows that, this model contains Antiparallel beta sheets, Parallel beta sheets, Collagen triple helix, Right-twisted beta sheets, Left-handed alpha helix and Right-handed alpha helix have been observed in the model.





Plot statistics

Residues in most favoured regions [A,B,L]	351	88.0%
Residues in additional allowed regions [a,b,l,p]	44	11.0%
Residues in generously allowed regions [~a,~b,~l,~p]	3	0.8%
Residues in disallowed regions	1	0.3%
Number of non-glycine and non-proline residues	399	100.0%
Number of end-residues (excl. Gly and Pro)	3	
Number of glycine residues (shown as triangles)	27	
Number of proline residues	16	
Total number of residues	445	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

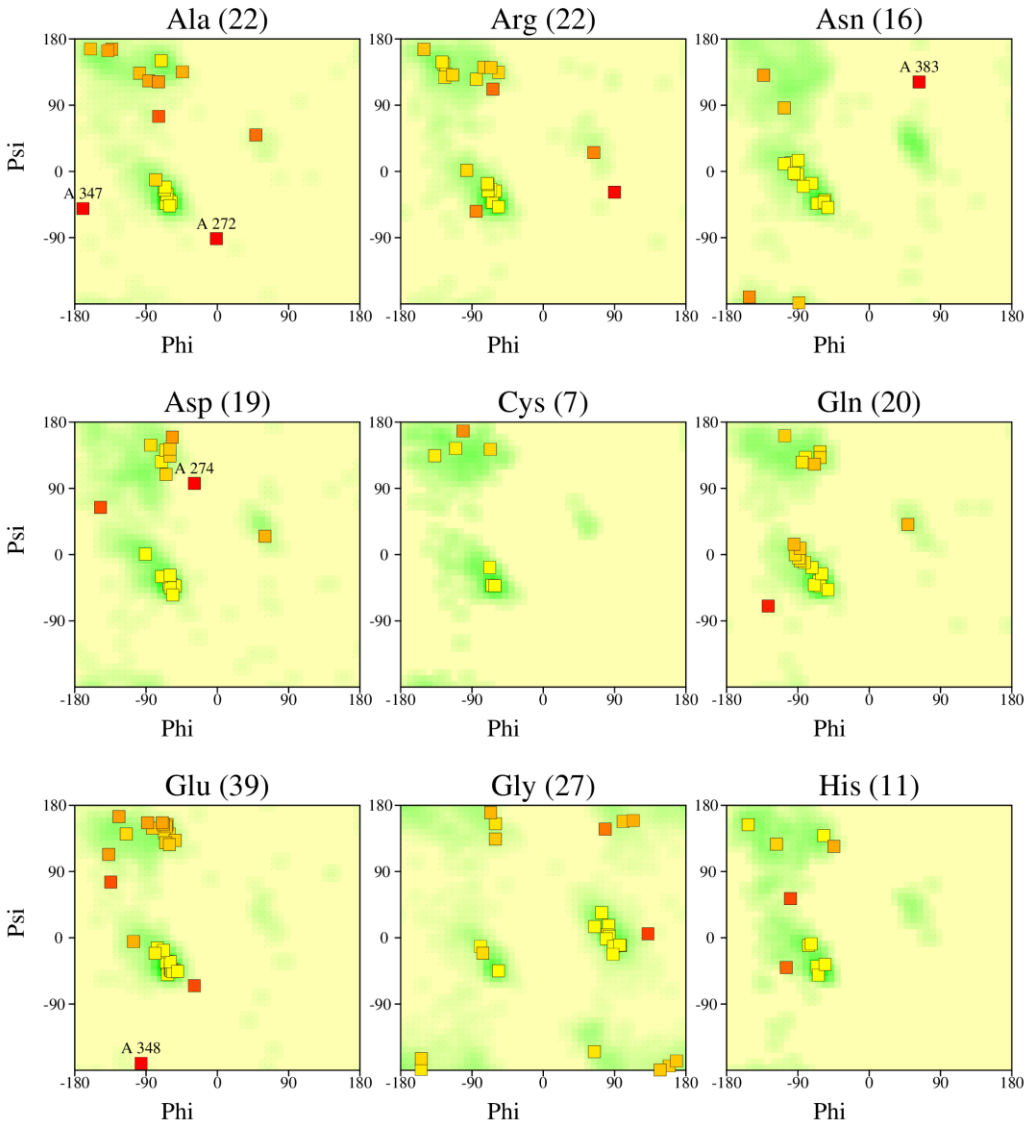
saves_01.ps

PROCHECK

Page 1

Ramachandran plots for all residue types

saves



Numbers of residues are shown in brackets. Those in unfavourable conformations (score < -3.00) are labelled.
Shading shows favourable conformations as obtained from an analysis of 163 structures at resolution 2.0A or better.

saves_02.ps

Conclusion:

I have learnt :

1.Protein modelling, types of protein modelling and homology modelling.

2. I have also described and learnt the steps of homology modelling(manual and computerised).

3. How protein modelling is done in Swissmodeller.

4. Ramachandran plot and it's analysis.