Project: **Homology modeling**

Course Title: **Structural Bioinformatics**

Course Code: **Binfo-605**

SUBMITTED TO: **Doctor. Saima**

SUBMITTED BY: **Ruqia Arshad (2020-ag-6295)**

                                  BS(BI)Section A

UNIVERSITY OF AGRICULTURE, FAISALABAD

Homology Modeling of unknown 3D Protein

Homology modeling of unknown 3D protein:

1.0 Abstract:

Open UniProtKB to find the unknown protein A0A8D5WHE7 of the wheat virus for modeling, which has no PDB structure. I used SwissProt, Robetta, MODELLER, and I-Teaser to create models of this unstructured 3D protein PDB. Select high GMQE, coverage, and identity template scores from SwissProt and build the model. Select the high confidence score model from Robetta. MODELLER did not have an ortholog according to NCBI. In the i-teaser, select high z-score and c-score templates. Firstly, select 5 models of Swiss-Prot, check their ERRAT, verify 3D, and Procheck scores, and select one model (6t0b.1.b) out of the 5 models of Swiss-Prot. Secondly, select 5 models of I-teaser, check their ERRAT, verify 3D, and ProCheck scores, and select one model (Model 3) out of the 5 models of I-teaser. Third, select 5 models of Robetta, check their ERRAT, verify 3D, and ProCheck scores, and select one model (Model 2) out of the 5 models of Robetta. Then, for the final selection of one model from each structure prediction tool, compare the results of these 3 selected models, one from Swiss, one from i-Teaser, and one from Robetta, by checking the results of ERRAT, Verify 3D, and ProCheck. Choose only one model from all these models that has a good result in ERRAT, Verify 3D, and ProCheck. Select the I-teaser model because it gives a good Errat score (70.5357), Verify 3D score (88.33), and ProCheck score (-1.29). However, the i-teaser PDB did not run on WinCoot or MolPribity. So, I decided to move to SwissProt. Then further work will be done on (SwissProt).Minimize the SwissProt model from Chimera and from WinCoot; remove rotamers,outliers and correct phi and psi angles.Got very good result in Swiss-Prot after refinement. But the residue length will reduce. So, I decided to move on to Robetta. Robetta's molprobity result has one outlier; Ramachandran favors 91%. BY Minimize structure from Chimera.It produced a poor result from chimera, with Ramachandra favoring only 87.29%, 3 outliers present, 1 poor rotamer, and a favored rotamer at 96.19%.Put thet Chimera PDB result in WinCoot for correction of ramachandran,Molprobity results. Achieved Ramachandron favored 94.07%, 0 outliers, 0 poor rotamers, and 100% favored rotamers from WinCoot. It is the story of my refinement.Original model of robetta is best so,choose Galaxy for refinement.The Galaxy refine of Robetta result give me Ramachandra favored 94.07%, 0 outliers, 0 poor rotamers, and favored rotamers 99%, which is the final result.

2.0 Extract the protein:

- Go to the advance option
- Select taxonomy (Wheat virus Q [2859709])
- Used not operator and find proteins that have no PDB structures.
- Select 120 amino acid protein ID is **A0A8D5WHE7**
- Gene is TGB2

■ Entry ▲		Entry Name ▲	Protein Names ▲	Gene Names ▲	Organism ▲	Length ▲
☐ A0A8D5WD45	📄	A0A8D5WD45_9VIRU	Replicase	RdRP	Wheat virus Q	2,069 AA
☐ A0A8D5WHE7	📄	A0A8D5WHE7_9VIRU	Movement protein TGB2[...]	TGB2	Wheat virus Q	120 AA

2.1 Protein Sequence:

from Uniprot

>tr|A0A8D5WHE7|A0A8D5WHE7_9VIRU Movement protein TGB2 OS=Wheat virus Q OX=2859709 GN=TGB2 PE=3 SV=1

MPLRAPQDNTIIVKYFCIVACVCGVLFLLTRDNLPHIGDQTHSFKHGGFYQDGTKRAVYC
GPRRNNPSSNLFSGLGSTFWIVTIFTLCLIYATRRSVSNRRYSCTCFHFLCCDCAPVTAA

Protein Function:

Plays a role in viral cell-to-cell propagation, by facilitating genome transport to neighboring plant cells through plasmosdesmata.

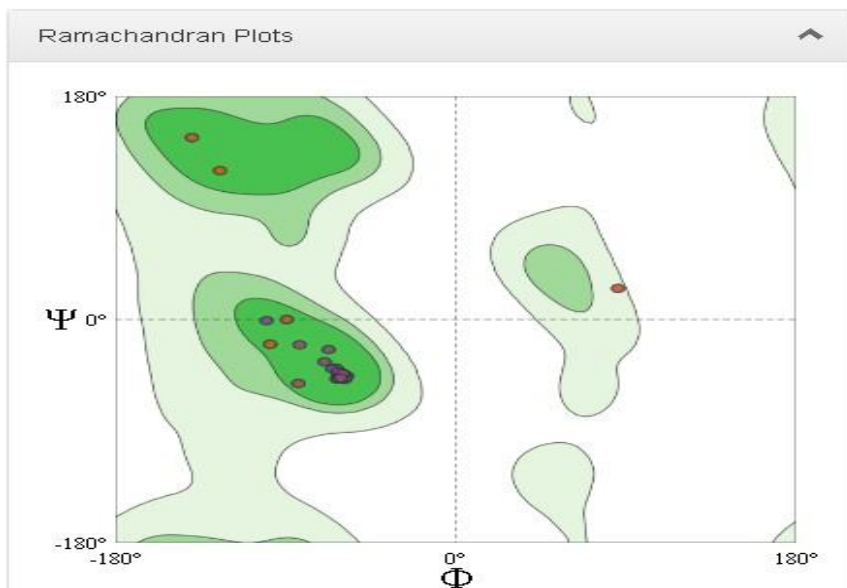## 3.0 Swiss-Model:

- Sequence paste in the swiss model.

3.1 Scores:

- Template use: 6t0b.1.b
- **Seq Identity**   23.33%
- GMQE   0.06
- Coverage   0.2

Ramachandran plot:

Molprobity result:

MolProbity Results		
MolProbity Score	0.95	
Clash Score	0.00	
Ramachandran Favoured	92.86%	
Ramachandran Outliers	3.57%	Y6 PRO
Rotamer Outliers	0.00%	
C-Beta Deviations	1	Y7 GLN
Bad Bonds	0 / 239	
Bad Angles	2 / 325	Y5 ALA, Y6 PRO

Model-template alignment:

Template
6t0b.1.b Cytochrome c oxidase polypeptide 5B, mitochondrial
The III2-IV(5B)2 respiratory supercomplex from S. cerevisiae

Seq Identity 23.33%

Coverage

Model-Template Alignment

```
Model_04  MPLRAPQDNTIIVKYFCIVACVCGVLFLLTRDNLP  35
6t0b.1.b  --VHGKGDVAFITKGVFLGLGISFGLFGLVRL---  95

Model_04  HIGDQTHSFKHGGFYQDGTKRAVYCGPRRNNPSSN  70
6t0b.1.b  -----------------------------------

Model_04  LFSGLGSTFWIVTIFTLCLIYATRRSVSNRRYSCT  105
6t0b.1.b  -----------------------------------

Model_04  CFHFLCCDCAPVTAA  120
6t0b.1.b  ---------------
```

Comparison with Non-redundant Set of PDB Structures

3.2 Predicted Models of Swiss Prot:

- I choose 4 templates to build a model which has good GMQE, Coverage & identity.
- And verify their results from structure validation tools.
- I choose one model i.e 6t0b.1.b from these 4 models with the best score.

3.2(a) Structure validation of swiss-prot model from ERRAT, verify 3D, Procheck

Models	ERRAT	Verify 3D	Procheck
6t15.1.Y best	100 result	0.00>=0.2 result	-0.14result
6t0b.1.b	100result	0.00>=0.2result	-0.13 result
6ymy.1.E	50 result	0.00>=0.2result	-0.12result Ramachandran pass
8age.1.H	75 result	2.44>=0.2result	-0.42result

3.2(b) Procheck results of swiss-prot:

Models	Favoured rejoin	Additional allowed	Generously allowed	disallowed
6t15.1.Y result	92	7	0	0
6t0b.1.b result	92	3	0	3
6ymy.1.E result best	85.2	14	0	0
8age.1.H result	67.7	25	3	3

6

Selected model:

- I select 6t0b.1.b model for further prediction.

4.0 I-taser:

- First register in the I-taser account & Paste the sequence in I-taser and give mail.
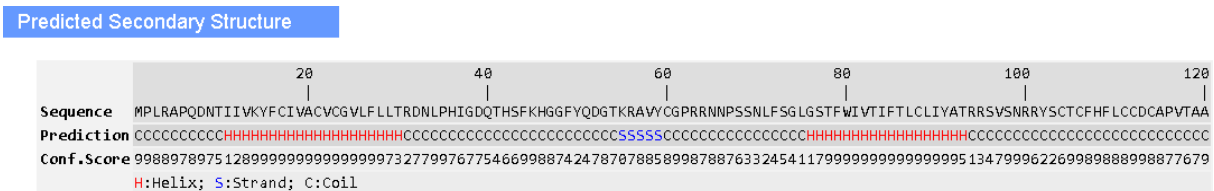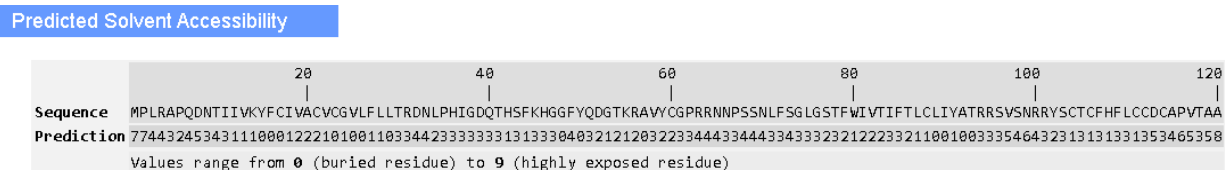

4.1 Results:

- The **confidence score** is high.

998897897512899999999999999997327799767754669988742478707885899878876332454117 99999999999999951347999622699898889998877679

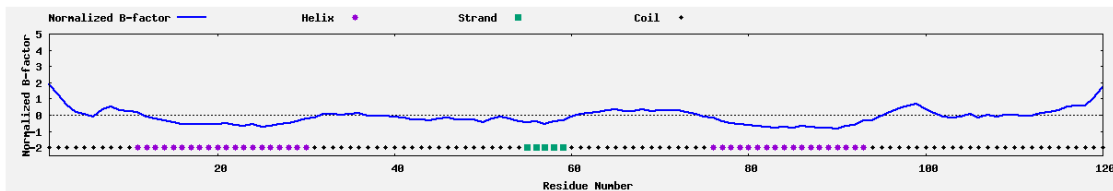- **Coils are more present** in prediction.

- More hydrophobic residues value range close to 0.
- Highly exposed are hydrophilic and Buried are hydrophobic

● **Predicted B-factor:**

Predicted normalized B-factor

(B-factor is a value to indicate the extent of the inherent thermal mobility of residues/atoms in proteins. In I-TASSER, this value is deduced from threading template proteins from the PDB in combination with the sequence profiles derived from sequence databases. The reported B-factor profile in the figure below corresponds to the normalized B-factor of the target protein, defined by B=(B'-u)/s, where B' is the raw B-factor value, u and s are respectively the mean and standard deviation of the raw B-factors along the sequence. Click here to read more about predicted normalized B-factor)
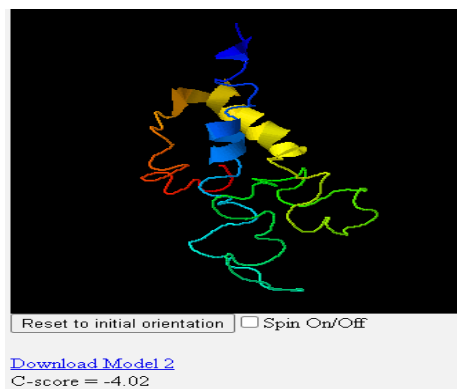

Scores:

Rank	PDB Hit	Iden1	Iden2	Cov	Norm. Z-score	Download Align.
1	2ndjA	0.10	0.20	0.84	1.01	Download
2	6t15	0.24	0.22	0.46	1.22	Download
3	6hu9	0.20	0.22	0.46	1.22	Download
4	6ymx	0.19	0.21	0.45	1.11	Download
5	7rpmA	0.14	0.27	0.57	0.49	Download
6	7s71A	0.04	0.23	0.95	0.83	Download
7	2moA	0.10	0.16	0.92	0.79	Download
8	6ymx	0.21	0.21	0.28	0.93	Download
9	2ndjA	0.11	0.20	0.86	0.77	Download
10	6ta9A	0.07	0.22	0.99	0.72	Download

- The more z-score mean good prediction . For good model all scores are high.
- C-score is positive.
- ™-score close 1
- Function prediction using GO-scores

I-teaser model:

Reset to initial orientation   ☐ Spin On/Off

Download Model 2
C-score = -4.02

8

4.2 **Predicted models of i-teaser:**

4.2(a) C-score and Structure validation tools of i-taser:

	c-score	ERRAT	Verify 3D	Procheck overall score
Model 1	-4.04	33.0357 result	69.17>=0.2result	-1.45 result
Model 2	-4.02	48.2143 result	62.50>=0.2 result	-1.12 result best
Model 3 best	-4.38	70.5357result	88.33>=0.2result pass	-1.29 result
Model 4	-4.60	41.9643 result	67.50>=0.2 result	-1.17 result
Model 5	-4.38	62.5 result	75.83>=0.2 result	-1.34 result

4.3(b) PROCHECK results for i-teaser:

	Favoured rejoin	Additional allowed	Generously allowed	disallowed
Model 1result	22	56	14	6
Model 2result best	52	13	12	3
Model 3result	49	13	17	2
Model 4result	36	44	12	6
Model 5result	41	34	15	8

Selected model:

- **I select i-taser 3 model for further prediction.**

5.0 Robetta:

• Register in robetta account. Give the target name and protein sequence then submit

Submit a job for structure prediction

Please do not submit jobs under different user accounts. Such jobs will be removed.

Required	
Target Name ❷	A0A8D5WHE7
Protein sequence ❷	MPLRAPQDNTIIVKYFCIVACVCGVLFLLTRDNLPHIGDQTHSFKHGGFYQDGTKRAVYC GPRRNNPSSNLFSGLGSTFWIVTIFTLCLIYATRRSVSNRRYSCTCFHFLCCDCAPVTAA
	or upload FASTA Choose File No file chosen

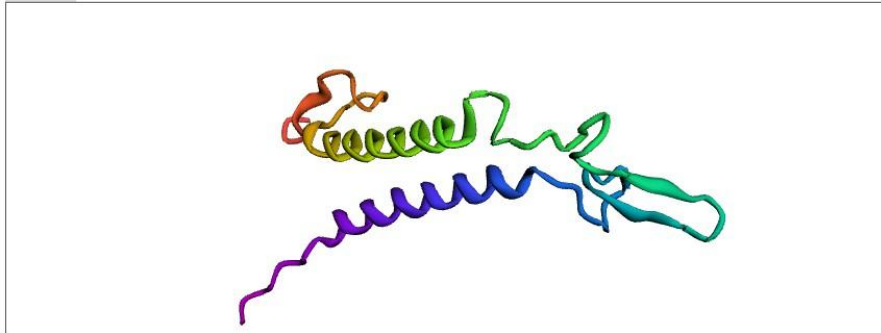| Optional	
RoseTTAFold ❷ ☐ CM ❷ ☐ AB ❷ ☐ Predict domains ❷ ☐	
Upload PDB template ❷ 🖹 Choose File No file chosen	
or enter PDB + chain IDs ❷ [ ] range ❷ [ ]	
Open constraints panel ❷ Open fragments panel ❷	

Submit   3 + 2 = 5   Keep private ❷ ☐
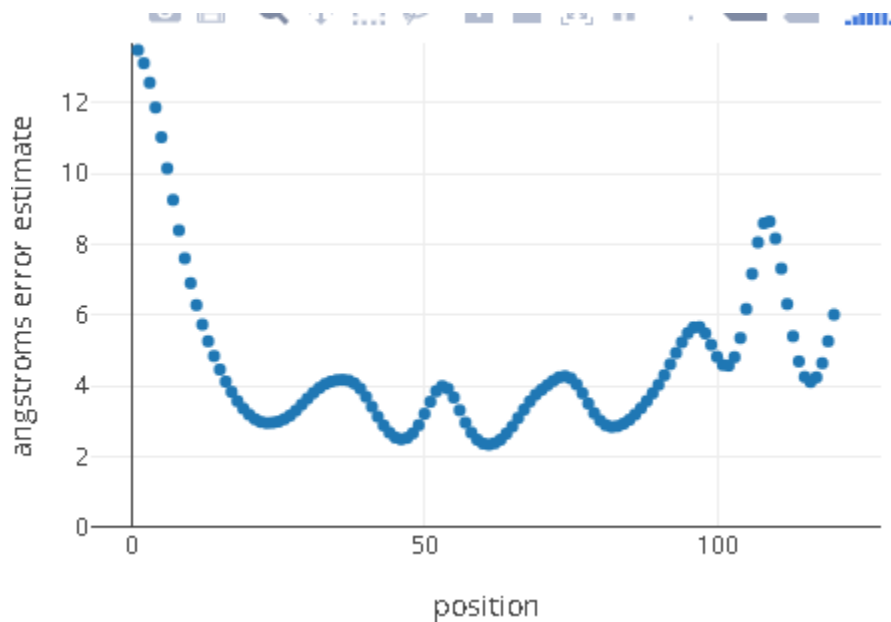
5.1 Results:

• 5 models are given by robetta.

Robetta Model:

Model 1   Model 2   Model 3   Model 4   Model 5

Confidence score:

	Model 1	Model 2	Model 3	Model 4	Model 5
Confidence score	0.63	0.63	0.63	0.63	0.63

5.2 Predicted models of Robetta:

5.2(a) Structure validation tools of robetta:

	ERRAT	Verify 3D	Procheck
Model 1	93.396 result	30>=0.2result	0.29result
Model 2best	89.1089 result	51.67>=0.2result	0.21result
Model 3	90.8257 result	31.67>=0.2result	0.22result
Model 4	86.5385 result	31.67>=0.2result	0.20 result
Model 5	98.9796result	30.83>=0.2result	0.28result

5.2(b) Procheck result for robetta models:

	Favoured rejoin	Additional allowed	Generously allowed	disallowed
Model 1result best	87	8	1	1
Model 2result	82	13	1	1
Model 3result	84	13	1	1
Model 4result	80	14	2	1
Model 5result	84	10	1	9

Selected model:

- **I select robetta model 2 for further prediction.**

6.0 Modeller:

Homology modeling with modeler:
NCBI did not give me any ortholog. No significant similarity was found

7.0 Final Selection of model:

I Select the i-teaser model rather than the Swiss-prot and robetta model because it gives a good score in errat, verify 3D and procheck score.

Models	ERRAT	Verify 3D	Procheck overall score
swiss-model(6t0b.1.b)	100result	0.00>=0.2result	-0.13 result
i-teaser(model 3)	70.5357result	88.33>=0.2 result	-1.29 result best
Robetta(model 2)	89.1089 result	51.67>=0.2result	0.21result

7.1 I-teaser not run on Wincoot & MolProbity:

ERROR: MODEL/ENDMDL card mismatches

It appears that the PDB you provided has a formatting error.
MolProbity believes the formatting error has to do with mismatched MODEL and ENDMDL cards.

Each MODEL card must be uniquely numbered and have a matching ENDMDL card.

The most common error is 'MODEL 1' existing with no ENDMDL card.

If you have a single-model structure, the **easy and likely correct solution is to remove the MODEL card**.
For further info please see the PDB's MODEL and ENDMDL documentation.

If you continue experiencing problems please contact us
using the feedback page which you can access on the left-hand
navigation bar from the main page.

Click here to return to the main page.

7.2 Switch to SwissProt:

After I-teaser,The second best result achieve from SwissProt.So, I done work on SwissProt.

MolProbity results:

The moleprobity result of Swiss 610b.1.b give outlier and Ramachandran not favored result.

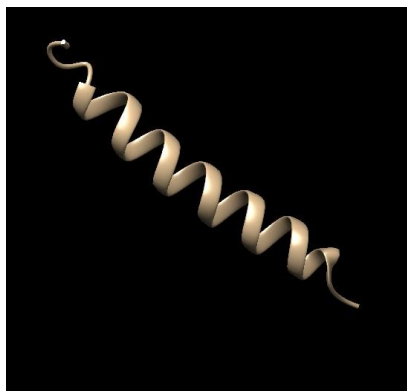|			
Poor rotamers	0	0.00%	Goal: <0.3%
Favored rotamers	27	100.00%	Goal: >98%
Ramachandran outliers	1	3.57%	Goal: <0.05%
Ramachandran favored	25	89.29%	Goal: >98%
Rama distribution Z-score	1.89 ± 1.65		Goal: abs(Z score) < 2
Cβ deviations >0.25Å	1	3.45%	Goal: 0
Bad bonds:	0 / 239	0.00%	Goal: 0%
Bad angles:	2 / 325	0.62%	Goal: <0.1%
Cis Prolines:	0 / 1	0.00%	Expected: ≤1 per chain, or ≤5%
CaBLAM outliers	1	3.8%	Goal: <1.0%
CA Geometry outliers	0	0.00%	Goal: <0.5%
Chiral volume outliers	0/42		

gives the raw count, right column gives the percentage.

Procheck result:

Summary	
Ramachandran plot	**Error**
All Ramachandrans	**Pass**
Chi1-chi2 plots	**Pass**
Main-chain params	

Minimize Swiss structure:

- File open in **chimera**
- Tools>structure editing>minimize structure>conjugated gradient steps (100)minimize>ok>steric only>ok>Gastegier
- favorites>reply log>save PDB

Structure validation:

Procheck result after minimize.

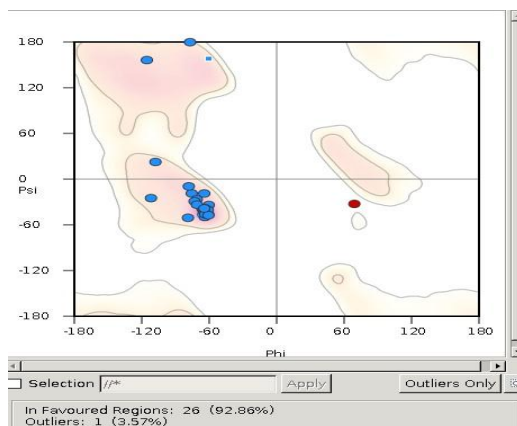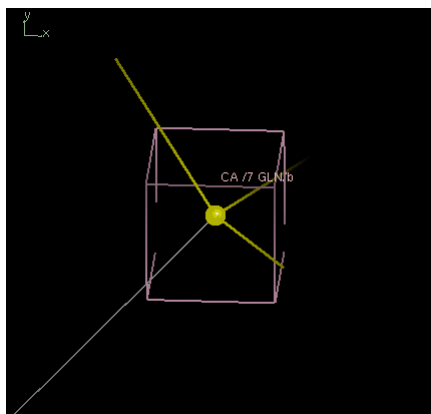## Wincoot:

File>open coordinates>validate>ramachandran plot>select outlier>from side bar click phi-psi>chage angel>ok

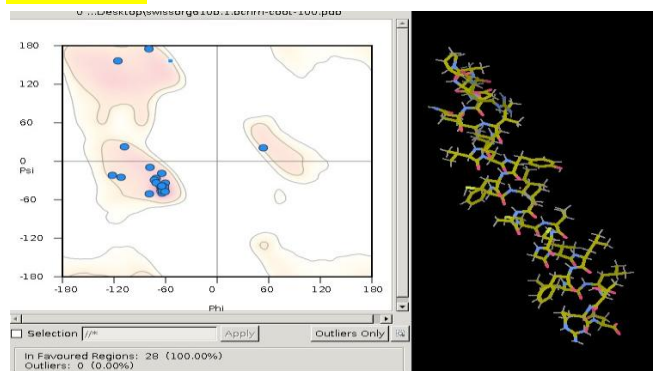- 1 outlier
- 92 favored region

Wincoot outlier:

In Favoured Regions: 26 (92.86%)
Outliers: 1 (3.57%)

Zoom-in outlier:

Remove outlier:

7.2(a) MolProbity after wincoot:

Poor rotamers	0	0.00%	Goal: <0.3%
Favored rotamers	27	100.00%	Goal: >98%
Ramachandran outliers	0	0.00%	Goal: <0.05%
Ramachandran favored	28	100.00%	Goal: >98%
Rama distribution Z-score	-0.35 ± 1.64		Goal: abs(Z score) < 2
Cβ deviations >0.25Å	0	0.00%	Goal: 0
Bad bonds:	0 / 239	0.00%	Goal: 0%
Bad angles:	3 / 325	0.92%	Goal: <0.1%
Cis Prolines:	0 / 1	0.00%	Expected: ≤1 per chain, or ≤5%
Twisted Peptides:	1 / 29	3.45%	Goal: 0
CaBLAM outliers	0	0.0%	Goal: <1.0%
CA Geometry outliers	0	0.00%	Goal: <0.5%
Chiral volume outliers	0/42		

Conclusion:

When i done work with Swiss-Prot the residue length reduce.So,I decide to move on Robetta.

7.3 Switch to Robetta:

MolProbity:

It give 1 outlier,ramachandran favoured only 91%,rotamers present.

_model_465759_2_robetta1_trimmed.pdb

Poor rotamers	0	0.00%	Goal: <0.3%
Favored rotamers	104	99.05%	Goal: >98%
Ramachandran outliers	1	0.85%	Goal: <0.05%
Ramachandran favored	108	91.53%	Goal: >98%
Rama distribution Z-score	0.22 ± 0.71		Goal: abs(Z score) < 2
Cβ deviations >0.25Å	0	0.00%	Goal: 0
Bad bonds:	1 / 967	0.10%	Goal: 0%
Bad angles:	5 / 1314	0.38%	Goal: <0.1%
Cis Prolines:	0 / 6	0.00%	Expected: ≤1 per chain, or ≤5%
Cis nonProlines:	1 / 113	0.88%	Goal: <0.05%
CaBLAM outliers	7	6.0%	Goal: <1.0%
CA Geometry outliers	2	1.72%	Goal: <0.5%
Chiral volume outliers	0/147		

Refinement:

MolProbity after refinement from chimera::

chimRobetta1_clean_trimmed.pdb

Poor rotamers	1	0.95%	Goal: <0.3%
Favored rotamers	101	96.19%	Goal: >98%
Ramachandran outliers	3	2.54%	Goal: <0.05%
Ramachandran favored	103	87.29%	Goal: >98%
Rama distribution Z-score	0.07 ± 0.82		Goal: abs(Z score) < 2
Cβ deviations >0.25Å	0	0.00%	Goal: 0
Bad bonds:	0 / 967	0.00%	Goal: 0%
Bad angles:	9 / 1314	0.68%	Goal: <0.1%
Cis Prolines:	0 / 6	0.00%	Expected: ≤1 per chain, or ≤5%
Cis nonProlines:	1 / 113	0.88%	Goal: <0.05%
CaBLAM outliers	11	9.5%	Goal: <1.0%
CA Geometry outliers	4	3.45%	Goal: <0.5%
Chiral volume outliers	0/147		

Result after Chimera minimization:


```
0 ...s\Hadia\Desktop\robetta\chimRobetta1.pdb
```

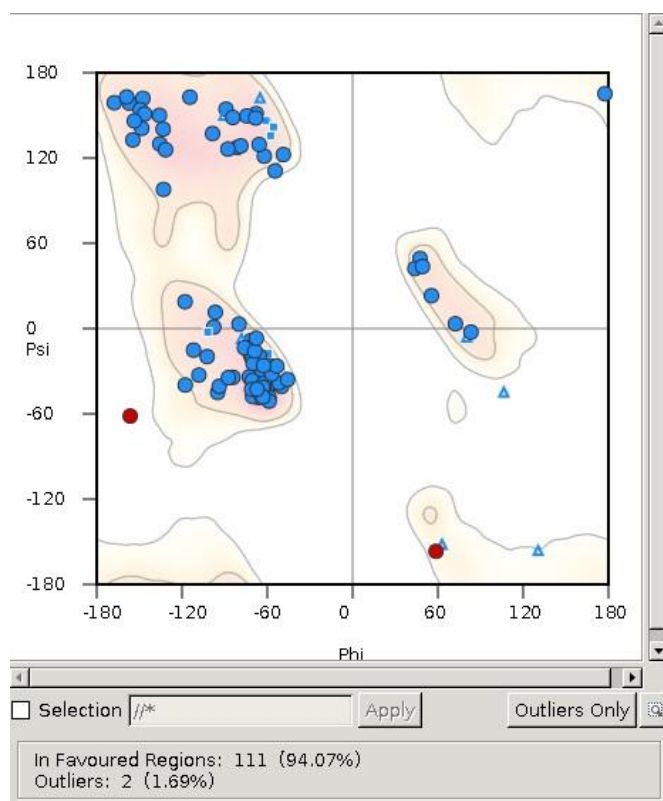In Favoured Regions: 99 (83.90%)
Outliers: 11 (9.32%)

MolProbity after refinement from Wincoot:

From chimera,ramachandron favoured only 87.29%,3 outliers present,1 poor rotamer,favored rotamer 96.19%.From wincoot,Ramachandron favored 94.07%,0 outliers,0 poor rotamer,favored rotmer 100%

94r.pdb_clean_trimmed.pdb

Duke Biochem
Duke University School

Poor rotamers	0	0.00%	Goal: <0.3%
Favored rotamers	105	100.00%	Goal: >98%
Ramachandran outliers	0	0.00%	Goal: <0.05%
Ramachandran favored	111	94.07%	Goal: >98%
Rama distribution Z-score	0.43 ± 0.81		Goal: abs(Z score) < 2
Cβ deviations >0.25Å	17	15.18%	Goal: 0
Bad bonds:	4 / 967	0.41%	Goal: 0%
Bad angles:	51 / 1314	3.88%	Goal: <0.1%
Cis Prolines:	0 / 6	0.00%	Expected: ≤1 per chain, or ≤5%
Cis nonProlines:	1 / 113	0.88%	Goal: <0.05%
Twisted Peptides:	2 / 119	1.68%	Goal: 0
CaBLAM outliers	13	11.2%	Goal: <1.0%
CA Geometry outliers	4	3.45%	Goal: <0.5%
Tetrahedral geometry outliers	3		

In Favoured Regions: 111 (94.07%)
Outliers: 2 (1.69%)

7.3(a)Galaxy refinement:

Analysis output: geometry for
alaxy1model2robetta1_trimmed.pdb

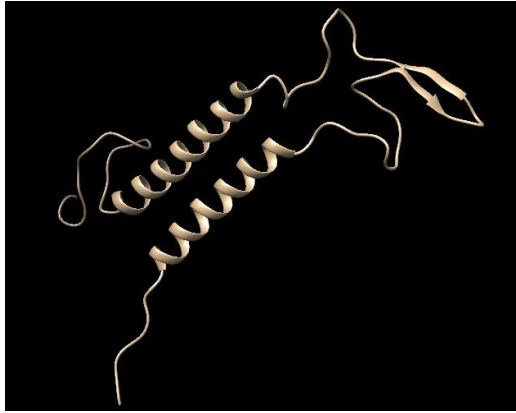| Poor rotamers	0	0.00%	Goal: <0.3%
Favored rotamers	104	99.05%	Goal: >98%
Ramachandran outliers	0	0.00%	Goal: <0.05%
Ramachandran favored	111	94.07%	Goal: >98%
Rama distribution Z-score	0.56 ± 0.60		Goal: abs(Z score) < 2
Cβ deviations >0.25Å	5	4.46%	Goal: 0
Bad bonds:	7 / 967	0.72%	Goal: 0%
Bad angles:	12 / 1314	0.91%	Goal: <0.1%
Cis Prolines:	0 / 6	0.00%	Expected: ≤1 per chain, or ≤5%
CaBLAM outliers	6	5.2%	Goal: <1.0%
CA Geometry outliers	1	0.86%	Goal: <0.5%
Chiral volume outliers	0/147		

Ramachandron favored 94.07%,0 outliers,0 poor rotamer,favored rotmers 99% which is the final result.

Final PDB:

8.0 Conclusion:

Robetta model refined from Galaxy.From Galaxy Ramachandron favored 94.07%,0 outliers,0 poor rotamer,favored rotmer 99% which is best result.

 The **ERRAT** score is above 70, **Verify 3D** is above 80, **MODELLER** MolProbity is high & DOP score is low good. **Procheck** overall score value is more in negative giving the best result.BY minimize the structure from **Chimera** the Ramachandran plot by procheck pass. From **WinCoot** the outlier is removed by adjustment of phi and psi angles.