GENE 433 – Term Project Part II

Modeling the query sequence using Modeller

Template 1PQ1 – Bcl-2-like protein 1 from Mus musculus

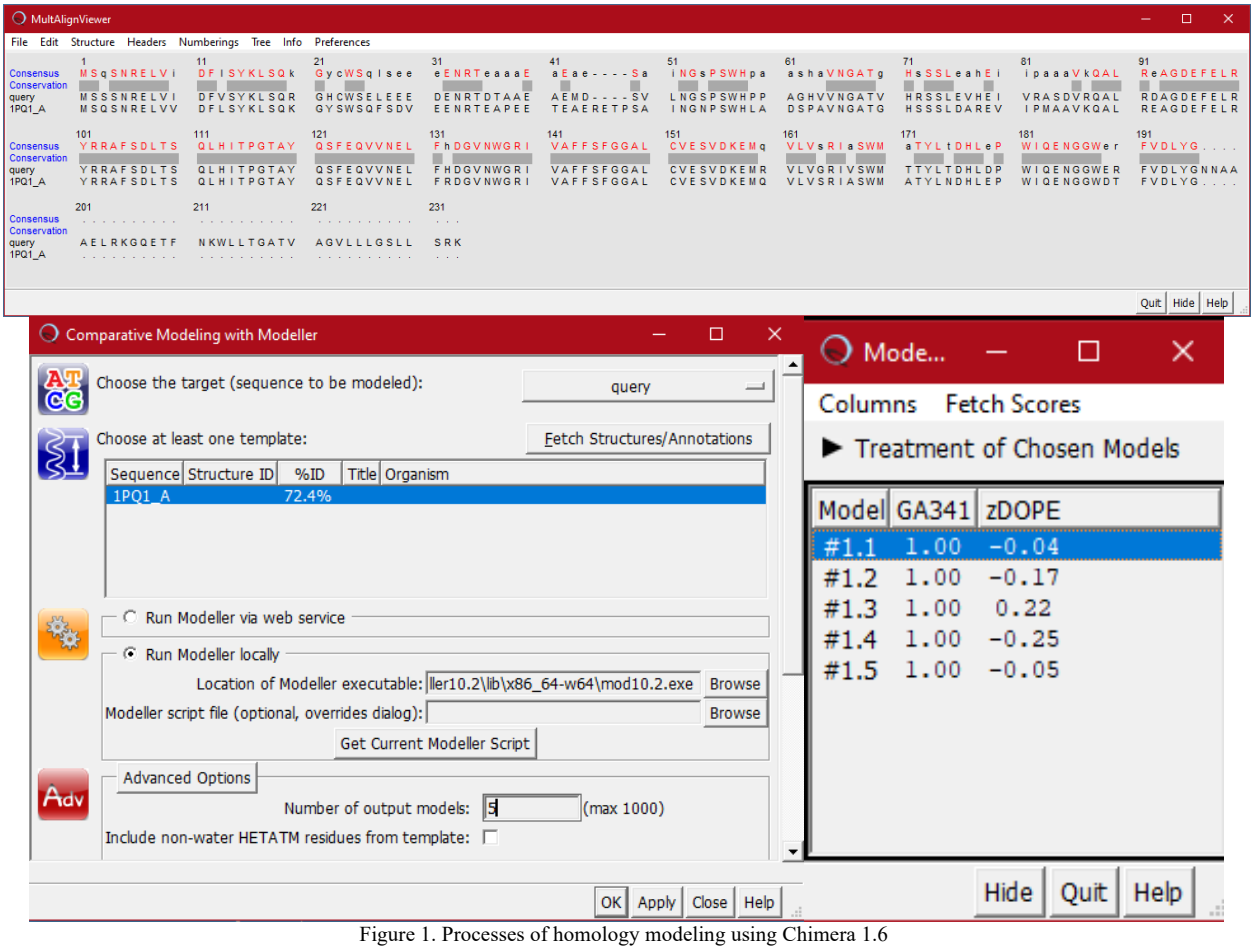

Figure 1. Processes of homology modeling using Chimera 1.6

The sequence of Bcl-2-like protein from Gallus gallus is introduced to Blast in Chimera, and 1PQ1 is selected for pairwise alignment. Then, modeler is used. 1PQ1 has 72.4% identity with the target sequence, which is quite enough for homology modelling. After calculations, 5 different model are constructed, and significance of model number four is the highest. Thus, it is selected for modeling the Bcl-2-like protein from Gallus gallus. Result can be seen below.

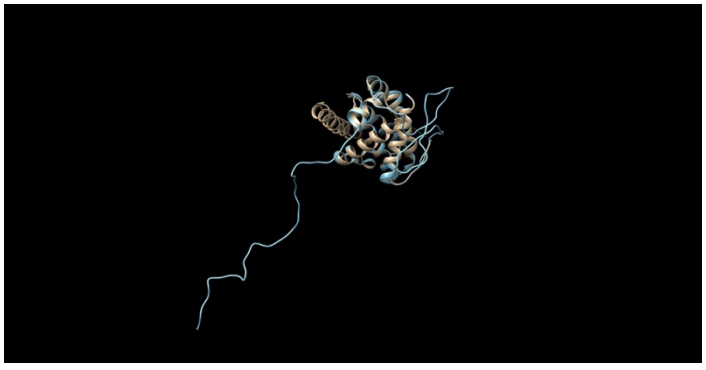

Figure 2. Construction of model from Bcl-2-like protein from Mus musculus

Exact process is done for 1G5J that is Bcl-2-like protein 1 from Homo sapiens. Result is given below. Identity score is 78.3%, and zDOPE score is -0.03.

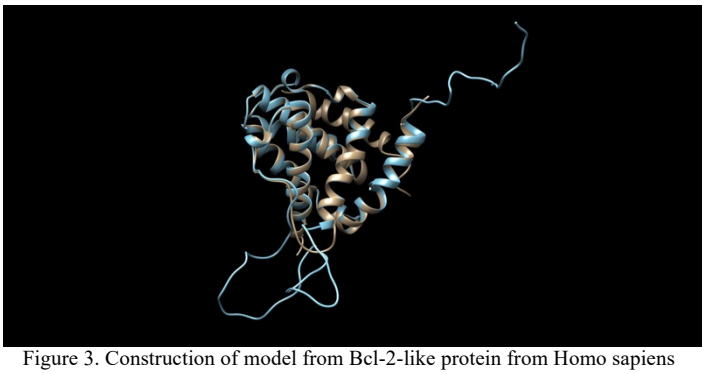

Figure 3. Construction of model from Bcl-2-like protein from Homo sapiens

Comparisons

There will be five different comparisons among templates and models. Models are named corresponding to their template protein. For example, Model-H is modelled with the help of the protein from Homo sapiens, and Model-M is from Mus musculus. Additionally, template proteins from Mus musculus and Homo sapiens are named as Template-M and Template-H, respectively.

1- Model-M vs. Model-H

Firstly, sequences of two model are matched and aligned by using Chimera Tools of Matchmaker and Matchalign, respectively.

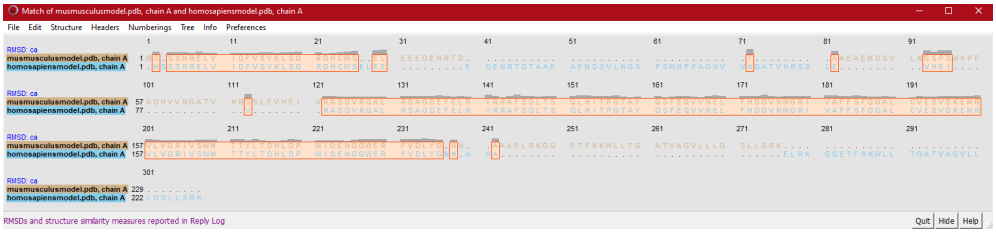

Figure 4. Match-align results of Model-M and Model-H

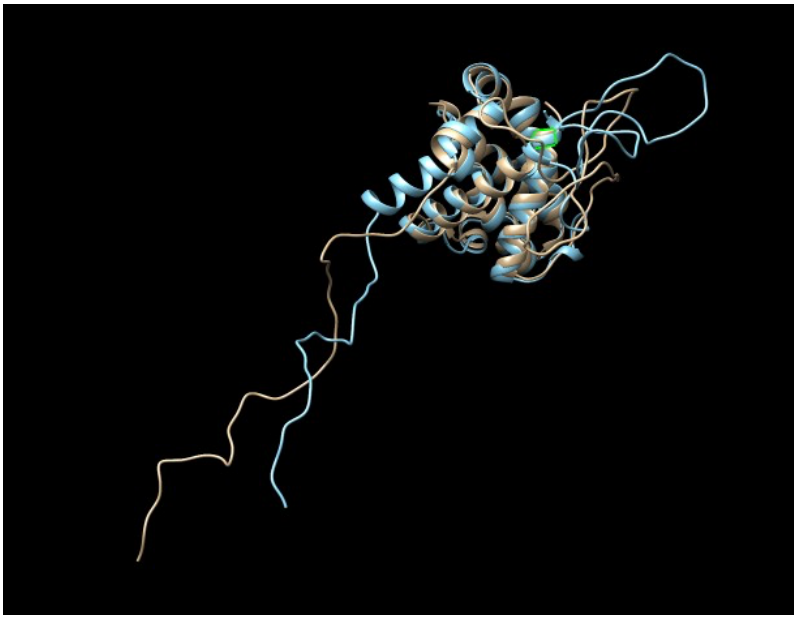

Figure 5. Matchmaker results of Model-M (light brown) and Model-H (blue)

RMSD values of Model-M and Model-H is 1.449, and overall RMSD is 1.449. Additionally, identity of alignment between models is 61.56%. Their overall folding are quite similar, and unfolded regions are likely. Secondary structures of both models are located very similar, most of which are matched and aligned as in figure above.

2- Model-M vs. Template-M

Same process is done for Model-M and Template-M, and results are given below.

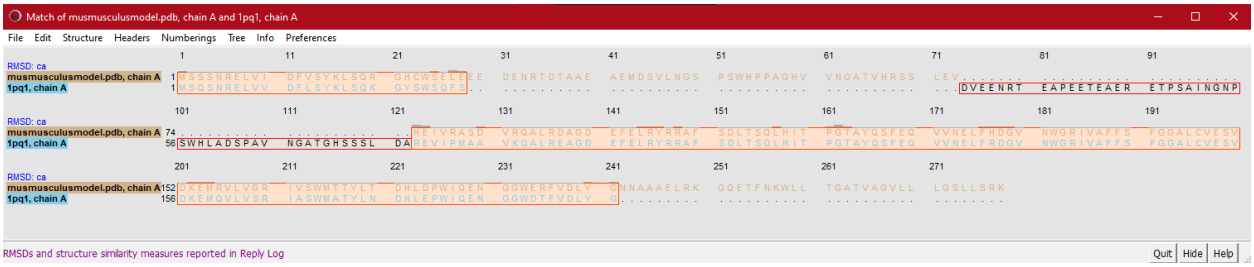

Figure 6. Match-align results of Model-M and Template-H

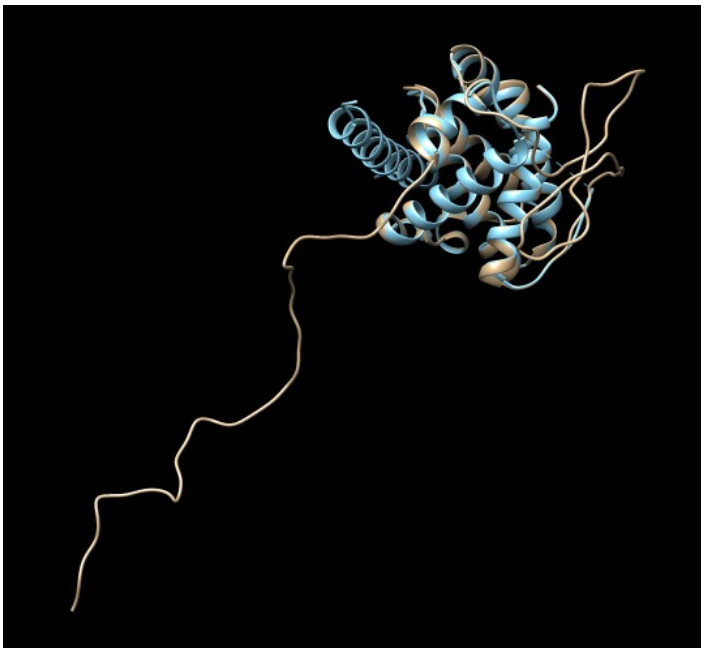

Figure 7. Matchmaker results of Model-M and Template-H

RMSD values of Model-M and Template-M is 0.206, and identity of alignment between models is 61.22%. Although template-M does not have unfolded region, the identity score is quite close to the first comparison. With that knowledge, we can estimate that aligned regions in this comparison and the first comparison are similar, and unfolded regions in the first comparisons are not aligned.

3- Model-H and Template-H

Same process is done for Model-H and Template-H, and results are given below.

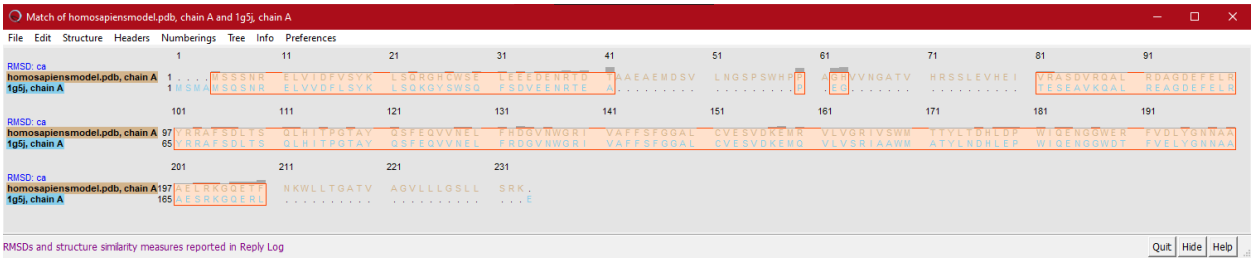

Figure 8. Match-align results of Model-H and Template-H

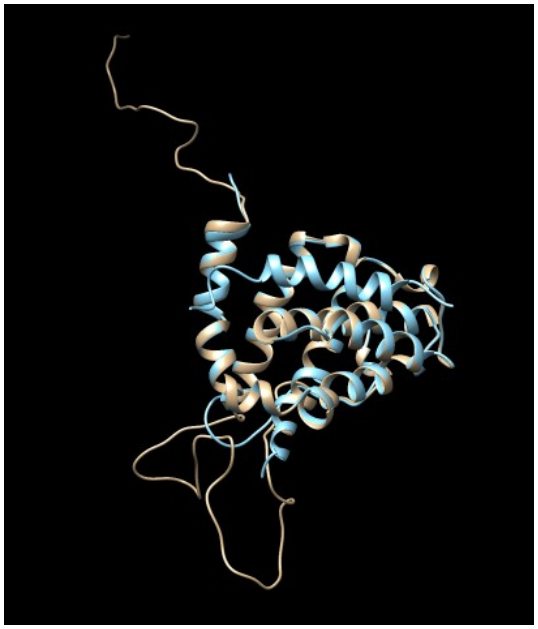

Figure 9. Matchmaker results of Model-H and Template-H

RMSD values of Model-H and Template-H is 0.474, and identity of the alignment between models is 76.00%. In this case, identity score is higher than first two comparisons. The reason might be about template selection since homolog protein from Homo sapiens identity score between target protein was higher than homolog protein from Mus musculus. Also, aligned regions are folded; thus, this template's structure is more useful in terms of modeling.

4- Template-M and Model-H

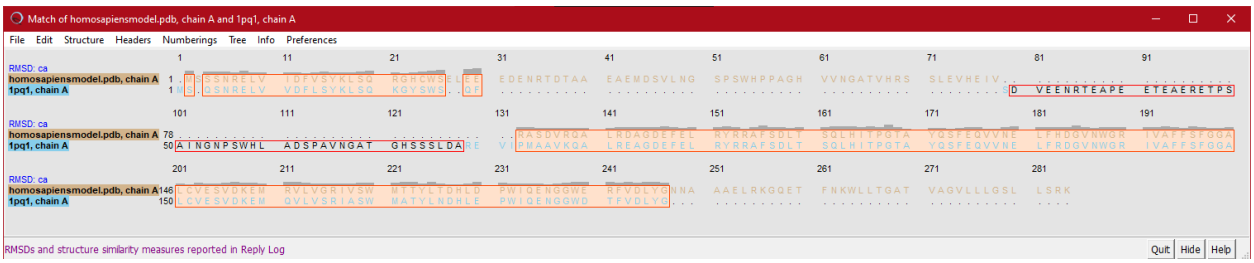

Figure 10. Match-align results of Template-M and Model-H

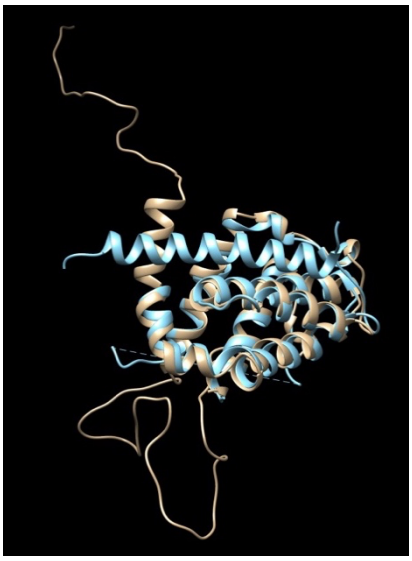

Figure 11. Matchmaker results of Template-M and Model-H

RMSD values of Template-M and Model-H is 1.310, and identity of the alignment between models is 59.69%. Here, Bcl-2-like protein from Mus musculus and a model that is constructed from the Bcl-2-like protein from Homo sapiens are compared, and alignment score is the least among other comparisons.

5- Model-M and Template-H

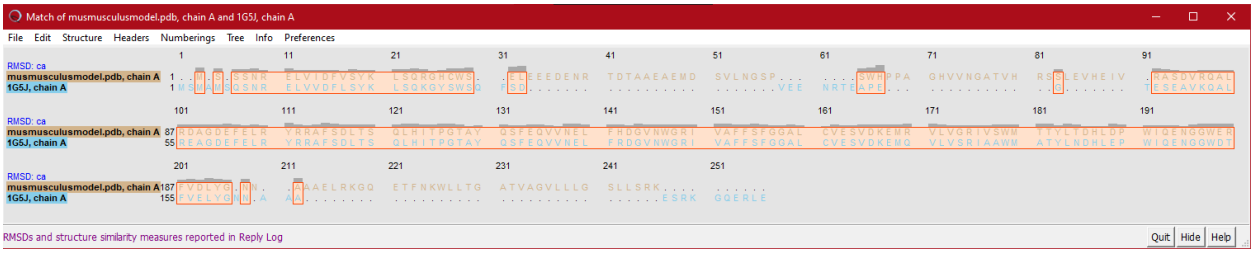

Figure 12. Match-align results of Model-M and Template-H

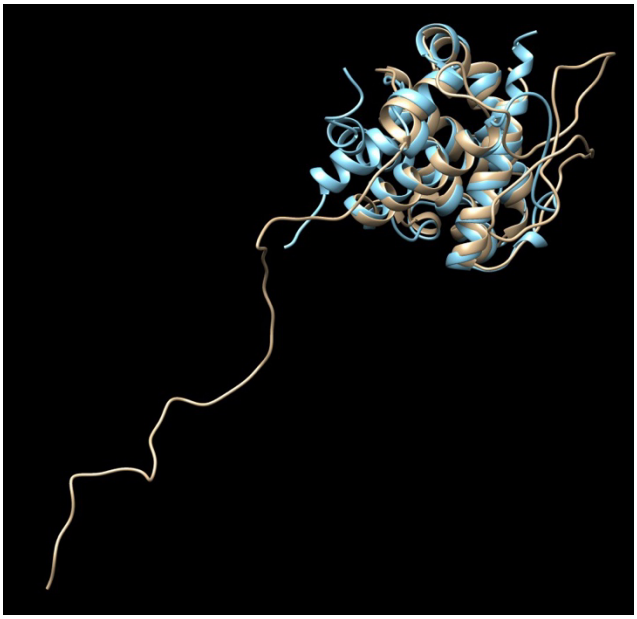

Figure 13. Matchmaker results of Model-M and Template-H

RMSD values of Model-M and Template-H is 1.392, and identity of the alignment between models is 67.43%. Here, Bcl-2-like protein from Homo sapiens and a model that is constructed from the Bcl-2-like protein from Mus musculus are compared. The alignment score is significantly higher than fourth comparison even though they are both mixed comparison, which means template and model are not come from same organism.

Multiple Sequence Alignment (MSA)

MSA is created by using Chimera, and order of the sequences is Model-M, Template-H, Model-H, and Template-M, respectively.

Figure 14. Multiple Sequence Alignment Results of four sequences

In MSA, percentage of aligned regions is less than pairwise alignment of models and template comparisons. The reason behind it is MSA is based on global alignment whereas pairwise alignment is based on local or global alignment. Thus, decreased percentage of aligned regions in MSA is expected. However, with MSA, detection of conserved regions is possible, and homology between sequences can be interpreted. Here in figure above, there are two significant alignments between four sequences, which are Model-M, Model-H, Template-M, and Template-H.

Final model with all templates

Model-M and Model-H were constructed from Template-M and Template-H, but for each model, only one template was used. Here, a new model is constructed from both Template-M and Template-H, and construction process and result are given below.

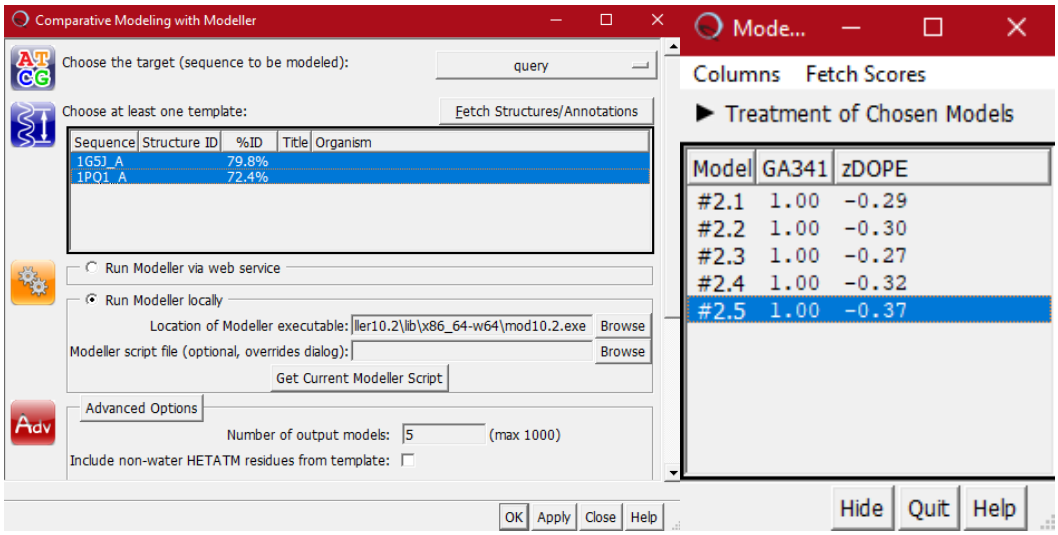

Figure 15. Construction of Final model processes from Chimera 1.6

After modeler calculations, model number five - final model, is selected and shown below.

Figure 15. Final model for Bcl-2-like protein from Gallus gallus

Comparison of Models and Final Comments

All the models are given below for the sake of comparison. Model-M is located on left, Model-H is located on the right, and all the models – Model-M is red, Model-H is blue, and Final Model is purple – are located on middle in the figure below.

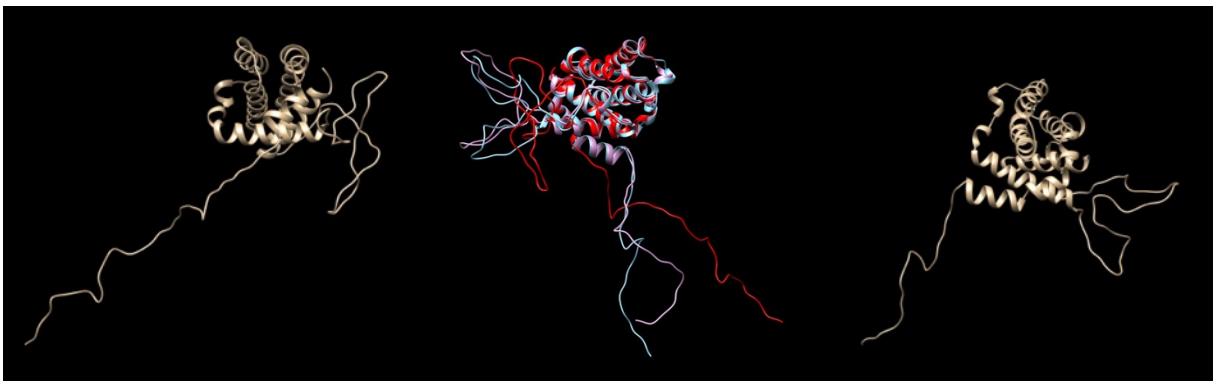

Figure 16. Presentation of Model-M, all models, and Model-H, respectively

Folded regions of all models are quite identical, but minor differences are available. However, unfolded regions of Model-M (red) are less similar than Model-H (blue). Reason behind it might be identity percentages of templates of Model-M and Model-H. High percentage identity from of template from Homo sapiens is observed.

Comparison of Final Model and Templates

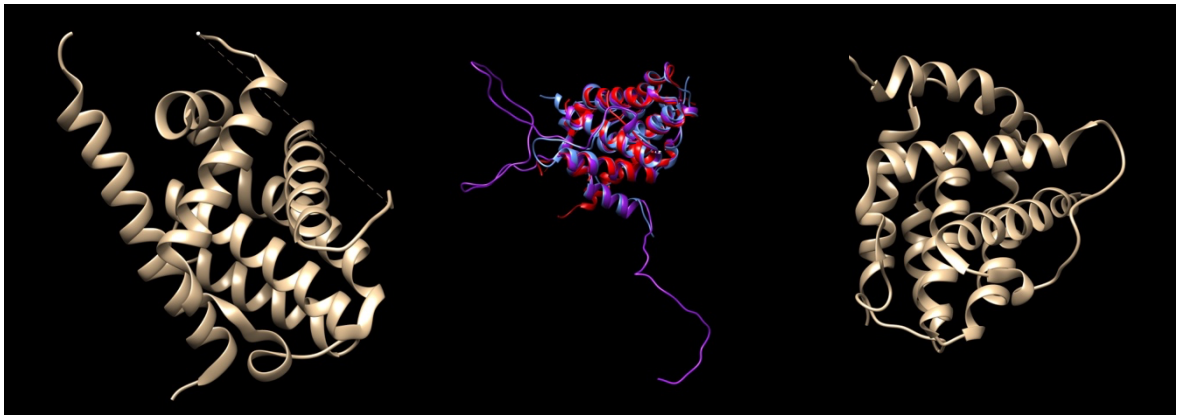

Figure 17. Presentation of Template-M, templates and final model, and Template-H, respectively

In the figure above, Template-M on the left, Template-H on the right, and Final model with templates are shown. Same colors are used for this figure, which is Template-M is red, Template-H is blue, and Final Model is purple. As observed, Template-H is more identical to Final Model, which some folded regions in Final Model do not match with the Template-M.

Final Comments in terms of all models

Observed values such as identity percentages of templates to our target protein sequence and RMSD outputs results will provide clues about choosing the best model for the research. Explanation of the values are given below.

Identity Percentages: Template-M has 72.4% identity, and Template-H has 79.8% identity. 7.4% difference between templates is available. Thus, using Template-H on applications in terms of modelling is more accurate in theory. As expected, Model-H gives more identity percentage to both templates.

	Model-M vs. Model-H	Model-M vs. Template-M	Model-H vs. Template-H	Template-M vs. Model-H	Model-M vs. Template-H
ID Percentage	61.56%	61.22%	76.00%	59.69%	67.43%

Table 1. Identity percentages of Direct and Cross Comparisons

RMSD outputs: According to ScienceDirect, more similar the two structures have smaller RMDS values. Here in the table below, models that are constructed with their corresponding organism gives more significant value rather than cross-comparison as expected.

	Model-M vs. Model-H	Model-M vs. Template-M	Model-H vs. Template-H	Template-M vs. Model-H	Model-M vs. Template-H
RMSD Values (A)	1.449	0.206	0.474	1.310	1.392

Table 2. RMSD Values of Direct and Cross Comparisons

After investigating the Model-M and Model-H, investigation of all models including Final Model is done, and results are given below. Investigations contain ID percentages and RMSD values.

Final Model Vs.	Template-M	Model-M	Template-H	Model-H
ID Percentage	60.20%	64.19%	76.00%	74.24%
RMSD Values (A)	0.803	1.200	0.942	1.181

Table 3. Identity Percentages and RMSD Values of Final Model Comparisons

In the table above, increased RMSD values of final model with respect to templates, the reason is that final model is constructed from both templates. This results in a mediation of RMSD values.

Overall, for the sake of the research, Model-H should be chosen, which has the highest ID percentage when it is compared with templates and final model. In addition, it has significant RMSD value among all comparisons. Moreover, both Template-M and Template-H are used for construction of Final Model, but identity percentage of Template-H with our target protein sequence was higher, and it covers nearly all regions of Template-M that are having homology with our protein of interest. Thus, preferring the Template-M or both for homology modelling is decrease the significance of the structure prediction, which can be observed at RMSD values of Table 3, and the model that is constructed only from Template-H is the best choice for the research, which is the Model-H shown below Figure 18.

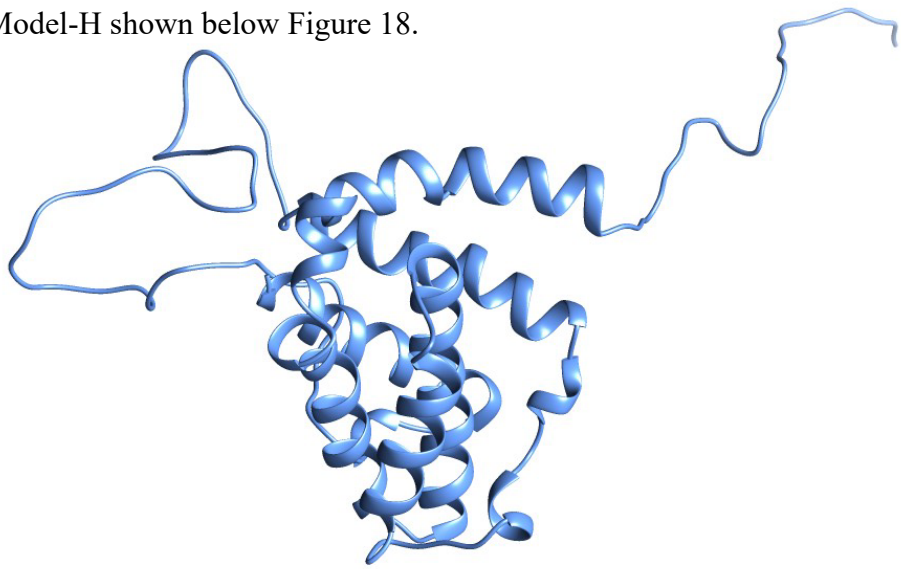

Figure 18. Presentation of Model-H

References

Bank, R. C. S. B. P. D. (n.d.). *1G5J: Complex of bcl-XL with peptide from bad*. RCSB PDB. Retrieved June 19, 2022, from https://www.rcsb.org/structure/1g5j

Bank, R. C. S. B. P. D. (n.d.). *1PQ1: Crystal Structure of bcl-XL/bim*. RCSB PDB. Retrieved June 19, 2022, from https://www.rcsb.org/structure/1pq1

Martz, E. (2001, June). *Homology Modeling for Beginners*. Homology modeling. Retrieved June 19, 2022, from https://www.umass.edu/molvis/workshop/homolmod.htm

Reva, B. A., Finkelstein, A. V., & Skolnick, J. (2005, July 2). *What is the probability of a chance prediction of a protein structure with an RMSD of 6 Å?* Folding and Design. Retrieved June 19, 2022, from https://www.sciencedirect.com/science/article/pii/S1359027898000194

UniProt ConsortiumEuropean Bioinformatics InstituteProtein Information ResourceSIB Swiss Institute of Bioinformatics. (2022, May 25). *BCL-2-like protein 1*. UniProt ConsortiumEuropean Bioinformatics InstituteProtein Information ResourceSIB Swiss Institute of Bioinformatics. Retrieved June 19, 2022, from https://www.uniprot.org/uniprot/Q07816