Indraprastha Institute of Information Technology Delhi (IIITD)
Department of Computational Biotechnology

BIO213 – Introduction to Quantitative Biology

ASSIGNMENT-3 (April 26, 2023)

Instructions:

1. You are required to submit the assignment by **May 3, 2023** (Wednesday).
2. Submit a single pdf file with all the properly labelled answers.
3. Use your Roll number and name to save the file.
4. Requests for extension of submission deadline will not be entertained.

Objective of this assignment is to get you acquainted with Modeller (https://salilab.org/modeller/), one of the most popularly used homology modelling tool.

The sequence given below is a part of human E3 ubiquitin-protein ligase for which structure has not been solved and therefore you will be developing a structural model for the same using the following instructions. During the modelling process answer the questions given below.

```
>protein
MALPAGPAEAACALCQRAPREPVRADCGHRFCRACVVRFWAEEDGPFPCPECADDCWQRA
VEPGRPPLSRRLLALEEAAAAPARDGPASEAALQLLCRADAGPLCAACRMAAGPEPPEWE
```

STEP 1. *Search for homologous proteins with solved structure to be used as template*

Firstly, use the NCBI blastp program available at the link https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp. Input the given FASTA sequence, choose 'Protein Data Bank proteins (pdb)' as the database and blastp as the algorithm. Now submit the job, wait for the results and then analyze it carefully.

Question 1. Which sequence can serve as the best template for modelling the E3 ubiquitin-protein ligase structure? Give reason for the same. Use the parameters like score, identity, similarity, query coverage, E-value, etc. to make the choice. **(10 marks)**

Question 2. Show the alignment of the chosen template with your query protein. Is there any region of the query that is not being covered by the template? If yes, mention the residue numbers. Use the graphical summary on the results page to check if any other sequence can serve as a template for the uncovered region or not. **(5 marks)**

STEP 2. *Retrieve the PDB structure of the chosen template*

Note the Accession of the chosen template from the blastp results page. The first four characters correspond to the PDB ID, and the character after the underscore represents the chain of PDB aligned to your query sequence. Copy the PDB ID (only the first four characters) and search for

the structure in the Protein Data Bank at https://www.rcsb.org/. Open and explore more about the template structure. Download the structure in PDB format.

Question 3. Which experimental method was used to solve this structure? How many total chains are there in the structure? Are the other chains different from the chain of your interest?
(5 marks)

STEP 3. *Prepare the files in format supported by Modeller to carry out further steps*

You will be using the basic modelling approach here. Follow the tutorial available at https://salilab.org/modeller/tutorial/basic.html. Download the example input and output files, which can be modified and used for the modelling of your protein ligase. As described in the tutorial prepare the protein sequence file in PIR (.ali) format.

Note: build_profile.py and compare.py help in selecting the template as you did in the previous steps. Skip these for now as you have already chosen the template protein.

STEP 4. *Align the query and template using align2d.py*

This help in the alignment of the query to the template sequence, taking into consideration its structural information as well. Follow the instructions provided in the tutorial for alignment of the two sequences.

STEP 5. *Model building*

As described in the tutorial, use model-single.py to build the model. By default, 5 models will be generated. Check the summary in the log file to get more information related to the generated models. Choose the best model.

Question 4. Which were the two default parameters or objective functions on which you chose the best model here? Give the significance of both. **(6 marks)**

STEP 6. *Visualization of the developed model*

Though any viewing platform can be used, you will be using Chimera that is freely available at https://www.cgl.ucsf.edu/chimera/download.html. Open the best model in Chimera.

Question 5. Compare the structure of your model to the PDB structure (3D view is also available with each entry). Does it carry similar structural folds? Provide a screenshot of the modelled structure. **(10 marks)**

STEP 7. *Evaluation of the developed model*

We discussed different methods for evaluating the correctness of the structural models. Here you will be generating only the Ramachandran Plot to assess if any amino acids fall in the disallowed region.

Use the PDBsum service (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) to generate the Ramachandran plot for your model. Upload the .pdb file of the best model generated

using Modeller and provide your e-mail ID. The link to the results page will be shared with you once ready. Click on the image depicting the Ramachandran plot to get further information.

Question 6. Provide the plot obtained and briefly discuss the results. Do you think it can be a reliable structure that can be used for other studies? **(10 marks)**

Advanced modelling

Question 7. Do you think using multiple templates (or multi-template homology modelling) could have resulted in a better structure? Justify your answer. **(4 marks)**