



# LEARN PYTHON & R FOR BIOINFORMATICS

## Prerequisites:

### 1. Modeller:

You must install the Modeller program on your PC before moving on. It is a commercial program that is why we cannot help you with installation.

### 2. Alignment and structure prediction scripts.

### 3. Code Editor.

### 4. UCSF CHIMERA or any other tool for protein structure visualization.

## Prerequisite terminologies:

In order to have a thorough understanding of our main topic, you should have the basic concept of the following terminologies:

### 1. Target sequence:

The query sequence whose structure is unknown.

### 2. Template sequence:

The protein sequence whose structure is known experimentally.

### 3. Homology Modelling.

## Introduction:

**MODELLER** is used for homology modelling that is used to build tertiary and quaternary structures of proteins. The user provides an alignment of the sequence to be modeled with known related structures and **MODELLER** automatically calculates a model of your target protein sequence.

## Steps:

- Prepare your query/target protein sequence in a FASTA file.

[You should make a folder where you can place all the files you need for Homology Modelling of your protein through MODELLER, i.e., FASTA sequence files of Template and Target proteins, Protein structure of your Template protein in PDB format, Alignment and structure prediction scripts.]

- To select the most suitable template sequence that will get aligned with our target protein sequence, we'll carry out Protein BLAST (blastp) and we'll select following parameters:

<b>Query sequence</b>	FASTA file of Target protein sequence.
<b>Database</b>	Protein Data Bank (pdb)

[For other options go with default parameters if you don't have to change them according to your research purpose].

- Click on BLAST.
- From the result page of BLAST, select a protein homolog who has:

Query coverage value	$\geq 60\%$
Percentage Identity	$\geq 30\%$

- Click on the suitable homolog and from the resulting page click on the 'Structure' option present on the right side of the page. [It'll directly take you to the MMBD page where it provides you the protein structure of your selected protein.]  
**Note:** To have a better understanding of BLAST results, kindly watch our tutorial on BLAST.
- On the right side of the page it provides you the PDB ID of your template (homolog) protein, just click on the accession number present against that PDB ID.
- Click on the 'Download' option and select the FASTA format and then place this FASTA file in the folder you created before (i.e., for Homology Modelling through MODELLER).
- Again click on the 'Download' option and select PDB format and place this file in the same folder.  
**Note:** For better understanding of different file formats and their uses kindly watch our videos from 'File Format' courses.
- Open your Template sequence file (FASTA format) in the Code Editor and remove the chain of the protein that did not get aligned with the target sequence (if any).
- Open your template sequence file (pdb format) in UCSF Chimera or any other tool for protein structure visualization and check the number of chains of your template protein and remove the one which didn't get aligned with your target sequence (if any). To do so:

- Click on the 'Select' option in the menu bar of Chimera, then click on 'Chain' and then non-aligned chain.  
[It will highlight the selected chain.]
- Then click on 'Actions' , then 'Atom and Bonds' and then 'Delete'. And save it or replace it with the original file.
- Now remove the ligands that are attached to your template protein by clicking on 'Residue' then 'All non-standard residues'.
- Then click on 'Actions' , then 'Atom and Bonds' and then 'Delete'. And save it or replace it with the original file.  
[For better understanding of UCSF CHIMERA, kindly watch our tutorial on it.]
- Copy your Alignment script and Structure prediction script files into the folder where you stored all your template and target sequence files, so that you can make a preparatory environment to do Homology Modelling through MODELLER.
- Open all the files (i.e., FASTA sequence files of template and target proteins, template protein sequence in pdb format and the script files) in the Code Editor.
- Now to make a query alignment file, open your query.ali or query alignment file in Code Editor and then paste your query sequence there. Make sure there is an asterisk (\*) at the end of your query sequence, that marks the end of the sequence.
- To make template alignment file, open your template.ali file or template alignment file and then paste your FASTA sequence file of your template protein there. Also put your template PDB accession there.
- To make the Alignment script of your template and target sequences, open align2D.py file (a Python script) in the Code Editor and then paste the PDB accession number wherever it is required.

[It will align the query alignment file (target sequence file) against the template alignment file and provides output in *PIR* and *PAP* format]

**Note:** For better understanding of file formats, kindly watch our tutorial on 'File Format' courses.

- To make the Structure Prediction script, open script.py file (Python script) and edit the variable "knowns='templateStructurePDB'", cut its value and paste the PDB accession number of your template protein there. And save the file.

[To have a better understanding of Python scripts and how it works, you should join our GOLD Bioinformatics courses.]

- Copy your folder address, open MODELLER, paste your folder address there after the chain directory (cd).

[In this way your directive will change into the environment you created in the specific folder where you saved all your files for Structure Prediction through MODELLER].

- To align your Target and Template sequences through MODELLER, type "mod9.20 align2D.py" (without double quotes) and press ENTER.

[It will provide you 2 new files 'output.ali' and 'output.pap' in the same Environment folder you've created]

- Open both output files(i.e., output.ali and output.pap) in the Code Editor, it will show you the alignment between your target and template sequences in PIR and PAP formats respectively.

- To build protein models of your Target protein sequence through MODELLER, type "mod9.20 script.py" (without double quotes) and press ENTER.

[It will start creating new files containing the possible Target Protein Structures predicted via MODELLER, in the environment folder you've created.]

- You can open all your predicted protein models in UCSF CHIMERA to visualize your predicted structures.
- The last step is **Model Evaluation**, you can evaluate your models using different tools but MODELLER also has its own model evaluation method that is stored in 'script.log' file.

- Open 'script.log' file on your Code Editor and search for the molpdf folder. Here it shows different values that are assigned by the MODELLER to each of the protein structures predicted.
  - The least value represents the most good predicted model that MODELLER defines itself.  
[ So, in this way you can predict thousands of models and select one based on the least value given by MODELLER evaluation.]

### **Summary:**

In this video tutorial of MODELLER for protein structure prediction, we came to know about the basic procedure to predict the protein structure from a Target protein sequence, in a professional way. We also learned about the Evaluation method of MODELLER to find out the most optimal and good protein structure predicted.