Practical courses 4 and 5
3D structure prediction

This practical course focuses on the modelling of protein structures by comparative modelling and by fold recognition. You will create different models for the acyl carrier protein of *Rhodospirillum centenum*. The quality of your models will be evaluated with the global Qmean score and Procheck. These tools are available here:

Qmean: https://swissmodel.expasy.org/qmean/

Procheck, via "PDBSum Generate":

https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html

What is the Qmean score based on and how to interpret its value?

1. **Comparative modelling of the 3D structure of the acyl carrier protein of *Rhodospirillum centenum*: manual approach**

The uniprot code of the acyl carrier protein of *Rhodospirillum centenum* (ACP) is B6IN76; its sequence is available in FASTA format on Uniprot (http://www.uniprot.org). In this first section, you will do manually each step of a comparative modelling: search of possible templates, selection of a template, sequence alignment between the target and the template, modelling and evaluation of the quality of the model.

A] Perform a Blast (http://www.ncbi.nlm.nih.gov/blast/; use "Protein blast") on the ACP sequence to identify a template to model the structure of this protein (**choose the appropriate database to scan with BLAST**).

B] Select a template among the hits that have been identified by Blast. For that purpose, take into account the percentage of sequence identity, the percentage of query cover and the quality of the structure of the template (see the tools used in TP1). Download the sequence of the template from the PDB website.

C] Perform a sequence alignment between the template and ACP. For that purpose, use the the Stretcher global alignment program (https://www.ebi.ac.uk/Tools/psa/emboss_stretcher/, choose the "Markx3" output format in the "More options" menu).

D] Submit the sequence alignment obtained in the section 1.C. to the Modeller server (https://toolkit.tuebingen.mpg.de/#/tools/modeller). The license key to use Modeller is "MODELIRANJE". The sequence alignment must be provided in PIR format. There is on the

virtual university a document that explains how to convert the Markx3 format obtained from the sequence alignment into a PIR format that must be submitted to Modeller.

Save the PDB file of the model.

E] Analyze the quality of this model.

2. **Comparative modelling of the 3D structure of the acyl carrier protein of *Rhodospirillum centenum*: semi-automatic approach**

The HHPRed server combined to Modeller (https://toolkit.tuebingen.mpg.de/#/tools/hhpred) will be used.

A] Submit the ACP sequence to the HHPred server. Describe the first step performed by HHPred. Several templates are proposed. Compare these templates to those identified with Blast in section 1.B. (sequence identity, quality of the template structure, …).

B] Select one template and click on "Model using selection". HHPred will align the 2 sequences. Then click on "Forward to Modeller", use the Modeller-key "MODELIRANJE" and click on "Submit job".

Save the PDB file of the model (you will find a "Download PDB file" tab).

C] Analyze the quality of this model.

3. **Comparative modelling of the 3D structure of the acyl carrier protein of *Rhodospirillum centenum*: automatic approach**

Use the SwissModel server (https://swissmodel.expasy.org) to build a model of ACP. Use the "Build Model" tab.

Identify the template that has been used. Save the PDB file of the model.

Analyze the quality of this model.

4. **Modelling of the 3D structure of the acyl carrier protein of *Rhodospirillum centenum*: by a fold recognition approach**

A] Submit the sequence of ACP to Sparks-x (http://sparks-lab.org/yueyang/server/SPARKS-X/). What is the best template according to Sparks-x? Download the model obtained with this template (click on "MOD1" to download the model).

Analyze the quality of this model.

B] Submit the sequence of ACP to genTHREADER (http://bioinf.cs.ucl.ac.uk/psipred/, **choose only GenTHREADER-Rapid Fold Recognition**). Select the first template proposed and build a model. Download the PDB file of the model.

C] Analyze the quality of this model.

5. **Protein Model Portal**

The Protein Model Portal (http://www.proteinmodelportal.org) is a website that collects models that are already available for a protein.

A] Search the Protein Mode Portal for existing models for ACP (Uniprot code B6IN76). How many models are there?

B] Select all the models and click on "Start analysis of structural variability" to compare the structure of the different models. Analyze the results.

C] Download the PDB file of model 5. For that purpose, click on "Show", then on "Model provided by NESG", then on "download: model". Analyze the quality of this model.

6. **Comparison of the models**

You have now 6 models for ACP: 3 models obtained by comparative modelling, 2 models obtained by fold recognition and 1 models from the Protein Model Portal.

A] Compare the quality of these models.

B] Open the 6 PDB files of the models in a same window in Pymol.

Compute the rsmd between all pairs of structure. For that purpose, you can use the "super" command in pymol. To superimpose the main chain atoms of struc1 to those of struc2 and to compute the rmsd, for instance, use the command "super struc2 and name ca+c+o+n,struct1 and name ca+c+o+n,cycles=0". Read the documentation about the "super" command. Note that Python can be used to write scripts in Pymol or to define new functions. Writing a script could be a good idea to perform this repetitive task. You will find a lot of information about this on internet (Pymol tutorials, existing scripts, …).

Group the models according to their structural similarity.

C] Use the "super" command, to superimpose all the models on one the 6 models and to identify the regions where the structural differences are the largest.