Exercise 1.

We are going to create a model for a protein that has no known structure.

Our target protein is the Arginine N-succinyltransferase (AST) protein from the species Yersinia pestis. The protein id is ASTA_YERPE.

First, check out the annotation for this protein in the Uniprot database, does it have a known structure? Search for the protein at uniprot.

Check if there is a model for this protein in the MODBASE database.

There is a link from the Uniprot page, but you can also directly search from the ModBase home page using the Uniprot id.
http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi

What are the characteristics of the model?
- What was the template?
- What part of the target sequence is covered by the model?
- What is the sequence similarity?

You can get more information about the parameters by clicking on the question mark next to each features.

The model quality is characterized by two parameters MPQS and Z-DOPE. Based on the information, do you think the model is reliable? (Click on the help).

MODBASE created multiple models. You can see their properties if you move the mouse over the structures. Are all those models based on the same PDB? What about the id%? What is

the advantage of getting several models?

Now we will look and compare all the models produced for this target in Chimera.
You can download the models using chimera

Chimera
Fetch Stuctucre by ID
ModBase → írja be a uniprot kódját
Fetch

Or you can directly open chimera by selected the three models and select
perform action...
chimera structural alignment

In which regions the models differ the most?

How reliable are regions between positions 198 és 210?

Exercise 2.

Modelling the human cyclin A1 using the SWISSMODEL package.

Cyclins are eukaryotic proteins with fundamental roles in cell cycle regulations. Cyclins can be grouped into groups A, B, C, D. These bind different types of cyclin dependent kinases at different stages of cell cycle. The type A cyclins in mammals have two groups, A1 and A2. Cyclin A family consists of two members, Cyclin A1 and Cyclin A2. While Cyclin A2 is widely expressed in different tissues, Cyclin A1 is limited to male germ cells.

To date no experimentally determined 3D structure of Cyclin A1is available. Create a model using the Swissmodel portal.

http://swissmodel.expasy.org/

Find the sequence of human cyclin A1 at uniprot.
Upload the sequence to the swissmodel site and start the search for templates.

Which part of the protein do the templates match? What type of domains do they belong (e.g based on Interpro) ?

Click on the Sequence Similarity tag. How are template matches clustered? Which group do they belong to?

How good is the alignment? How many gaps are there?

Choose a structure and build the model. Repeat the model building with a different structure and compare the models.

If you would like to model how cyclin A1 binds ATP, which template would you choose?

Check model quality, which regions have lower model quality ?

Exercise 3.

HHpred
http://toolkit.tuebingen.mpg.de/hhpred

We will build a model for the human Gadd45β (O75293) using a powerful search technique called HHpred to identify possible protein templates with known 3D structure.
Hhpred (https://toolkit.tuebingen.mpg.de/hhpred)

Find protein the protein in uniprot. Can you find any structural information for it?

Model the protein using the HHPred method!

Paste your sequence into the search box. Basically the only option you can change is the template database. You would like to predict the structure, which one would you choose?

Start the search. In a few minutes you will get the results with the best hits.

Which is the best hit? What is the E-score and sequence identity for the best hit?
Is this a good enough similarity to build a model?
How well the secondary structure prediction for the template agrees with the secondary structure of target protein?

Build the model!
Choose the best alignment for this.

Modeller code: MODELIRANJE

Examine the model using different quality scores. are there regions with where the model seems less reliable?

Save the model. Read it into chimera, does it look like a good structure? For examples, can you see H-bonds? Are the hydrophobic amino acids buried, as expected?