Homology modelling with Swiss-Model

Structural bioinformatics in practice (Færøyene 2016)

Lars Skjærven, University of Bergen

Overview

This tutorial provides a basic introduction to homology modelling. Homology modelling is a computational approach for modelling the 3D structure of a protein using the knowledge of a related homologous structure which has been resolved experimentally. The technique effectively maps the *target* sequence (i.e. the sequence of the protein that you want to model) onto the homologous *template* protein structure. Homology modelling consequently relies on (1) the availability of a homologous structure, and (2) a relatively high sequence identity between the template and the target (> 30%).

In this tutorial we will use SwissModel [http://swissmodel.expasy.org] to build the structure of *Salmo salar* variant of DHFR. Note that you will find more extensive explanations here [https://swissmodel.expasy.org/docs/help].

Building the *Salmo salar* (Atlantic salmon) variant of DHFR

DHFR is an important drug target and the structure has been resolved for several species which are available through the protein data bank (PDB). Being a country dependent on salmon we're interested in learning more on how the structure of *salmo salar* DHFR looks like. This is potentially useful in designing new anti louse drugs that does not target the salmon, i.e. can we use the structures of salmon and salmon louse to design specific inhibitors that target and kills the louse?

Swiss-model provides a neat web interface for rapid homology modelling. Go to the Swiss-model website [http://swissmodel.expasy.org/], and click the **Start modelling** button. This will take you to the first step of the application - namely to provide the sequence of protein you want to model (the *target* sequence).

To build the salmon DHFR we therefore need to obtain the sequence of this enzyme variant. Fortunately we can find this at the UniProt web site [http://www.uniprot.org/]. Go there, and enter "salmon DHFR" in the search text field on the top of the site. From the list of results (Figure 1) copy the "Entry name" of *Salmo salar* (**B5XG28_SALSA**) and paste this in the **Target sequence** input box on the Swiss-model web site.

Entry	Entry name		Protein names	Gene names	Organism	Length
D3PJE0	D3PJE0_LEPSM		Dihydrofolate reductase	DYR	Lepeophtheirus salmonis (Salmon louse)	180
B5XG28	B5XG28_SALSA		Dihydrofolate reductase	DYR	Salmo salar (Atlantic salmon)	190
B5X910	B5X910_SALSA		Dihydrofolate reductase	DYR	Salmo salar (Atlantic salmon)	190

Figure 1: Results table of search for the salmon DHFR sequence in the Uniprot database.

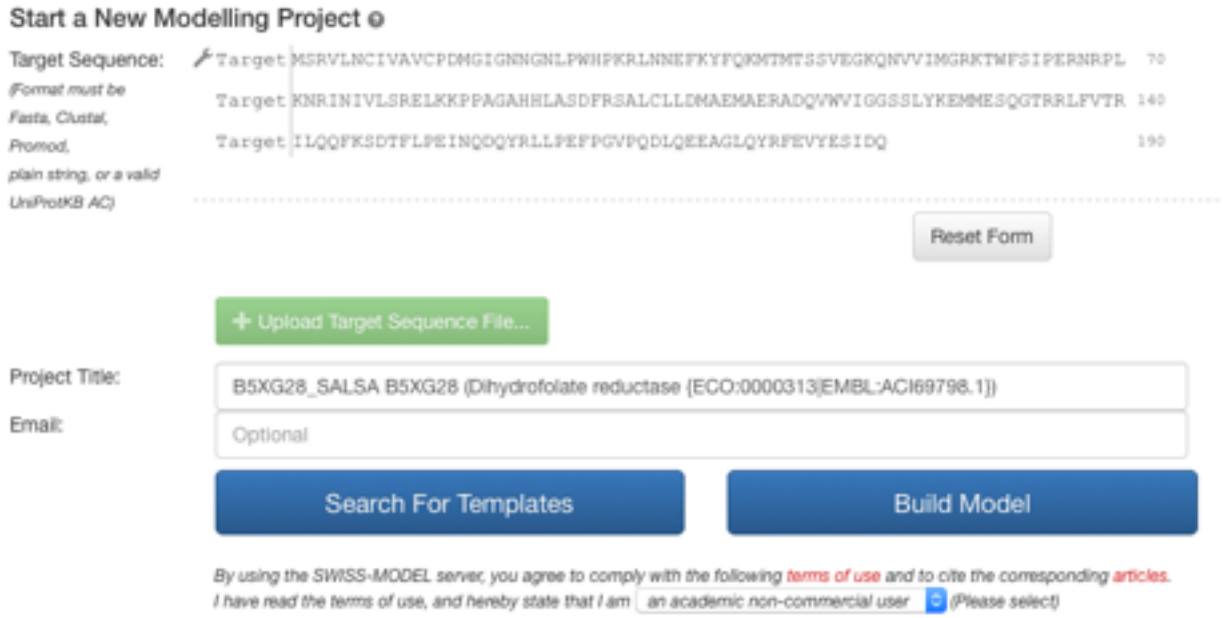


Figure 2: Here we have entered **B5XG28_SALSA** to the target sequence dialog. The sequence appears automatically when entering the uniprot identifier.

Observe that the sequence of the salmon DHFR appears in the target sequence input box. Next step is to search for template structures in which our model shall be based. Click the big blue **Search for templates** button. When the search is complete a list of potential template structures appears:

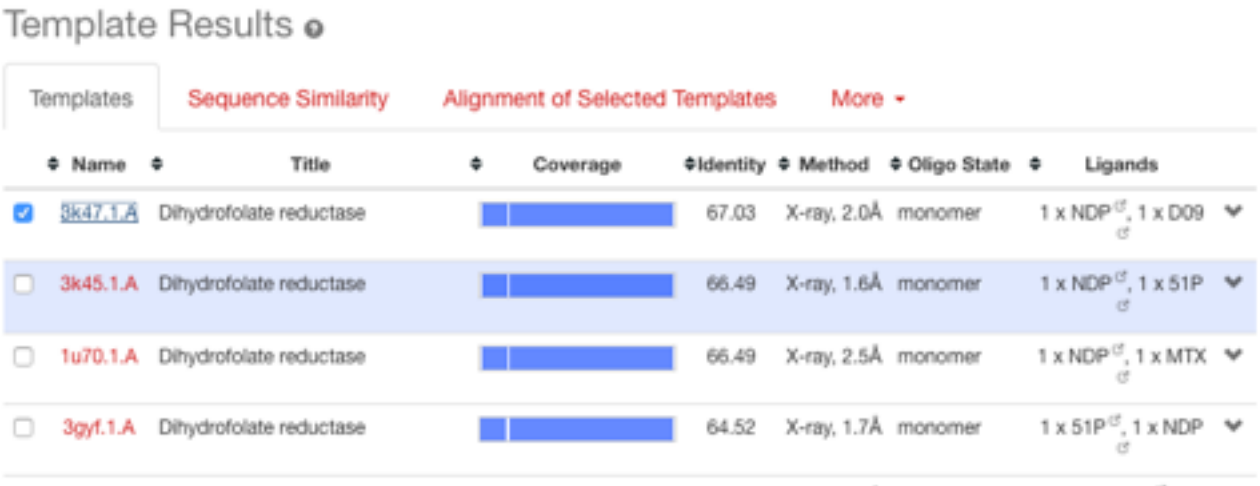


Figure 3: Results of template search. PDB ID 3k47 can be used for template for model building.

The columns provides information such as the sequence identity between the target sequence and the template structure. In this case a template structure with sequence identity of **67%** has been identified. Click the name / identifier of this top hit (**3k47**) to explore this structure in more detail, or go to pdb.org to figure out which species / variant of DHFR this is.

Make sure this entry is selected, before you continue by clicking on the "**Alignment of Selected Templates**" tab.

Figure 4: Sequence alignment between the target sequence (salmon DHFR) and the template (mouse DHFR).

As the title indicates this tab shows the alignment between our input sequence and the mouse DHFR structure which we have chosen to be our template for the model building. Note the two additional residues (Lys and Arg) at position ~30 in the target sequence which is not present in the mouse sequence.

Task: Inspect the sequence alignment of the target and template. Identify the area(s) of major sequence variation. Can you also point to a few columns in which equivalent positions have opposite charge?

Next, click the large blue **Build models** button on the right hand side. This will start the model building of the salmon DHFR. Once the model is complete the "Model Results" appear (**Figure 5**).

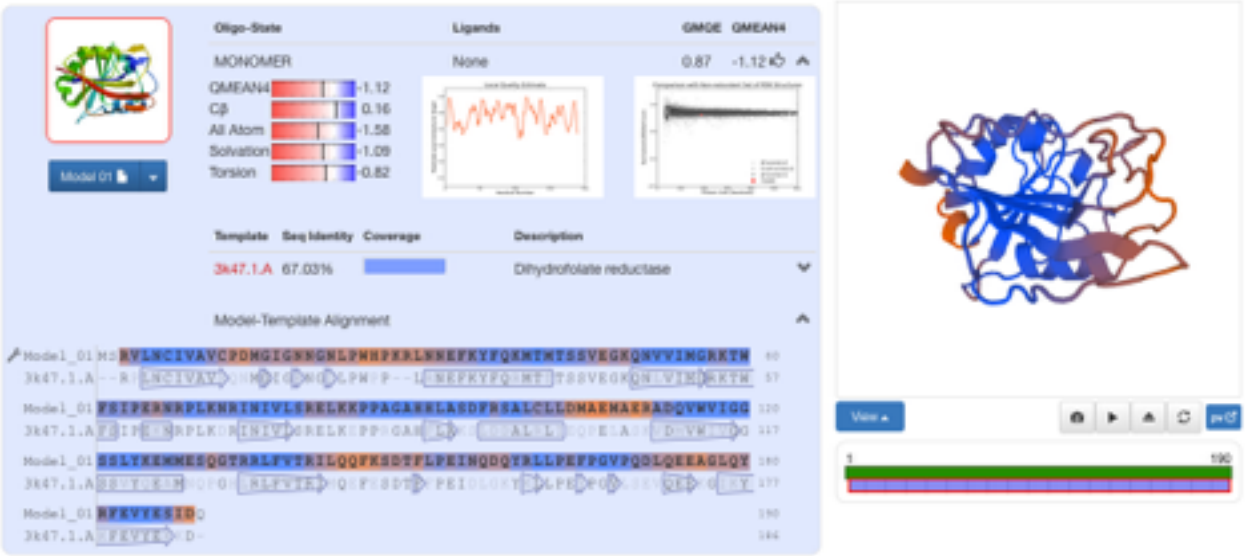
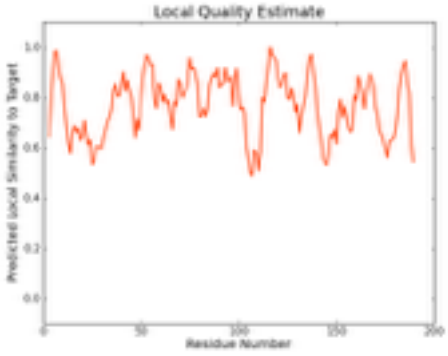


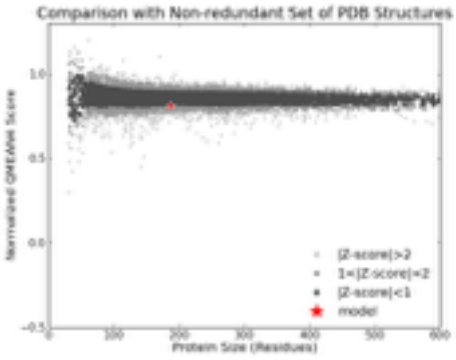
Figure 5: The model results section. The model is displayed on the right hand side coloured according to a local quality measurement.

In this section you can access the newly built model, as well as estimates of the global and local model quality. The calculated model is presented (on the right hand side) initially coloured by model quality assigned using the QMEAN measure [read more]. This allows instant visualisation of regions of the model which are well or poorly modelled. Here, blue depict areas of higher score (more reliable) while red depict lower score (less reliable). Note that you can change the color of the alignment and the corresponding model by clicking on the small adjustable spanner icon next to the alignment [see colouring scheme here].

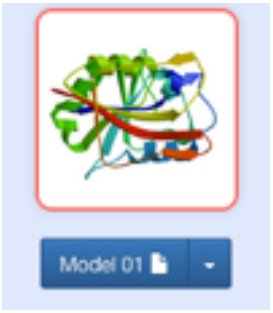
There are two additional plots giving information on the model quality:

- The first plot gives a local quality estimate per residue based on the QMEAN scoring function. This measure is from 0 to 1 where higher value indicate more reliable model.
- The second plot provides a comparison of the obtained global QMEAN score in relation to a large set of high-resolution PDB structures.





Download the model in PDB format by clicking on the blue button ("**Model 01**"). Go to **File → Save Page As**. Open this model PDB file in PyMOL together with the template (i.e. use command "fetch 3k47" after loading the model). Use your skills from the PyMOL tutorial to align the two proteins (**A → align → to molecule → 3k47**).



Task: Investigate the model of the salmon DHFR and identify the area of the protein with the inserted Lys and Arg.

Task: Zoom to the folate binding site (hint: look for inhibitor with name D09 in the 3k47 structure). What can you say about the conservation of amino acid in the folate binding site?

~~**Pre-computed results**: http://swissmodel.expasy.org/interactive/vY2Nqb/models~~

Optional: Salmon louse DHFR

Task: Model the salmon louse DHFR and compare the results in PyMOL.

Summary

In this tutorial we have used basic homology modelling with the SwissModel interface to generate a model of salmon DHFR. We used the mouse DHFR variant as template structure which shows a sequence identity of 67% towards the target sequence. We identified a few inserts and multiple mutations to charges residues. The resulting model shows a quality measure close to structures resolved by X-ray crystallography.