5BBB0226 Week 2
Tutorial: Protein Modelling

All online quizzes are only activated during the tutorial. A copy of the questions and suggested answers in the quizzes will be uploaded to KEATS for your revision.

In protein modelling, we use an existing experimental structure of a protein as a template in order to build a model of a different, related protein. Based on the fact that related protein domains have similar structures, relying on homology often yields prediction of the correct fold of proteins with unknown structures. Here, we are going to be using the **protein kinase** domain of **A-Raf** as our protein of interest.

Exercise 1: Navigating the UniProt database

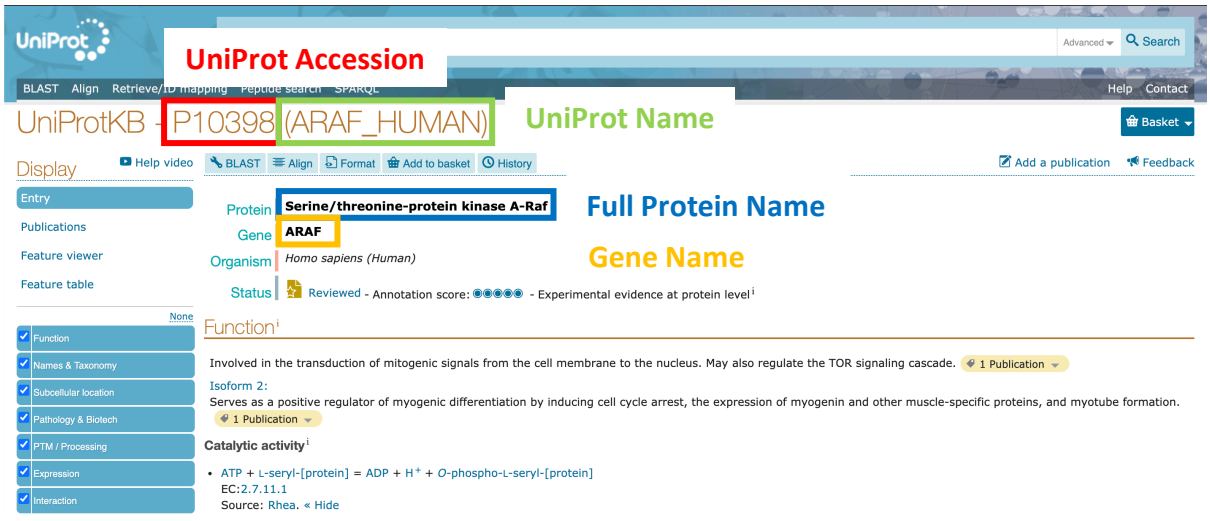
Please follow the steps below and answer this quiz as you go along:
https://PollEv.com/surveys/BcQZ9GHBxkmJ9nT3g6Jud/respond

In this exercise, we will use web-based tools to find and read about our protein of interest, and to find homologous protein structures that can be used as templates for homology modelling.

1. Go to the UniProt website (https://www.uniprot.org).

2. Use the Search bar at the top of the page to search UniProtKB for the human version of A-Raf. The specific protein should be named ARAF_HUMAN.

3. You should see a page like the one below. Note that there are several ways to refer to this protein:



Read through the ARAF_HUMAN page and see what information is included in this database. You will see a panel on the left-hand side allowing you to jump to any specific subsection.

4. Use the panel to jump to the "Family & Domains" section. You will see that the ARAF protein has been annotated with two domains. Click on the "Protein kinase" domain positions.

5. You should see a sequence in FASTA format similar to the one below:

```
>query
SEVQLLKRIGTGSFGTVFRGRWHGDVAVKVLKVSQPTAEQAQAFKNEMQV
LRKTRHVNILLFMGFMTRPGFAIITQWCEGSSLYHHLHVADTRFDMVQLI
DVARQTAQGMDYLHAKNIIHRDLKSNNIFLHEGLTVKIGDFGLATVKTRW
SGAQPLEQPSGSVLWMAAEVIRMQDPNPYSFQSDVYAYGVVLYELMTGSL
PYSHIGCRDQIIFMVGRGYLSPDLSKISSNCPKAMRRLLSDCLKFQREER
PLFPQILATI
```

*N.B. the sequence shown above is in the **FASTA format**. This is a standard format for writing biological sequences: we first specify a title with ">", followed by the sequence (represented by the string of characters printed without spaces).*

Comparing the sequence you see on UniProt and the one above, they may start and/or end at different positions – for this exercise we shall use the sequence printed above.

Save the above sequence as a separate FASTA format file (*.fasta) using a text editor.

N.B. note that some text editors may try to save your file with an additional .txt extension – this may cause problems down the line).

Now that we have located our query sequence to model, the next steps are to look for a suitable template and use it to perform homology modelling of our query protein.

Exercise 2: SWISS-MODEL automated mode.

In these tutorials, we will use the webserver SWISS-MODEL to perform homology modelling.

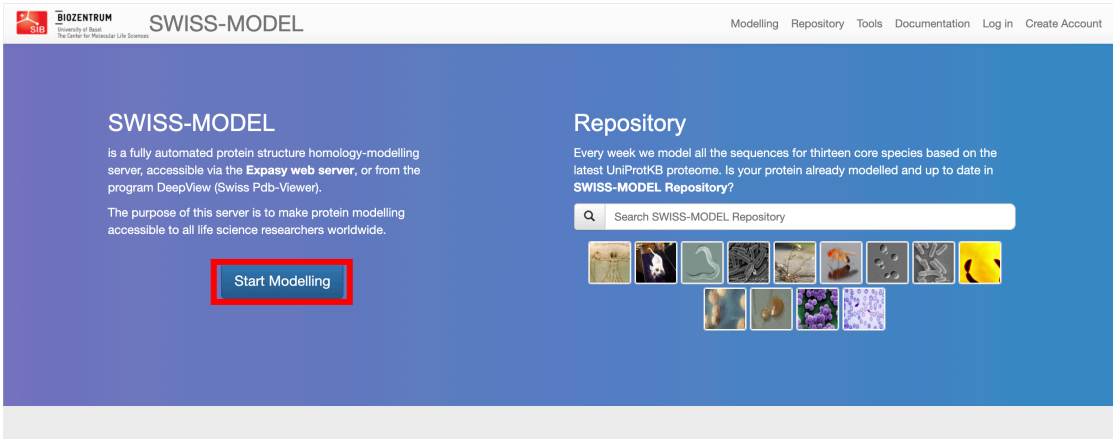
One method to obtain a homology model of your query is to use the 'automated mode' in SWISS-MODEL. Here, you submit your query sequence, and SWISS-MODEL will automatically perform a template search and build a model using one of the templates.

(N.B. As detailed in the slides, you could alternatively use SWISS-MODEL under the 'alignment mode'. We will cover this later & next week.)

Please follow the steps below and answer this quiz as you go along:
https://PollEv.com/surveys/WnwGQG4FJrNYjH1naDXRO/respond

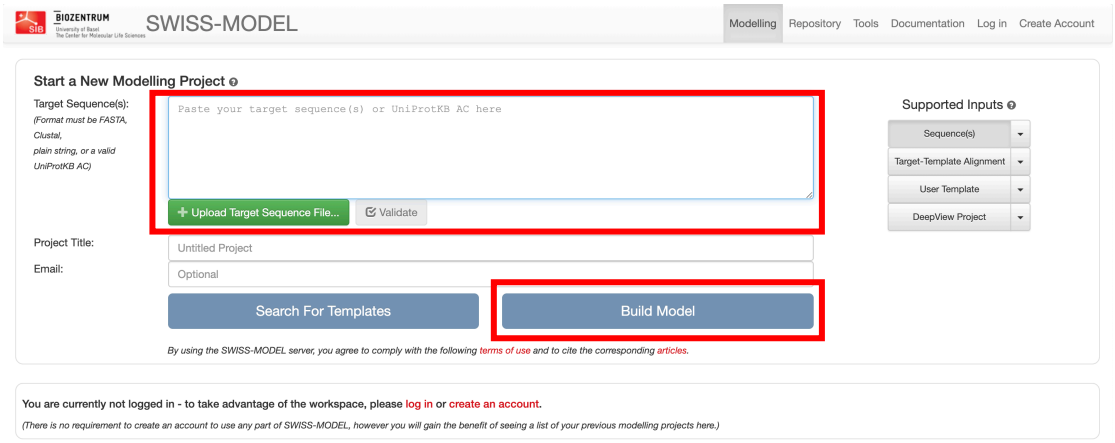
Here are the steps to perform modelling using the SWISS-MODEL automated mode:

1. Navigate to the SWISS-MODEL web server (https://swissmodel.expasy.org/).



Automated mode

2. Click the "Start Modelling" button. You will be directed to the page below:



3. Under "Target Sequence(s)", paste or upload the previous sequence (provided in Step 5 in the UniProt exercise) for the ARAF_HUMAN protein kinase domain in FASTA format and click "Build Model". Leave SWISS-MODEL to finish building the models while you continue the exercise (should take 5-10 minutes).

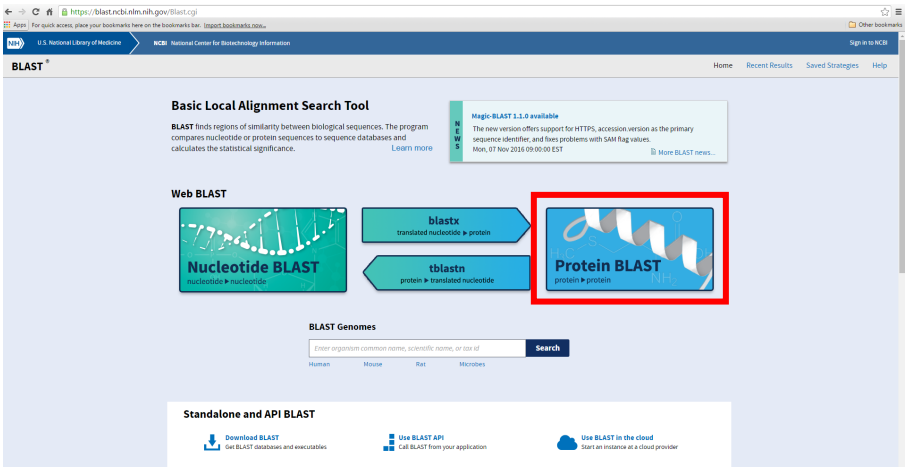
Take some time to study the results. SWISS-MODEL provides useful help pages to explain the meaning of the results – please click the 'Question marks' next to headings for help.

Exercise 3: Do your own Template Search using BLAST

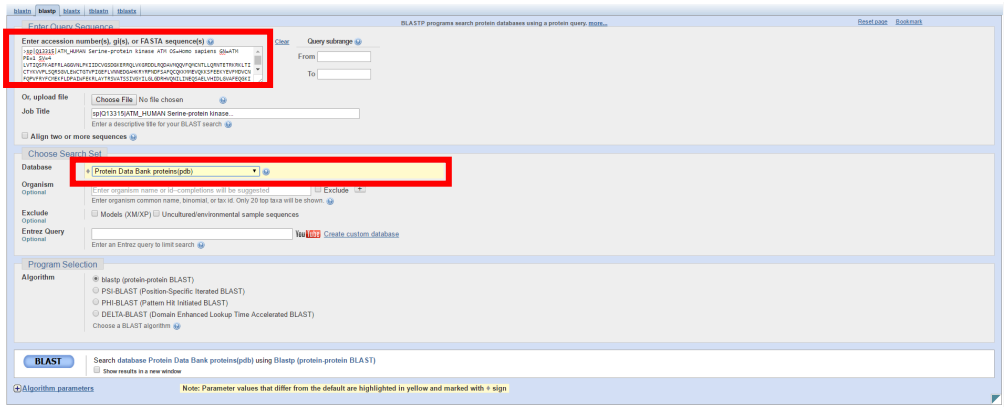
We have seen in the automated mode of SWISS-MODEL it internally performs a template search and selects (what the algorithm thinks) the 'best' template to build a model. Each of the steps 'hidden' behind SWISS-MODEL can be done on your own – this will help you understand the importance and implication of each of the detailed steps. In these tutorials we will break down these steps. For the remaining time in this tutorial we will show you how you can do your own template search using the Basic Local Alignment Search Tool (BLAST).

Please follow the steps below and answer this quiz as you go along:
https://PollEv.com/surveys/rO9LiipPhYWmos8GwTvXs/respond

1. We will use the Basic Local Alignment Search Tool (BLAST) to search for homologous sequences to our protein of interest. Navigate to the BLAST website (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

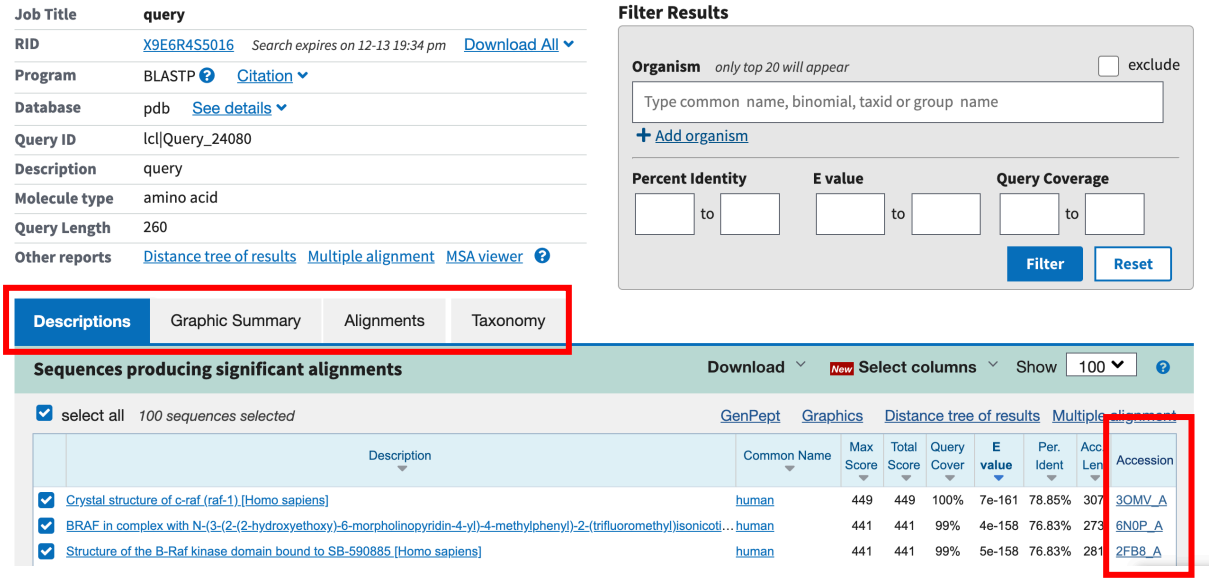


2. Click on "Protein BLAST" to be taken through to the blastp suite.



3. Under "Enter Query Sequence", paste or upload the previous sequence for the ARAF_HUMAN protein kinase domain in FASTA format, and add a job title (optional). Under "Choose Search Set", change the Database option to "Protein Data Bank proteins (pdb)". Note that several different algorithms and parameters are available.

4. Click "BLAST" and wait for the algorithm to finish running.



5. You will receive an output that looks like the one above. You should see a hit list under "Descriptions", a graphical summary of results, and individual alignments between query sequence and hits. Study the outputs carefully.

   All hits should be experimentally determined structures – you can read details about these structures from the entry page on the Protein Data Bank (PDB), the database which holds protein structural data. Each structure is represented by a four-character code under "Accession" in the hit list (highlighted in the screenshot above, e.g. ABCD_A = PDB structure *ABCD*, chain A). Use the PDBe (https://www.ebi.ac.uk/pdbe/) to find out more about the hit structures you have identified.