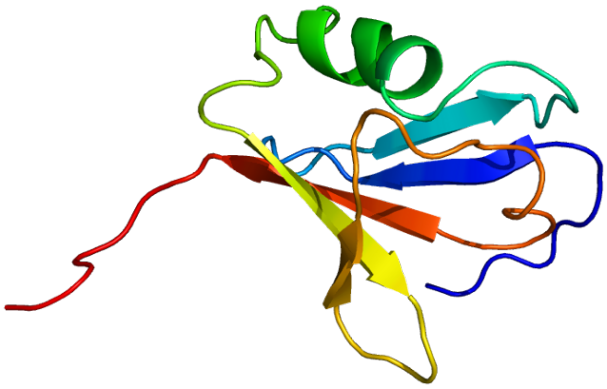
# Tutorial: Protein Modelling, Part 3

**All online quizzes are only activated during the tutorial. A copy of the questions and suggested answers to the quizzes will be uploaded to KEATS afterwards.**

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.

In the first part of this tutorial, we continue from Weeks 2&3 and use the **protein kinase** domain of **A-Raf** as our protein of interest. In the second part of the tutorial, you will be given the sequence of an unknown protein to model and investigate – this is an opportunity for you to revise the steps of homology modelling and the use of online protein bioinformatics tools on your own.



**Exercise 1: Model Assessment for Homology Models**

**Please follow the steps below and answer this quiz as you go along:**
**https://PollEv.com/surveys/75XSEQ0EZvA5IrmJsivmo/respond**

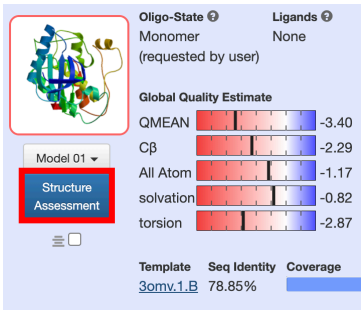
Last time, we generated models of the ARAF kinase domain using both the automated and alignment modes of the SWISS-MODEL server. In this exercise, we will use web-based tools and PyMol to examine the quality of the homology models you have generated, and to compare the outputs of the automated and alignment mode pipelines.

***You will need the homology models that you generated last week from the two pipelines in PDB format, as well as the PDB files for the structures used as templates. If you do not have these, please try the following:***

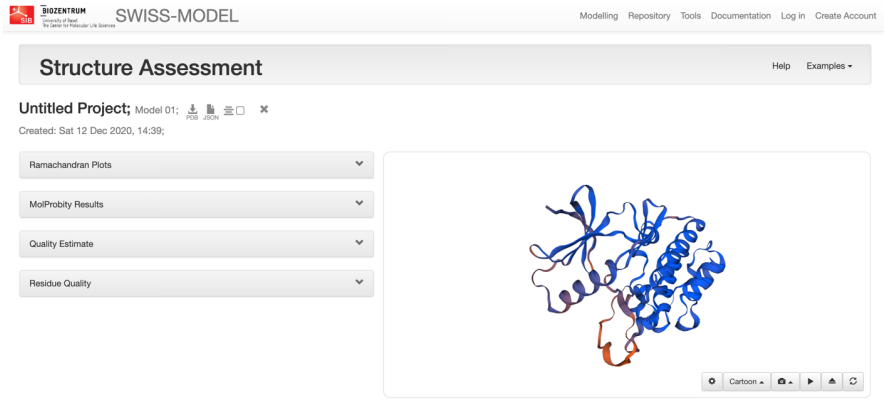
- Go to the submission page of SWISS-MODEL. If your browser cookies have not been cleared, you might see a section named "*Modelling Projects in Session*", which lists your previous projects. You can click into one of the projects and navigate to the 'Models' tab, which should show the Model Results page. Start from step (1) below. If your projects have expired, we have backup results available:
  **Automated Mode:** https://swissmodel.expasy.org/interactive/TSpRVX/models/
  **Alignment Mode:** https://swissmodel.expasy.org/interactive/GfsYem/models/

- As a reminder, we have pre-computed materials for the steps covered in previous tutorials: https://github.com/Fraternalilab/5BBB0226_Protein_Modelling_Tutorial

1. Go to the Model Results pages for the models generated in the previous exercise for the automated and alignment mode pipelines on the SWISS-MODEL web server.

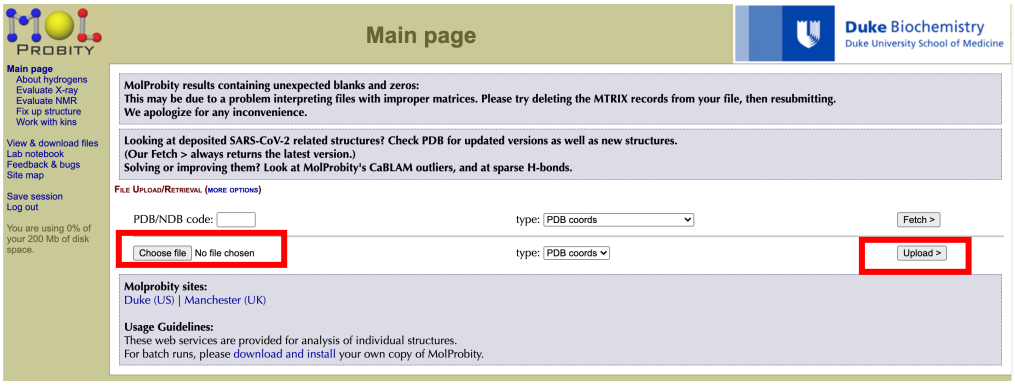



2. Select a model and click on "Structure Assessment".

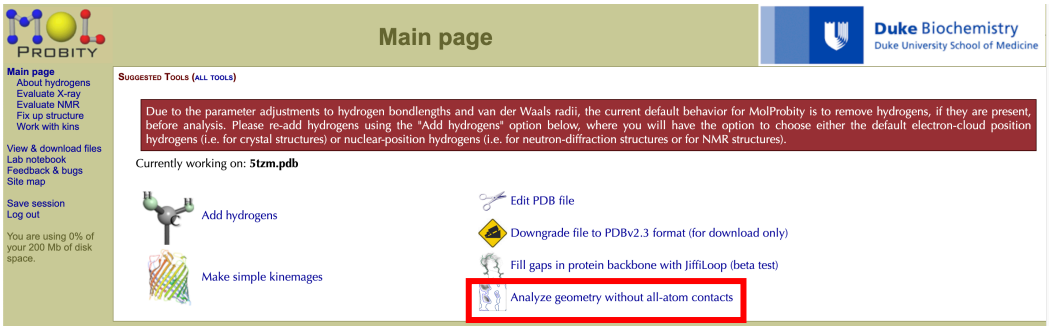



3. You should get an interactive page like the one shown above. SWISS-MODEL will call **MolProbity** to compute a score and assessments of the model. You can view multiple Ramachandran plots with different residue types, map the poorly modelled residues identified by MolProbity to the structure on the right, and also visualise individual residues of interest identified by using the Residue Quality tab. Play around a bit with the different features here.

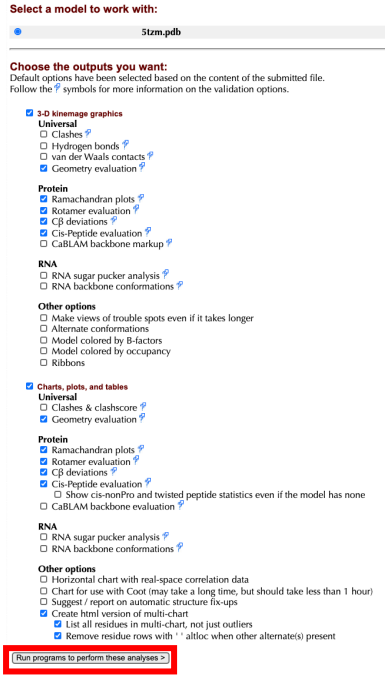
4. Step (3) provides interactive access to quality assessment of your models. Alternatively, you could use the MolProbity webserver itself, which yields identical results and is also more flexible (accepts queries of *any* PDB file as input). To download the assessments made here, we will go to the source. Navigate to the MolProbity web server (http://molprobity.biochem.duke.edu/).



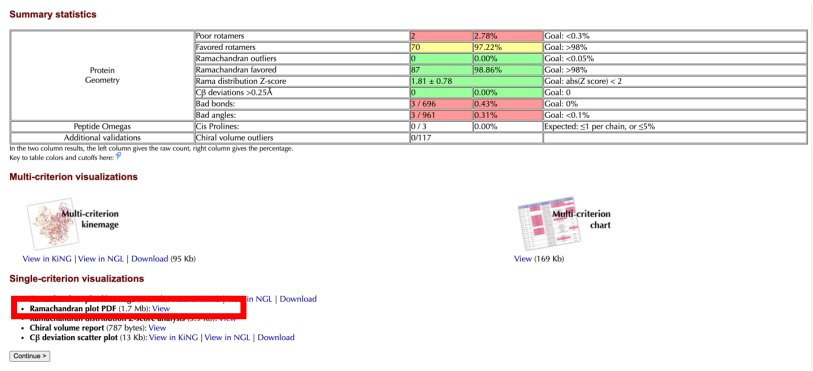

5. You should see a page similar to the one above. Upload one of the models you previously downloaded from SWISS-MODEL using the "Choose file" option, then click "Upload". If necessary, click "Continue" once your file has finished uploading.




6. Once your model has been uploaded successfully, click on "Analyse geometry without all-atom contacts" to begin the assessment protocol.

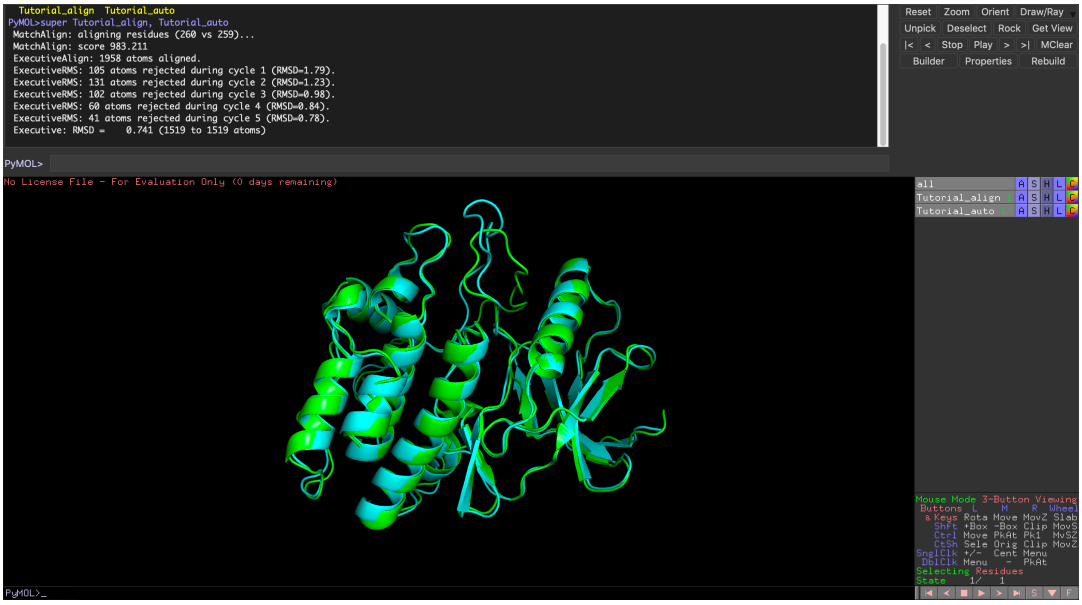



7. You should see a list of options for specifying the desired outputs. Leave the options as their defaults and click "Run programs to perform these analyses".




8. Take a look at the summary statistics and the defined thresholds; download the Ramachandran plot as a PDF.

9. Open the PDB files for the automated and alignment mode models generated by SWISS-MODEL in PyMol.




10. When both models have been loaded in the same session, superimpose the two models and compare the structures of each.

    *N.B. Use one of the following commands to superimpose:*
    `super [obj1], [obj2]`
    `align [obj1], [obj2]`

    Substitute [obj1] and [obj2] with the names of the objects you would like to superimpose. Your tutor will demonstrate superposition of structures on PyMOL in the tutorial.

    Note the Root-mean-squared deviation [RMSD] values on the top-left window.

    Alternatively, you could also compare your models with the templates you have chosen (for either the automated/alignment mode), or those you have omitted.

**Exercise 2. Analysis of Structure and Function of a given Protein Sequence**

**Please follow the steps below and answer this quiz as you go along:**
**https://PollEv.com/surveys/1L1mW9TAV3okU5m0xhYuV/respond**

In this exercise, you will work independently to analyse the properties of the structure and function derived from a given protein sequence. This is practice for the essay component of this module – you will be asked to describe the steps and results of homology modelling, as well as other bioinformatics investigations surrounding your assigned protein.

Below, we detail the steps you need to take to complete the exercise. In blue are questions for you to think about as you progress through the exercise; these may be helpful when answering the quiz or writing your essay. In addition, in *italics* are some hints both to help you to navigate the web resources and to make you aware of points to be considering when presenting your analysis.

**Save all results you have acquired along the way, including BLAST results, alignments, homology models, and any evaluation plots/statistics SWISS-MODEL returns for your model.**

You will need these to evaluate your results towards the end of this exercise.

1. You are given a FASTA sequence:

```
>Unknown
VTLQKRIGTGSFGTVFKGKWHGDVAIKILKVKEPTPEQLQAFKNEMQVLRKTRHVNILLF
MGFMTKPNFAIITQWCEGSSLYRHLHVTETKFDTMRRIDVARQTAQGMDYLHAKNIIHRD
LKSNNIFLHEGWTVKIGDFGLATVKYRWSGSQQVEQPSGSILWMAPEVIRMQDSNPYTFQ
SDVYGYGVVLFELMSGTLPYSNINNRDQILFMVGRGYLSPDLGKLCSTSPKSMKRLIIDC
LKFKREERPLFPQILVAIEQVQELM
```

   Retrieve the corresponding UniProt identifier. Identification of the protein name and organism. This is your query sequence.

> Q. Does your assigned sequence encode only part of a protein? If yes, which part of the UniProt entry matches with the given sequence?

*Hint: Check the format of the FASTA sequence.*

> Q. Study information from the UniProt entry on the function of your assigned protein.

*Hint: These features are typically covered in a UniProt entry:*
- *Function/Pathways*
- *Protein Domains*
- *Associated diseases (if any)*
- *Interactions (if any)*

2. Perform automated homology modelling via the web server SWISS-MODEL.

> Q. Study the table of templates SWISS-MODEL returns. Are there promising templates to be taken further to perform modelling of your target? Why? (Hint: Select at least 1 template here for automated homology modelling.)

*Hint: These are typical points to consider when evaluating the suitability of a template:*

- *Sequence identity*
- *Coverage*
- *Type of structure (crystal structures / NMR / Cryo-EM)*
- *Resolution*
- *What proteins are these templates?*
- *Oligomerisation state*

*Hint: It is bad practice to refer to your chosen templates only by quoting the full name (instead of using PDB ID, etc.)*

3. Examine the model(s) created by the server. Download the PDB file(s) for the template(s) used by the server, as well as the PDB file(s) of the model(s) created by the server.

4. Perform a BLAST search with the target sequence for structure modelling (using PDB database of protein structures). Retrieve sequences of suitable templates.

5. From your BLAST search, select the best template(s) for homology modelling.

*Hint: Select at least 1 template from the BLAST table.*

Q. Think about the criteria to select templates for modelling. Which template(s) would you select? Why?

Q. Compare the template(s) hits returned from SWISS-MODEL in step (2), and the one(s) you would have selected based on the BLAST search. Are they the same templates? Why (or why not)?

6. Perform T-Coffee alignments of your query sequence and the selected templates.

Q. Are there any notable differences between T-coffee alignments and SWISS-MODEL alignments?

*Hint: You can compare alignments with multiple criteria, including:*

- *Quantitative (scores from alignments)*
- *Qualitative (notable, observable differences in the alignments)*

7. Perform alignment mode homology modelling on SWISS-MODEL.

8. Examine your models, from both alignment-mode and automated-mode homology modelling using molecular visualisation software (e.g. PyMol). Perform a comparison (structural, 3D superimposition) with the corresponding templates, the model deposited in the ModBase database (if any), and/or others if you find other deposited models.

Q. Generate informative snapshots of your structural analysis. How do these results relate to other analyses you have performed so far, e.g. the alignments and the BLAST/SWISS-MODEL template hits?

*Hint: Visualise structures using PyMol, VMD etc. to compare these models (e.g. superposition of templates vs models, between models)*

Q. Study the model(s) from the metrics provided by SWISS-MODEL.

*Hint: A brief explanation of SWISS-MODEL metrics:*

- *GMQE – a measure of reliability. Expected accuracy of a model built with that alignment and template and the coverage of the target.*
- *QMEAN - estimator based on different geometrical properties and provides both global (i.e. for the entire structure) and local (i.e. per residue) absolute quality estimates on the basis of one single model.*
- *QMEAN Z-Score – comparison with existing, experimentally determined structures for proteins of a similar size.*

9. Evaluate the model(s) and template structures with MolProbity.

Q. Study the results from MolProbity. How do these results relate to the snapshots you have generated above, as well as your results on the template searches? Which model has a higher quality?

*Hint: Look at Ramachandran plots and summaries of geometrical assessment from MolProbity (number of outliers, bad bonds, bad angles etc.)*

10. Search for the functions of your assigned protein.

Q. Using public repositories and servers, comment on possible functions for the gene/protein corresponding to the selected sequence.

*Hint: Use resources like:*

- *Pfam*
- *InterPro*
- *UniProt*
- *Consult published work (primary research, review)!*

Final Note: Some analyses were covered in previous tutorials and not in this revision exercise (e.g. evolution of domain architecture in TreeFam). TreeFam entries are not always available for every UniProt protein.

Q. Why is that?

You will have learnt about ways to study evolutionary relationship of sequences in other lectures of this course (e.g. constructing phylogenetic trees from BLAST hits) – these could be alternative ways you could study the evolution of your sequence.