SSBI summer term 2023

Huiting Xu, Desirée Renschler-Sperl

2023

Assignment 05

1 Task 1 – Homology Modelling

SWISS-MODEL is an online-tool that uses four steps to do homology modelling of proteins. First, it searches a structural template for the protein to model by searching its own database called The SWISS-MODEL Template Library (SMTL). SMTL is based on Protein Data Bank (pdb) entries and the queries are done with BLAST or HHblits. It is also possible to query the AlphaFoldDB for templates. These results are ranked according to their sequence similarity and coverage. Second, pair-wise alignments between the template structure(s) and the target sequence are calculated. Depending on the template structure, the deletions and insertions in the alignment are improved by a heuristic step. For the following modelling step, the user can specify which of the templates to use or run the automated mode. The modelling process is based on ProMod3 and OpenStructure. A 3D model is generated for each template by using the atom coordinates of the template and referring to the associated alignment for correct positions in the model. Regions of insertions and deletions cannot be taken from the alignment but need to be modelled with constraint space programming or possible structures can be chosen from a loop library. The amino acid side-chains are then modelled starting with the conserved residues and using a rotamer library for possible conformations of the non-conserved side-chains. An energy minimisation is done using OpenMM Library and CHARMM22/CMAP force fields. As a last step, the model quality is evaluated by GMQE (Global Model Quality Estimate) which shows the expected quality of the model. Values range from 0 to 1, the higher the better the model. GMGE can also be calculated after the alignment step and is then used to chose the templates for modelling. The QMEAN score reports statistical potentials of mean force. QMEANDisCo uses the distances between atoms in the model and compares it with data from experimentally determined structures. The QMEAN score can be transformed to a Z-score for comparison with experimentally derived structures [6, 5, 1].

2 Task 2 – Evaluation of a Homology Model

To perform homology modelling with SWISS-MODEL, the structure of 6PP4 (referred to as protein A) is downloaded from the Protein Data Bank in fasta format (rcsb_pdb_6PP4.fasta, attached). This sequence is then uploaded to SWISS-MODEL and the search template option is chosen for analysis (https://swissmodel.expasy.org/interactive/bRQff5/, link for template search). The top template (SMTL ID : 5uo8.2.A, referred to as protein Z) is chosen to build the model. The model is downloaded as an archive. To compare the model to the X-ray structures of the proteins A and Z, the respective pdb files are used (model.pdb, 6PP4.pdb, 5uo8.2.pdb, attached).

2.1 a) Differences in secondary structures

The modeled structure of protein A mainly shows *alpha* helices and some *beta* sheets 1. The alignment for protein A and the template 5uo8.2.A (downloaded from SWISS-MODEL, alignment_model_a.png, alignment_model_b.png, attached) also shows that. The helices found in chain A are Leu44-Gln47, Ala79-Ser96, Ala105-Thr120, Glu127-Asn140, Arg147-Gln149, Ala164-Gly182, Pro225-Gln236, Pro267-Leu269, Glu281-Leu286, Ser319-Cys328, Leu336-Cys342, Thr350-Ser352, Trp354-Leu372, His380-Arg398, Trp405-Ile408, Gly414-Leu416. In chain B the helices are Leu44-Gln47, Leu69-Arg71, Pro80-Ser97, Gln104-Thr120, Glu127-Arg139, Arg147-Gln149, Ala164-Gly182, Pro225-Gln236, Pro267-Leu269, Glu281-Leu286, Ser319-Cys328, Leu336-Cys342, Thr350-Leu372, His380-Arg398, Trp405-Ile408, Gly414-Leu416. The model also shows ligands (figure 1).

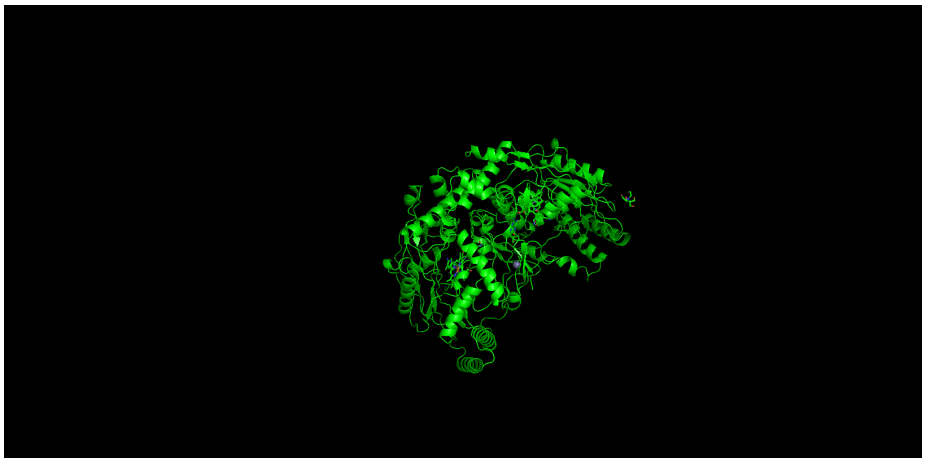

Figure 1: PyMol-view of model of protein A as predicted with SWISS-MODEL.

For superposition of the model with its X-ray structure (both protein A) the command super model_,6pp4 is used. The resulting view shows that the pre-

dicted structure and the X-ray structure are very similar. The predicted model is displayed in orange, the X-ray structure is shown in light-blue. The model shows a helix from positions Leu69-Pro73 (chain B) which is not found in the X-ray structure. The X-ray structure is a homo-2-mer, but the model is not. The template used for modelling is a monomer. That is why the structure of the model is only superpositioned on one of the dimers (figure 2).

Figure 2: PyMol-view of the model of protein A (orange) superimposed with 6PP4 (light-blue).

Next, the model (orange) is superpositioned with the template X-ray structure (protein Z, pink). Also those two structures are very similar. This results from the fact that their sequences are identical which allows good homology modelling. The same helix from position Leu69-Pro73 is only found in the model but not in the template structure, as well (figure 3).

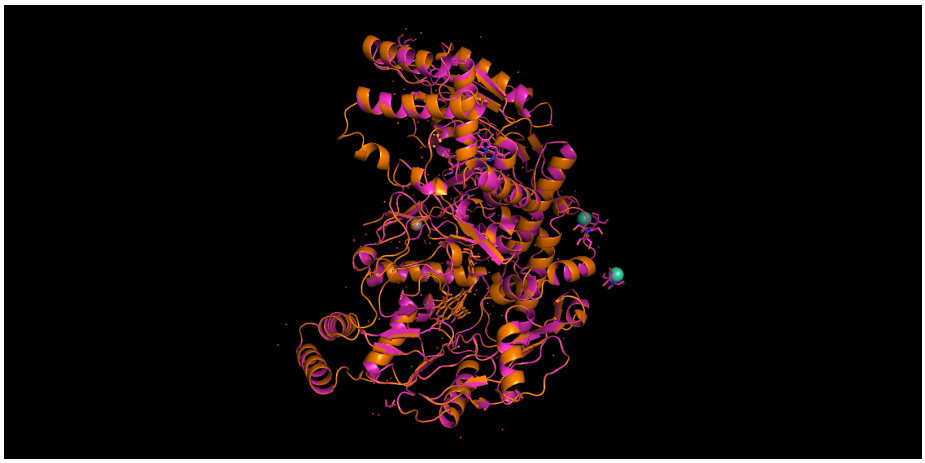

Figure 3: PyMol-view of the model of protein A (orange) superimposed with the template protein Z used for homology modelling with SWISS-MODEL (pink).

When the X-ray structures for the template and 6PP4 are superpositioned they show a very high similarity as well. So the template used is well suited for prediction of the protein model (figure 4).

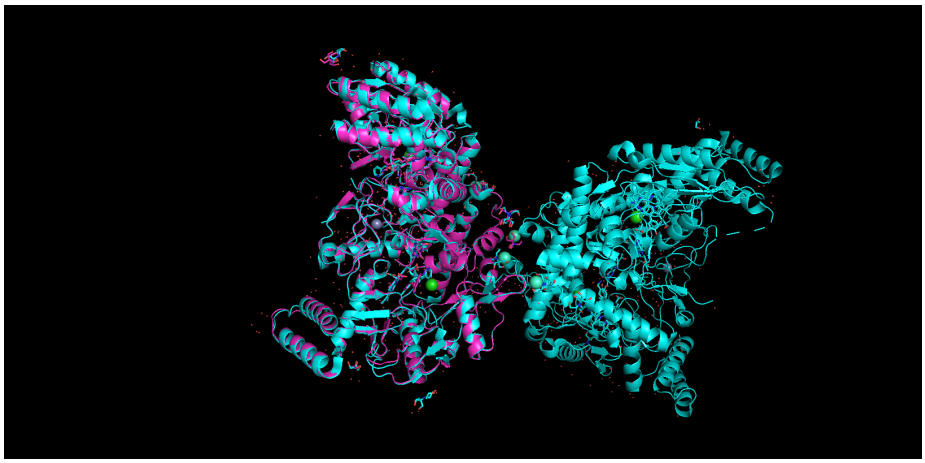

Figure 4: PyMol-view of the Xray-structures of the template (pink) and 6PP4 (light-blue) superimposed.

2.2 b) Molecular weight

The theoretical weight of the protein A according to its sequences is 49294.02 Da calculated by the addition of average isotopic masses of amino acids in the protein and the average isotopic mass of one water molecule [2, 3, 4].
The published weight is 204.9 kDa according to the pdb entry. This weight is for the homo-2-mer and also includes the ligands and not only the protein. The difference between bought weights is though big.
Even between two proteins of the same sequence the theoretical and the experimental weight can be different. For the theoretical weight calculation, average weight values for each amino acids are used. These can differ from experimental values. One should also consider the fact that the theoretical values take the weight of only one water molecule into account. When determining the weight experimentally the protein dry mass or protein with water can be used. This will change the molecular weight. In addition, often the weight of proteins together with ligands is determined for stability reasons. This should also be remembered when comparing molecular masses.

2.3 c) Hydrophobic regions

To color the amino acids according to their hydrophobicity the color_h script is used. To show the surface the command as surface, model_ is used (figure 5). The more hydrophobic the amino acid the darker the red color.

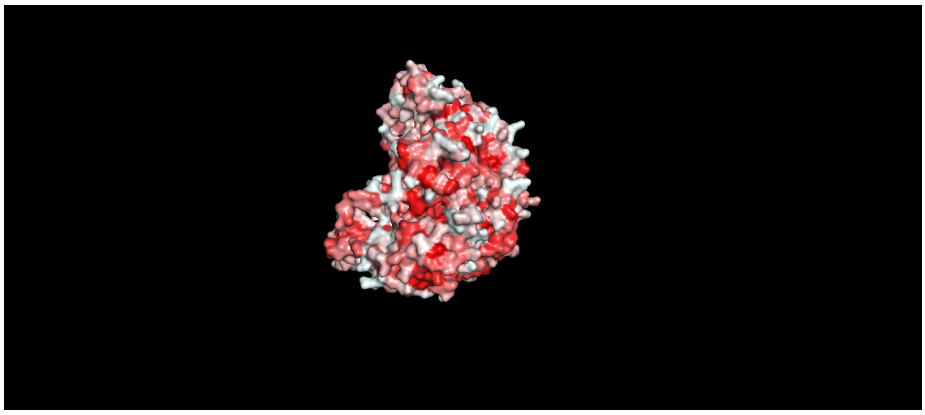

Figure 5: Surface of protein A colored by hydrophobicity.

The most hydrophobic amino acids and their hydrophobicity scale are Ile: 1.380, Leu: 1.060, Phe: 1.190 and Val: 1.080. The amino acids valine 64 and phenylalanine 65 are highlighted as an example (figure 6).

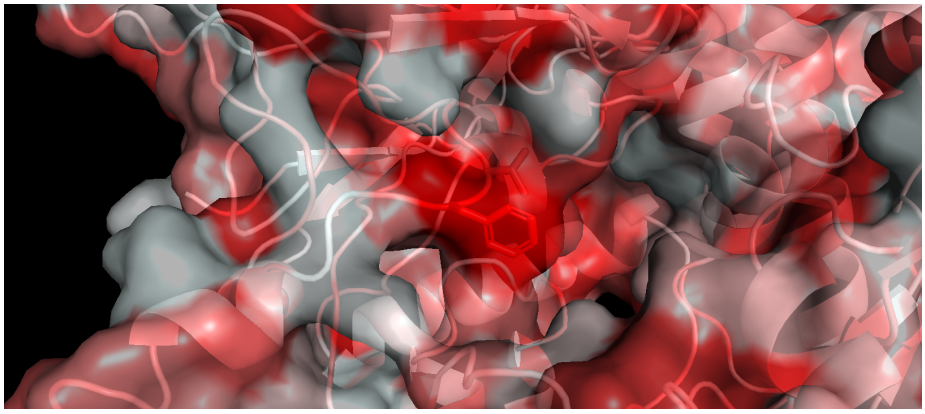

Figure 6: Hydrophobic amino acids valine 64 and phenylalanine 65 side chains highlighted as sticks (protein A).

3 Task 3 – RMSD

To calculate the root mean square deviation (RMSD) between two pdb-files, alignments of the pdb structures were made with PyMol for the model (protein A) compared to 6PP4 and the template (protein Z, 5uo8.2) and the model.

PyMOL align model_, 6pp4, object=model6pp4
Match: read scoring matrix.
Match: assigning 835 x 2189 pairwise scores.
MatchAlign: aligning residues (835 vs 2189)...
MatchAlign: score 4261.000
ExecutiveAlign: 6489 atoms aligned.
ExecutiveRMS: 103 atoms rejected during cycle 1 (RMSD=3.18).
ExecutiveRMS: 366 atoms rejected during cycle 2 (RMSD=0.91).
ExecutiveRMS: 314 atoms rejected during cycle 3 (RMSD=0.48).
ExecutiveRMS: 281 atoms rejected during cycle 4 (RMSD=0.38).
ExecutiveRMS: 185 atoms rejected during cycle 5 (RMSD=0.34).
Executive: RMSD = 0.315 (5240 to 5240 atoms)

According to PyMol the RMSD for the model and 6PP4 is 0.315 Å.

PyMOL align model_, 5uo8.2, object=model5uo8.2
Match: read scoring matrix.
Match: assigning 835 x 982 pairwise scores.
MatchAlign: aligning residues (835 vs 982)...
MatchAlign: score 4332.500
ExecutiveAlign: 6591 atoms aligned.
ExecutiveRMS: 14 atoms rejected during cycle 1 (RMSD=2.84).
ExecutiveRMS: 282 atoms rejected during cycle 2 (RMSD=0.14).
ExecutiveRMS: 276 atoms rejected during cycle 3 (RMSD=0.11).
ExecutiveRMS: 127 atoms rejected during cycle 4 (RMSD=0.10).
ExecutiveRMS: 55 atoms rejected during cycle 5 (RMSD=0.09).
Executive: RMSD = 0.092 (5837 to 5837 atoms)

According to PyMol the RMSD for the template and the model is 0.092 Å.

To calculate the RMSD, the structures of two proteins are superimposed. The averaged distance between all atoms of the aligned proteins is calculated. In our code $Renschler - Sperl\_Xu\_Assignment05$ the RMSD is calculated with the formula

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2} \tag{1}$$

In the script $Renschler-Sperl\_Xu\_Assignment05$, we used the package PyMol. First two pdb files are loaded, then *cmd.align* is used to align these two pdb

files and save them as *align* (figures 7, 8). Then the coordinates of all atoms in *align* of the two structures are chosen, respectively. In the end, the RMSD is calculated with the formula given. Between our model and 6pp4.pdb, we got a RMSD value of 0.31504671771958886 Å. Between model.pdb and 5uo8.2.pdb, we got a RMSD value of 0.09181821717876336 Å.

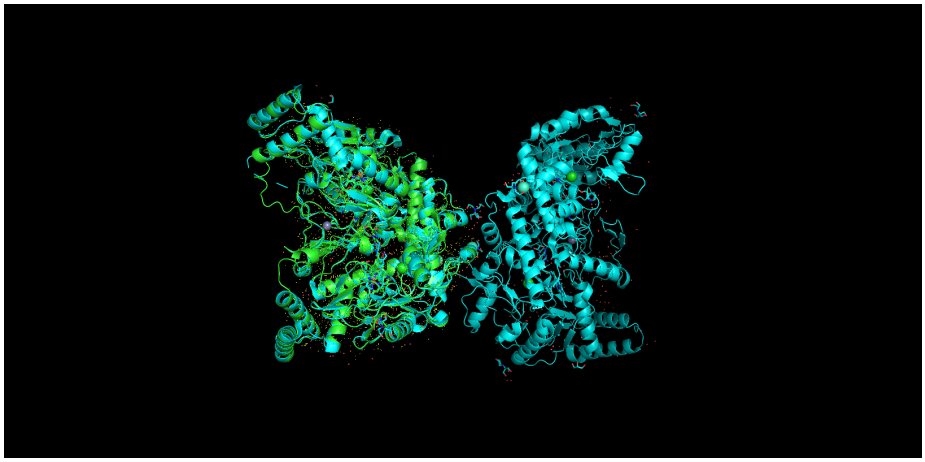

Figure 7: PyMol-view of the model (green) and 6PP4 (light-blue) aligned.

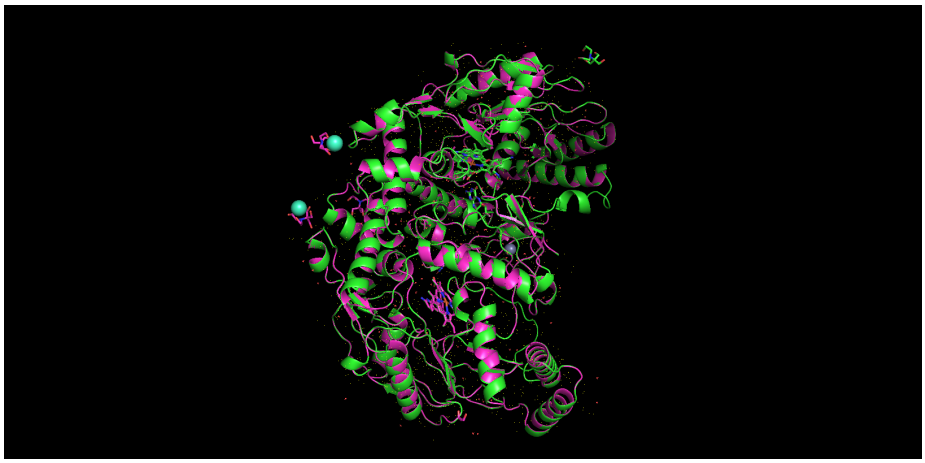

Figure 8: PyMol-view of the model (green) and 5uo8.2 (pink) aligned.

References

[1] https://swissmodel.expasy.org/docs/help. Accessed: 05-29-2023.

[2] https://web.expasy.org/compute$_p i$/. *Accessed* : $29 - 05 - 2023$.

[3] https://web.expasy.org/compute$_p i/pi_t ool - doc.html$. *Accessed* : $29 - 05 - 2023$.

[4] https://web.expasy.org/findmod/findmod$_m asses.htmlAA$. *Accessed* : $29 - 05 - 2023$.

[5] Torsten Schwede, Jurgen Kopp, Nicolas Guex, and Manuel C Peitsch. Swiss-model: an automated protein homology-modeling server. *Nucleic acids research*, 31(13):3381–3385, 2003.

[6] Andrew Waterhouse, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny, Florian T Heer, Tjaart A P de Beer, Christine Rempfer, Lorenza Bordoli, et al. Swiss-model: homology modelling of protein structures and complexes. *Nucleic acids research*, 46(W1):W296–W303, 2018.