Prof. Dr. Oliver Kohlbacher
Prof. Dr. Manfred Claassen
Anke King, Rosanna Krebs,
Dr. Kyowon Jeong, Hadeer Elhabashy, Dr. Samuel Wein
Faculty of Science
Department of Computer Science and
Institute for Bioinformatics and Medical Informatics (IBMI)

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

5th Assignment - Structure and Systems Bioinformatics

Hand in: 2023-06-08 20:00 CEST (source code and pdf in a single archive, uploaded via ILIAS)

Task 1 – Homology Modelling (10 P)

The first step in homology modelling is to find a protein of known structure that shows the highest sequence similarity to the target protein. Its experimentally determined structure is then used as a template to model the structure of the target protein by computational methods.

One automatic protein model generator is the program SWISS-MODEL (http://swissmodel.expasy.org/). Shortly describe the underlying method in your own words (max. 0.5 pages).

Task 2 – Evaluation of a Homology Model (50 P)

In this task you evaluate the obtained homology model of a protein A and compare it to the structure determined by X-ray crystallography. Furthermore, you compare the obtained homology model of protein A to the template model (protein Z) and the template model to the X-ray structure of the target protein A.

1. Find the protein structure of 6PP4 in the PDB database (http://www.rcsb.org/). This will be your protein A. Download the fasta-file. It is used to generate a homology model.
2. Use SWISS-MODEL (http://swissmodel.expasy.org/) to build the templates for your protein A from the fasta-file. Choose the top template model and create the homology model of Protein A for it.
3. Now analyze the results by having a detailed look at the three structures (your homology model of the target protein A, the X-ray structure of protein A and the X-ray structure of the template protein Z). One common way to analyse differences in structures are to use superpositions (PyMOL can do that by aligning the structures, you can use the 30 day trial version without a license). Answer the following questions:
   a) Explain the differences in the secondary structures you may observe. Determine the kind and location of secondary structure of your homology model (which parts form α-helices, which β-strands?) and compare it with both X-ray models. Is there any deviation between them (discuss!)?
   b) Determine the theoretical molecular weight of your protein A based on its amino acid composition. Find an published determined molecular weight of protein A to compare it with the theoretical weight. Comment on and discuss how well both numbers agree, especially if there are larger deviations! How exact are the

theoretical and experimental mass determinations? Is it possible that two proteins with identical sequences have different masses? Explain your answer.

c) Visualize the hydrophobic regions of protein A (please include an image of your visualization) (https://pymolwiki.org/index.php/Color_h) and state which amino acids are found in the hydrophobic parts of the surface.

Task 3 – RMSD (40 P)

Another common way to analyse differences in structures is to compute the RMSD.

Write a Python3 program to calculate the root mean square deviation (RMSD) between two pdb-files. What does the RMSD describe (explain, do not (only) give the formula)? What is the RMSD between your model and both X-ray structures (units in Ångström)?

- The program input should be a list of two PDB files. Please use a command line parser for providing input files.
- You can use Python libraries to parse the PDBs and get access to the structure elements (e.g., residues, residue atoms, atom coordinates).
- **However, you have to implement the computation of the required RMSD yourself.** You are not allowed to use external libraries/methods for this step.
- In your hand-in, also provide the code of your program. Upon running for provided input PDBs, it has to print the RMSD value to the console.

Think about the necessary pre-processing steps before calculating the RMSD. E.g. how would a sequence insertion or a structure rotation influence the RMSD calculation of a protein to its original?

For hand-in, zip all three pdb-files, your RMSD-script and your answers (PDF). Questions can be directed to ssbi-ss23@informatik.uni-tuebingen.de or the ILIAS course forum. We highly encourage you to use ILIAS for communication.