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

3rd Assignment - Structure and Systems Bioinformatics

Hand in: 2023-05-18 10:00 CEST (source code and pdf in a single archive, uploaded via ILIAS)

Task 1 Protein Data Bank II (**8** P)

Look up the proteins with the PDB identifier 1IGT and 5IRE. For both of the proteins, answer the following questions.

1. What experimental method was used to determine the structure? (**2** P)
2. Explain why the structure of 5IRE does not contain hydrogen atoms while the structure of 1IGT does. (**2** P)
3. The protein of 1IGT reports the so-called R value in the experimental section. Please provide a short description of the R value. How would you conclusively judge the quality of this model? (**2** P)
4. The structure of 1IGT does not only contain protein residues, other molecules can be found in this file as well. Describe, in 2-3 sentences, what these molecules are. What type of interaction do these molecules have with the protein? (**2** P)

Task 2 Investigating secondary structure elements (**70 P**)

In this task, we will have a closer look at various properties of different secondary structures. You will find a random selection of 200 PDB files in the supplementary material. You will need to write a Python program to parse them and extract the information needed to answer the following questions. Please hand in your answers together with the source code and make sure to include information on how to execute the code.

Note: There might be multiple models of the same structure in one file. Do not forget to address this for computing frequencies.

a) How abundant are sheets, right-handed α-helices, and right-handed 3_{10}-helices in these proteins? In order to find that out, compute the aggregate length of each of the above secondary structures and divide them by the aggregate length of all chains in the provided PDB files. Refer to page 157 of the Protein Data Bank Contents Guide (ftp://ftp.wwpdb.org/pub/pdb/doc/format_descriptions/Format_v33_A4.pdf) to find out how different types of helices are encoded in PDB files. (**20 P**)

Secondary structure	Abundance
⋮	⋮

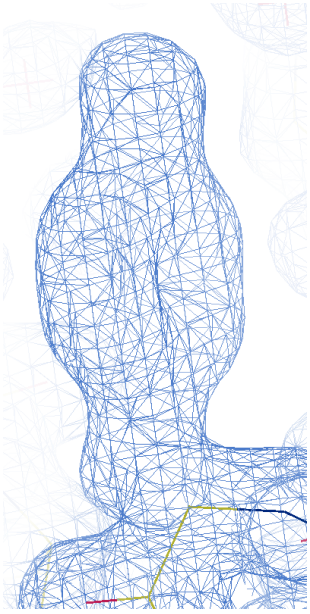
b) Considering only right-handed α-helices and sheets, what are the propensities of the different amino acids to form these structures? To answer this question, provide a list of amino acids ordered by the frequency of their occurrence in the respective structure for both types of secondary structures. Please include the propensity of each amino acid in the table. (**20 P**)

Right-handed α-helices		Sheets	
Amino acid	Propensity	Amino acid	Propensity
⋮	⋮	⋮	⋮

c) For all proteins, compute the distances between backbone N atoms and the O atom of backbone $C = O$ groups four residues earlier. Plot all these distances in a histogram. Next, do the same for right-handed α-helices only, and plot them in a second histogram. What do you observe? (**30 P**)

Task 3 Electron Density Map (**12 P**)

- Which amino acid(s) match(es) the electron density map below? Explain your answer and try to mark the heavy atoms in the electron density map (in this very figure).
- The image below shows the backbone as a stick model. What are potential problems that arise if the protein structure is inferred from an electron density map? **Please keep your answer around 2-3 sentences.**



Task 4 Force Fields (**10 P**)

In the lecture, you learn about the concepts of Force Fields to describe energetic contributions from different kinds of interactions. In particular, you learned about the AMBER family of force fields. **In the following, provide short answers of around 2-3 sentences each.**

- Please find another common Force Field. Provide its name, main citation, and highlight the main differences of this Force Field when compared to AMBER.
- Assuming you need to use a Force Field in your work, what reasoning would you use to decide for or against your found Force Field vs. AMBER?

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.