EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

BIO-4371
Structure-Based Drug Design
Summer Semester 2024

Institute for Bioinformatics
and Medical Informatics

SBDD Project

Submission deadline: July 12, 2024, 6 p.m.

The project has to be performed in **groups of size three**. Currently it seems that we need one group of size two.

You need to hand-in a **report in PDF format** with a maximal length of 4 pages in total. It has to follow the structure of a scientific article: introduction, materials and methods, results, and discussion. The focus should be on brief but concise descriptions of the methods (how you did what) and the results. Up to three meaningful figures are allowed. The introduction should give a brief overview over the project and the targets. The discussion should comment on the findings you made and to draw conclusions.

In the lecture on July 17 every group has to give an **in class presentation** of 6 minutes. The presentation with supporting slides shall give a summary of your work for your colleagues.

Grading: Report 10 pt, Presentation 10 pt, Content 20 pt.

It is time to make hands-on experience in the application of the arguably most important computer-aided SBDD method: **protein-ligand docking**. The aim of this project is to learn how to do protein-ligand docking. In this real world use-case you have the chance to assess the quality of a docking tool with respect to its ability 1) to predict the correct binding pose of a known ligand and 2) to perform virtual screening (VS) on selected targets.

We have prepared several **datasets** for this project, which are available as .zip files. The 4-character prefix of every filename is the PDB ID of the target structure. In the .zip files there are two files in SDF format: one with 40 known ligands and one with 1200 so-called decoys (= known non-binders).

The project can be split up into **five parts**: 1) setting up an environment for protein-ligand docking, 2) learning how to do protein-ligand docking, 3) assessing the pose-prediction quality of the docking tool, 4) performing VS using protein-ligand docking, and 5) assessing the VS quality. To learn these steps and to set up the required workflows for VS and its evaluation (scripting/automation is highly recommended) please work as a team and use the target *Neuraminidase* with the PDB ID `1a4g` (zip-file 1a4g_na.zip).

When you have set up your workflows and successfully finished the *Neuraminidase* target **every group member freely chooses one further target** (no duplicates) from the available datasets and repeats parts 4) and 5) individually (target and ligand preparations, VS, VS quality assessment). In the report, every target must be briefly described by the responsible person and the VS evaluation metrics must be included. Please indicate in the report, which group member worked out which target.

1) Setting up an environment for protein-ligand docking

In the following descriptions we name tools and give links that lead to a software environment centered around the docking tool `AutoDock Vina`. Besides the docking tool itself you usually require further software to do file format conversions, to add missing hydrogens or partial charges, etc. (Please note: you have the freedom to use any protein-ligand docking tool and required additional software that are available to you if you like). One helpful command-line tool is `obabel` and also PyMOL is a tool that can assist you with several tasks you need to do.

Please find helpful resources and links here:

- https://autodock-vina.readthedocs.io/en/latest/index.html
- https://open-babel.readthedocs.io/en/latest/Command-line_tools/babel.html
- https://github.com/forlilab/Meeko
- https://anaconda.org/bioconda/autodock-vina
- https://anaconda.org/conda-forge/openbabel
- https://anaconda.org/conda-forge/meeko

If required, we can give you access to our institute's computers where *Autodock Vina* works. Especially, if you encounter problems of running *Autodock Vina* on your computers please let us know as soon as possible and make use of this option!

2) Learning how to do protein-ligand docking

For *Autodock Vina* users we recommend the tutorial 'Basic docking' from the documentation I've referred to. There are alternative ways to do receptor and ligand preparation and one of them is described here. Despite using another docking tool in the end these preparation steps can also be used for docking with *Autodock Vina*. Finally, we recommend to use the Vina force field (in the AutoDock world force field is the *scoring function*).

During receptor preparation you may remove all solvent molecules. The ligand to be used for docking should be the experimentally resolved ligand that is bound to the target. Please carefully inspect the used structure (especially the ligands) to decide, which of the chains to use in the end.

3) Assessing the pose-prediction quality of the docking tool

If docking worked you want to assess the quality of the prediction result. As you performed a so-called 're-docking' of a known ligand with known binding pose the interesting question is: how close is the docking pose on the first rank to the experimentally observed binding pose (ligand in the crystal structure)? To answer this question you can use the root-mean-square deviation (RMSD) as a similarity metric.

A commonly used upper threshold to accept a predicted docking pose as correct is 2.0 Å. If the energetically most favourable candidate is not acceptable, is there one on a lower rank? What is the pose with best RMSD and what is its rank?

4) Learning how to do VS using protein-ligand docking

The second interesting application is to use protein-ligand docking for VS. Thus, docking of a library of compounds against the target in order to enrich hit candidates at the top of the list. To assess if our docking tool indeed can produce a good ranking, we do a retrospective benchmark study using a small set of known ligands and a large set of decoys (described above).

To do the VS, you prepare and dock both sets of compounds (ligands, decoys) and you merge the best prediction of every docked compound in a final result list for ranking. Hint: `obabel` can convert SDF to PDBQT.

5) Assessing the VS quality

To assess the VS quality of the docking tool you now analyze the ranked list from VS using statistical measures. In this project you should use ROC and AUC. Please familiarize yourself with both and use them to the assess the VS quality of your docking tool for your targets individually.

IMPORTANT Please keep in mind that docking of a single compound can easily take 2 minutes. Every dataset contains 1240 compounds. You can do the maths yourself. Thus: 1) begin with your project immediately and start the VSs as soon as possible. 2) Parallelization of docking is very simple (docking in batches). Consider to make use of that. 3) VS and report writing can also be parallelized quite easily.

▷ Please use Slack to discuss problems in the first place

▷ If you have confidential questions, don't hesitate to drop by or write me an e-mail