Lab Exercise 6: Docking - Biophysical Chemistry

April 16th, 2018

1 Introduction

The computational cost for sampling relevant conformations of most protein systems through for instance molecular dynamics (MD) is rather high. Larger pharmaceutical companies haven't until quite recently started to invest in the adequate resources for doing large-scale MD. One should also note that although MD is a general tool for investigating most of the relevant biophysical phenomena, there are other specialized methods far more efficient for certain analysis.

The side effects of a receptor modulating drug can for instance be limited by minimizing it's interaction with other molecules similar to the target receptor. Hence a common part of the toxicological analysis of new potential drugs is to investigate its binding affinity to a large set of relevant molecules. The number of receptor-ligand pairs that needs to be studied in this case easily become very large, making a computational approach an attractive option. However, even a general method like MD is likely too expensive for a large study. Instead it is common to use less precise but more specialized methods.

Docking is a popular method specialized in approximating interactions between a pair of molecules, usually a ligand and a receptor. Here, the calculation of physical dynamics is abandoned in favour of improving the computational efficiency. Today, there is a great variety of tools for doing this that are extensively used by the pharmaceutical industry as well as governmental institutions, predominantly for drug discovery and toxicological analysis.

Let's think of ligand binding as a "lock-and-key" problem (figure 1). We thus need to find out how likely a ligand is to find and interact with one or more sites on the receptor. The geometrical fit as well as the energetic stability of each site needs to be accounted for. Roughly speaking the geometrical fit is used to filter out bad or impossible confirmations, whereas the energetic stability is used for sorting all the feasible conformations by likelihood. A fit that maximizes the contact area also maximizes the possibility of interactions and hence the energetic stability. On the other hand a narrow binding site decreases the flexibility of the ligand and thus imposes a entropic penalty, countering the potential gain in energetic stability. As we have learned in previous assignments, accounting for entropic effects requires some sort of sampling of the local

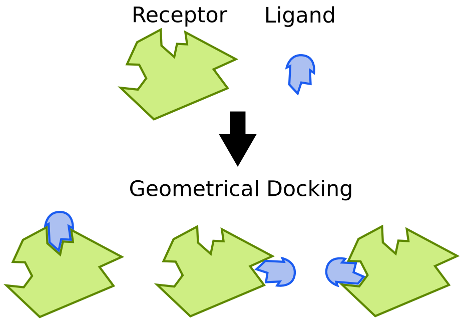

Figure 1: Geometrical docking of a ligand-receptor pair. Imagining the ligand as a key that goes into the lock on the receptor, the geometrical part of docking looks for a binding-site that maximizes the contact area without steric clashes.

conformations, similar to dynamical sampling done in MD. Due to the simplified model used in most docking methods this effect is hard to account for and thus mostly ignored.

To account for all of these factors and being able to sort amongst the possible conformations most docking tools use a scoring function. A decent geometrical fit and a low enthalpy scores higher and puts such binding site higher in the sorting. The user is then presented with a list of binding conformations ordered from most likely to least likely. To reduce the search space the ligand is usually allowed to have some flexibility, whereas the receptor is assumed to be in a rigid and active conformation.

Tasks

1. As describe above, the ligand is the only flexible part in docking. What is the draw back of this limitation?
2. To find potential binding sites an approximation of the relative free energy difference between the sites must be calculated. It is easy to imagine that the enthalpic contribution for a specific conformation can be calculated by counting interactions (e.g. hydrogen bonds), but can the entropic part also be estimated in docking? Describe in a few sentences how you would do it in if you were to write your own docking program, based on what you have learned in previous exercises.

2 Getting Started

AutoDock is a free docking tool that is fast and commonly used in research. We will be using a variation of this tool that is faster and have been show to give better results, which is called *AutoDock Vina*. We won't however discuss how the algorithm is implemented in any greater detail.

For completing this assignment you will need the structure-files that comes in a tar-ball named `structures.tar.gz`.

Compared to other biophysical computational tools, like MD, docking tools in general are not that complicated to use. The user is not required to understand much about the underlying algorithm and there are very few parameters that needs to be defined. To get an understanding of what we are looking at however, we will have to use a molecular visualization tool. There are a large set of alternatives, but in this exercise we will use *Pymol*.

3 First Docking

Before becoming drug development/discovery experts you will have to get acquainted with the tools using a test system and its ligand. For this we will use the crystal structure of a domain from the Abelson tyrosine kinase system, which is shown to be involved in cell differentiation and hence a target for cancer treatment drugs. The ligand we'll use in the first docking is an inhibitor and thus an example of such a drug. **Open up the files `intro/ABL_kinase.pdbqt` and `intro/ABL_ligand_ref.pdb` in Pymol and inspect the structures (note the different file formats).** You should see the system with the ligand at the position where it was found in a crystal structure complex, which we assume is the "correct" binding pocket.

To define the search region for docking, Vina needs to be provided a boundary box. We have written a script that can be used in Pymol for drawing a cube by specifying a center point and its dimensions. The script is called (`drawBox.py`). Go ahead and load the script into Pymol by running the following command in Pymol:

```
run /path/to/drawBox.py
```

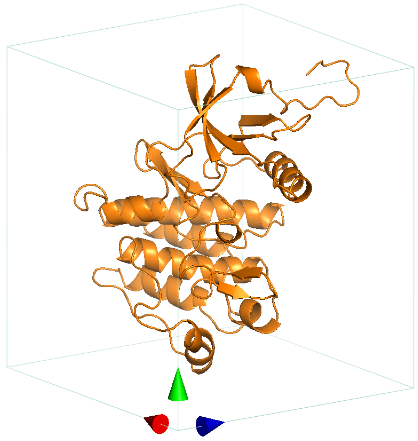

Figure 2: A box created around the protein using the `drawBox`-command in Pymol. The three arrows show the direction of the spacial coordinates; **red is for x, green is for y and blue is for z.**

Note that you'll have to specify the actual path to the script instead of ”`/path/to/`”. This command will run a regular python script that communicated with the Pymol platform and can be used to add extra functionality, which can be very useful in most projects requiring some specialized visualization. The script is very simple, so don't be afraid to take a look inside.

Now try creating a box with the new command by running:

```
drawBox(15,90,57,55,60,52)
```

This will call the function that we defined in the script file and create a cube with the the center point at $(x = 15, y = 90, z = 57)$ and the dimensions $(55, 60, 52)$, all units are in Å. Together with the box there are a three cones showing the direction of the spacial coordinates; red is for x, green is for y and blue is for z. See figure 2. **Make sure this makes sense to you before moving on, by playing around with the parameters.** The coordinates that define the box will also define the region within which the docking tool will preform the search. Note that the box you created surrounds the whole complex, which means that the docking will cover all exposed surfaces of the protein.

To start docking we will need the binary for AutoDock Vina. Go ahead and download the correct binary from `vina.scripps.edu` and ”untar” it in an easily accessible path.

Run the binary without any flags and check the output. Now open up `intro/ABL.conf` in a texteditor and fill in the coordinates from the box above replacing the ”???”. This file will be used as an input configuration file. Run it with Vina inside the `intro` directory via the command

```
path/to/vina --config ABL.conf
```

and as previously replace ”`path/to/`” with whatever path to the relevant file.

You will get an output structure file named `ABL_ligand_out.pdbqt` that contains the most likely position of the ligand together with a command line output with a list of the corresponding affinity energies. View the output structure file in Pymol together with the reference structure (`intro/ABL_ligand_ref.pdb`) to check how close you got. **Note that `intro/ABL_ligand.pdbqt` does not show the reference position of the ligand.** You can browse through all the structures with the arrows in the lower right part of the Pymol-window. The first structure under frame one is the one that the algorithm scored the highest.

Now check how your results differ if you give a smaller bonding box. In Pymol instead try

```
drawBox(11,95,57,22,24,28)
```

and update the corresponding numbers in the Vina configuration-file. Now the search volume is localized to a more probable region of the receptor. The algorithm will do a more thorough examination of the correct region and should therefore find it more easily. Rerun Vina!

Tasks

1. Include an image of the best fit of the ligand in the receptor binding pocket in your report.
2. Would you be able to limit the search region if you didn't know the exact binding site? Think for instance of a transmembrane protein.

4 Drug Design

Let's say that we have a molecule which is know to have a certain physiological effect. We now want to increase the effects of this substance by engineering a new similar substance with a stronger effect and sell it as a pharmaceutical drug. We don't know exactly how it works but we have some preliminary data that shows that there must be at least one amongst three receptors of the same family that it interact with, see figure 3. The idea is to try and find the receptor that it binds to and then try to use the structure of that receptor to come up with a new substance that binds even stronger to it.

4.1 Pin Down the Receptor

In the first part of this section we will try to pin down the target receptor. The structure file of the three receptors are named `part1/receptor_A.pdbqt`, `part1/receptor_B.pdbqt` and `part1/receptor_C.pdbqt` and the ligand `part1/unknown_ligand.pdbqt`. These receptors are transmembrane proteins, hence their outer surface won't be exposed. **Think about this when you create your bounding box.**

If you want to do a finer search (with increased computational time) you can set the `exhaustiveness`-option in the configuration-file to a higher value.

Task

Try to figure out which of the receptors the ligand likely belongs to and include an image of the bound ligand in the binding site.

4.2 Improve Binding Affinity

Now that we know the likely receptor, let's try to find out which of the three candidate drugs that has the highest affinity to the receptor (figure 4). The structure files for the three ligands are `part2/Lig045.pdbqt`, `part2/Lig412.pdbqt` and `part2/Lig973.pdbqt`. Remember to reduce the search volume a bit to include the binding pocket, but also a bit of the surrounding regions if the modified drug binds to a slightly different region.

Task

Write down which drug has the largest affinity to the receptor binding site and its affinity energy.

4.3 Which Ligand & Receptor is it

Task

Which receptor-ligand-pair are we looking at? If you do biology, you might recognize the receptors as part of a famous membrane protein family. If not, you should be able to find it by doing a ligand search on the many freely available databases out there.

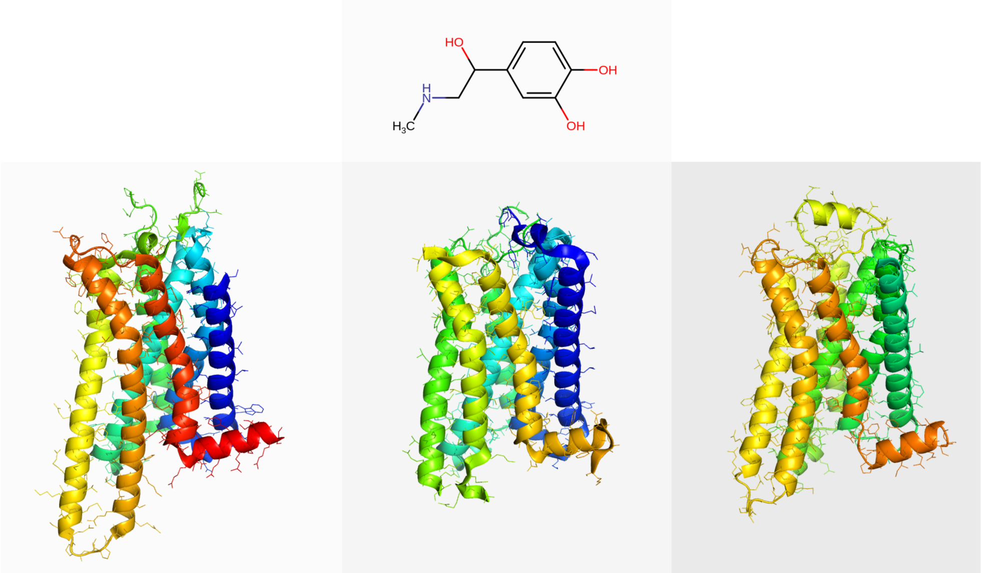

Figure 3: The unknown ligand (top) and the three receptor proteins (bottom).

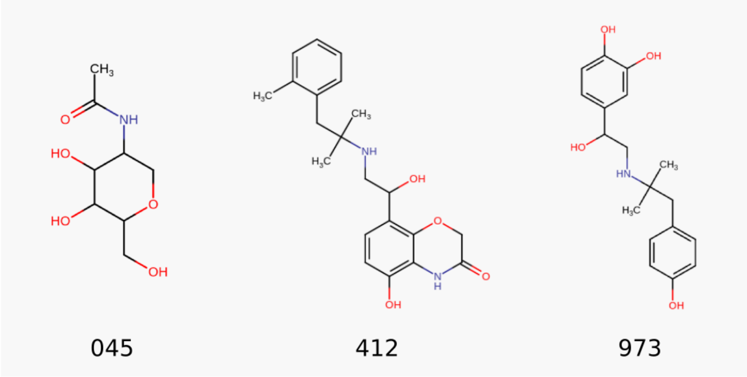

Figure 4: The three candidate drugs and their corresponding serial number.