**Virtual screening tutorial**


This tutorial is inspired by the article of Chris de Graaf and co-workers [1]. Histamine is an important biogenic amine (decarboxylation product of histidine) involved in allergy. Antagonists and inverse agonists of the histamine receptor can be used as antihistaminics to treat allergy. Relatively recently solved 3D structure of this receptor [2] makes it possible to discover new antihistaminics by structure-based drug design. In the article of Chris de Graaf and co-workers, authors virtually screened 108 790 compounds. They selected, purchased and tested 26 of them, 19 turned out to be active with $K_D < 10$ μM. We can try a simplified version of this virtual screening.

The authors took 108 790 compounds and removed compounds that were too big. This resulted to 95 147 compounds. These compounds were docked into the binding site of Histamine 1 receptor and the authors also calculated "interaction fingerprints" (in this tutorial we will use only docking). Based on charge of the molecule, docking scores and fingerprints they selected 611 compounds. After further filtering and visual inspection they selected 26 compounds. These compounds were purchased and tested experimentally. Out of them, 19 showed strong binding to histamine 1 receptor.

In a simplified virtual screening procedure we will dock 21 compounds identified and experimentally confirmed in Ref. 1 as ligands of histamine 1 receptor. Then we will download another set of compounds randomly picked from ZINC database [3], we will dock them and compare them with known ligands. Finally we will evaluate the virtual screening procedure in terms of its ability to distinguish between ligands and inactive molecules.

Program PLANTS (Protein-Ligand ANT System) [4] is used for docking. It uses an algorithm called ant colony optimization, which is inspired by behaviour of real ants when they are trying to find the shortest path between food and ant nest. This program can be obtained under academic license from the web site of its developers.

Go to the PDB web site and find the structure of histamine 1 receptor in the complex with Doxepin [2]. Download the PDB file and make its copy. Open the copy in UCSF Chimera and remove all non-protein atoms. Then add hydrogen atoms. Use options "considering H-bonds" for protonation and "determined by the method" for His protonation. Then save the file in *mol2* formate. Addition of charges is not necessary.

Go back to original PDB file and view its text. Find the HETATM lines corresponding to the Doxepin and find its atom C6. This atom is approximately in the centre of the ligand and its coordinates can be used to define the centre of the binding site. Write down its Cartesian coordinates.

Take structures of known ligands from the articles. They could be supplied to you either as individual *mol2* files or a single *mol2* file. Place a ligand *mol2* file, receptor *mol2* file and PLANTS executable to a separate directory. Then copy there a configuration file from PLANTS example session.

Modify the configuration file:

```
# scoring function and search settings
scoring_function chemplp
search_speed speed2


# input
protein_file name_of_your_receptor_mol2_file
ligand_file name_of_your_ligand_mol2_file

# output
output_dir results

# write single mol2 files (e.g. for RMSD calculation)
write_multi_mol2 0

# binding site definition
bindingsite_center type here coordinates x, y and z of the binding site centre
bindingsite_radius 10.5000


# cluster algorithm
cluster_structures 10
cluster_rmsd 2.0
```

Next, type **./PLANTS --mode screen** *name_of_your_config_file* to the command line. Docking takes seconds to minutes depending on computer power, size and complexity of the ligand etc. When finished, go to **results** directory and view the file **bestranking.csv** to see predicted scores. Majority of real ligands should have the value of score bellow –90. You can also view predicted binding poses by viewing *mol2* files in VMD or UCSF Chimera together with the original PDB file.

Now we will chose and dock "decoys". Go to the web site of ZINC database and follow links *Subsets* > *Property* > *Clean fragments*. Then click on *sample molecules*. You will see the screen with structures of 50 molecules. By clicking to a structure you can add the molecule to a "shopping basket". Do this for an appropriate number of molecule. It may happen that there is a series of similar neighbouring molecules in the database. Therefore, try to select the set of decoys as diverse as possible. To do it, you can switch between screens by pressing *next* button or by typing page number. Once you made a selection, click to *Active cart: ...* and follow the links to download all structures as a single *mol2* file. Then make a new directory, place there decoys *mol2* file, other files and perform docking as described for active compounds.

Finally, we can evaluate the virtual screening procedure. Import files **bestranking.csv** from both docking runs (active compounds and decoys) to MS Excel or similar. Place the results of active compounds first, followed by decoys, to the same sheet. Delete all columns except the one with names and the one with the final scores (the first and second). You should have only two columns, for example **A** and **B**, one with name and one with scores. Now insert a column on the left side (column **A**) and type there values 1, 2, 3 … for each row. Now you have columns **A**, **B** and **C**. Now place charges of molecules to the column **D**. Place 1 to the column **D** for all active compounds (their charge is +1). For decoys you can use a special script to extract their charges. Place these charges to the column **D** for decoys. Then type 1 to the column **E** for all active compounds and 0 for decoys. This column will indicate whether the compound is active or not (we assume that decoys are inactive, however, it may happen that some compound from the ZINC database may be active).

Now use the sort function to sort all columns **B**-**E** by the final score (column **C**). Select these columns and sort data by column **C** in an ascending order. To the box **F1** type that is equal to **E1**. To the box **F2** type that it is **=F1+E2**. Copy this box down. This column indicates how many real ligands scored better then the compound in the row. To the box **G1** type 1, to **G2** type 2, and so forth, until you reach 21 in **G21**. Then type 21 to **G22**, **G23** and so forth. This column shows how the column **F** would look like in the ideal word where the docking program can perfectly distinguish real ligands from decoys. Finally, to the box **H1** type **=21*A1/***total number of compounds** and copy this box down. This box will indicate how the column **F** looks like when the docking program gives totally random scores. Now make a plot with column **A** at the horizontal axis and columns **F**, **G** and **H** on the vertical axis.

You can repeat the procedure with charges as the major criterion. Sort again the same columns, now by charge (column **D**) in descending order. Then chose compounds with positive charge and sort them by score (column **C**). Finally select compounds with zero and negative charge and sort them by score. Look at changes in the plot.

**References**


1. de Graaf C, Kooistra AJ, Vischer HF, Katritch V, Kuijer M, Shiroishi M, Iwata S, Shimamura T, Stevens RC, de Esch IJ, Leurs R. Crystal structure-based virtual screening for fragment-like ligands of the human histamine H(1) receptor. *J Med Chem* 2011, **54**, 8195-8206.
2. Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW, Kobayashi T, Stevens RC, Iwata S. Structure of the human histamine $H_1$ receptor complex with doxepin. *Nature* 2011, **475**, 65-70.
3. Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 2005, **45**, 177-182.
4. Korb O, Stützle T, Exner TE. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf Model* 2009, **49**, 84-96.