1 Tutorial for Virtual Screening

This tutorial is a rework of the tutorial called Virtual screening tutorial (*https* : *//web.vscht.cz/* $\sim$ *spiwokv/struktbio/*04*_histamine.pdf*) made by Vojtěch Spiwok because the one he made was kinda obsolete and it didnt say some steps very good. I heavily recommend reading the previous tutorial first to atleast understand what we are trying to do, in the case the tutorial no longer exist this tutorial will still work, but you may miss some background information.

1.1 CopyPaste from previous "Virtual screening tutorial"

This tutorial is inspired by the article of Chris de Graaf and coworkers [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 structurebased drug design. In the article of Chris de Graaf and coworkers, authors virtually screened 108 790 compounds. They selected, purchased and tested 26 of them, 19 turned out to be active with $KD < 10\mu M$. 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.

2 Reworked Tutorial

Firt we are going to prepare the system, i used a computer with Ubuntu 22.04, NVidia graphics card (RTX 2070), 64 Gb of RAM and a processor with 10 cores. The only thing of this you really need is the nvidia graphics card, as some softwares use CUDA and dont work on AMD.

We are gonna first install the PLANTS software, obtaining it from the internet is kinda difficult but i will directly show you the download link through Ubuntu terminal (if it isnt on the github repository in which you are reading this):

wget https://github.com/purnawanpp/plants/raw/main/plants

Executing this line on the terminal will download the executable for plants on the path that the terminal is aiming at the moment. Now lets download "ChimeraX", the previous tutorial actually use "Chimera" but with ChimeraX it worked just fine and maybe when you are reading this tutorial there is a new Chimera, so just download the newer one.

Now you will go to the PDB database and download the Histamine 1 receptor in the complex with Doxepin (https://www.rcsb.org/structure/3RZE) then you will open it on Chimera.

Now you will do two things, remove the "non-protein atoms" and add hydrogen:

- Remove Non-Protein Atoms: (ON TOP BAR)
  $Select > Chemistry > Protein.$
  $Select > Invert > SelectedModels.$
  $Actions > Atoms/Bonds > Delete.$
  Congratulations, you have removed the non-protein atoms!!

- Add Hydrogen:
  $Tools > StructureEditing > AddHydrogen$
  Click on Options
  Method: Also considers H-bonds (slower)
  Choose "Unspecified (determined by method)"
  Click on OK
  Congratulations, you just add Hydrogen Atoms!!

Now save the "new" structure in format .mol2.
Now you will open the Histamine file with format PDB (NOT THE RECENTLY SAVED MOL2) with a text editor (can be just the notepad) and you will search for the HETATM lines, specifically the atom C6 and you will write its cartesians coordinates as you will be using them later. This atom is approximately in the centre of the structure and its coordinates can be used to define the centre of the binding site.

Now from the reference 1 you will download the known ligands that dock to the histamine, the ones i could find with its ID and download are the nexts:

Fragment	Known active	pKi	pIC50
ZINC00006157	Epinastine	8.85	-
ZINC00000504	Mianserin	8.38	-
ZINC00002227	Triprolidine	8.70	-
ZINC01249433	CHEMBL10602	5.29	-
ZINC01723265	CHEMBL609579	5.56	-
ZINC00010402	Promazine	-	7.9
ZINC00000931	Amoxapine	7.4	-
ZINC00020245	Imipramine	-	7.5
ZINC01530611	Desipramine	-	6.7

Figure 1: Table with known ligands

To be noted, you only need one to test this tutorial. You can download them from the ZINC 12 DATABASE (or ZINC 15 if you manage to make the page load), just to be noted, this database have a shopping cart mechanic, after you put the molecules in this car you would want to download them, well, the button to download is the one called "Refresh" on the page and choose mol2 format. (Yeah, the one who make the refresh button a download button was a complete imbecile)

Now you will put this downloaded mol2 file in the same directory where your plants executable is and now you will create a .txt file with the next text inside it (You can name this file how you like, but i will call it "plants_config"):

```
# Command molecular docking in google colab using PLANTS 1.2
# 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
```

You can change this file as you wish, but you need to atleast know what are you doing, mostly you will be changing the output and input part, but it is your decision.
Now come the fun part, the plants executable, the txt config file (plants_config), the histamine mol2 file and the recently downloaded mol2 molecules will be put on a separate folder all together, in this new path you will open a new terminal and will write the next line and you will execute it:

./plants - -mode screen name_of_your_config_file

Congratulations, you have already done your first virtual screening run!!

Now you will open the folder where the results go and you will find a file called **bestranking.csv**, open it with excel or whatever you have and now you will find a lot of information, what we care is the second column called **TOTAL_SCORE**, this is the score of the molecule on the histamine. If you use only the downloaded from Figure 1 you will find that all of this are less than -90, this is your greatest result because if your molecule succesfully docked this number will be less than -90.

Now from the same ZINC12 databse download random molecules, fifty will do, and test them the same way you did it with the ones from the table, this molecules result will be more than -90 (we assume that decoys are inactive, however, it may happen that some compound from the ZINC database may be active and viceversa).

Congratulations for you, you now know how to do (basic) virtual screening, you can now change the histamine 1 receptor for whatever protein/bacteria/molecule you want and if you do the same steps with a group of molecules that you choose for whatever reason you will get results. This method isnt perfect, remember that you have to take account of electric charge, helicity and size, but this is the

first step to great results. I hope this has help you in whatever you are trying to do and remember, Someone is always moving on the surface.

3 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 structurebased virtual screening for fragmentlike ligands of the human histamine H(1) receptor. J Med Chem 2011, 54, 81958206.
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 H1 receptor complex with doxepin. Nature 2011, 475, 6570.
3. Irwin JJ, Shoichet BK. ZINCa free database of commercially available compounds for virtual screening. J Chem Inf Model 2005, 45, 177182.
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.