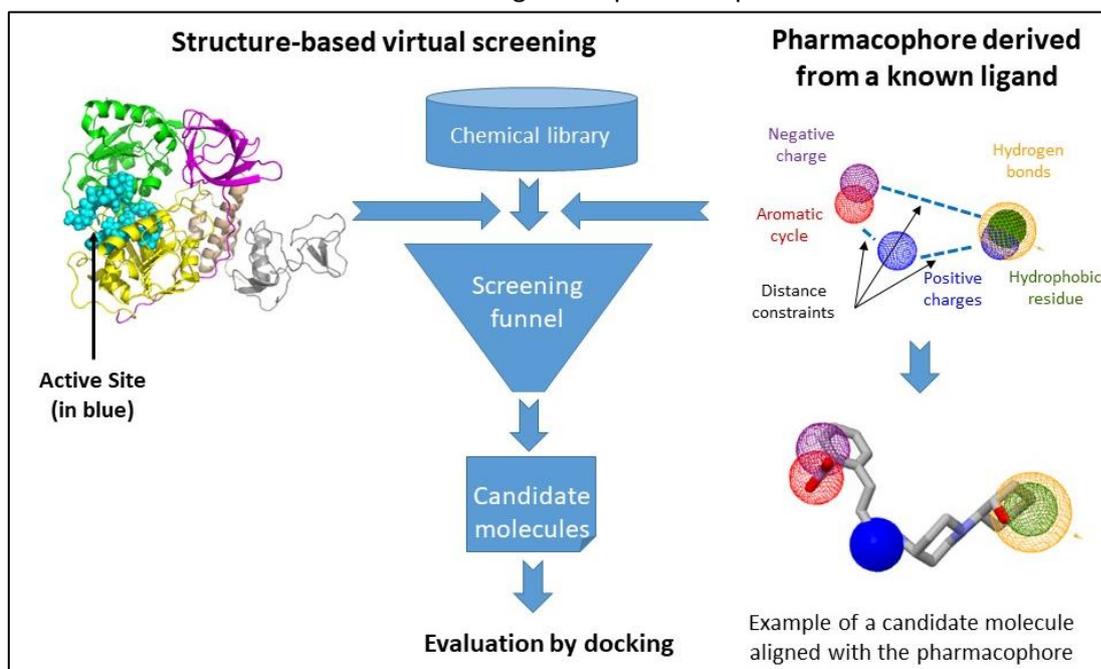# Tutorial : Virtual Screening with Pharmacophores using Pharmit (Pittsburgh University)

INTRODUCTION

Schema of structure-based virtual screening with a pharmacophore

Main reference for Pharmit : Sunseri J, Koes DR. Pharmit: interactive exploration of chemical space. Nucleic Acids Res. 2016 Jul 8;44(W1):W442-8. doi: 10.1093/nar/gkw287.

Official IUPAC definition, "a pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response."

For more information on pharmacophore see for instance Seidel T et al. Strategies for 3D pharmacophore-based virtual screening. Drug Discovery Today Technology 2010, 4:7.

I.   Connection to the Pharmit server online
https://pharmit.csb.pitt.edu/help.html

Take some time to read the first sections of the Help page (Getting started, Initiating a search, Visualization, Pharmacophore representations)

They are a good introduction to the functionalities of this server.

In particular, you will learn that you can initiate a search in three ways:

- from a PDB accession code,
- from receptor/ligand files (including both a receptor and ligand as well as a ligand by itself),
- or by using a pharmacophore generated from an external source.

In this tutorial we will start from a PDB accession code.

We will use the Covid-19 nsp13 protein that was studied in the article mentioned in the seminar (Md Kamrul ISLAM et al. Molecular-evaluated and explainable drug repurposing for COVID-19 using ensemble knowledge graph embedding, Scientific Reports13:3643, 2023. doi:10.1038/s41598-023-30095-z). Nsp13 is an helicase, critical for viral replication, and very much conserved across coronavirus strains.

The PDB structure used here comes from the PDB entry 7nn0.pdb (3.04 A resolution) which possesses a ligand (AMP-PNP, Adenylyl imidodiphosphate, a non-hydrolysable ATP analogue) in the ATP binding pocket. It is different from the one used for GOLD docking (7nio.pdb) which does not contain any ligand (also known as the 'apo' form). Attention, in the case of 7nn0 the last character is the number zero not an 'o'.

The interactions of the ligand with residues from both domains 1A and 2A are well described in the article by Newman JA et al. 2021).
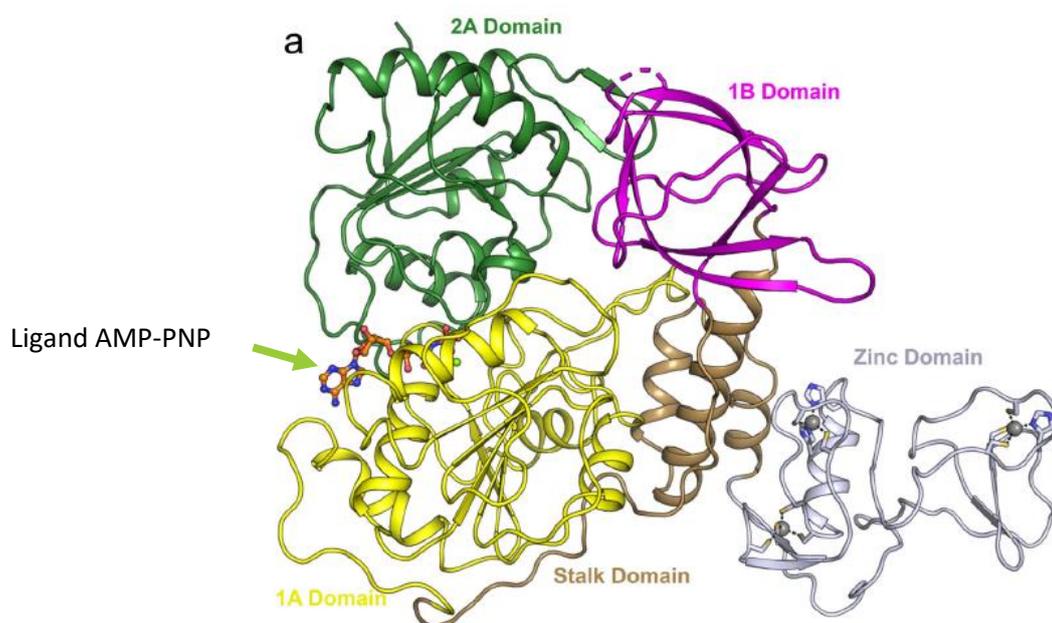

Figure 1: 3D model of SARS-CoV-2 nsp13 corresponding to PDB:7NN0 (Newman JA et al., Nature Comm 12:4848, 2021, doi:10.1038/s41467-021-25166-6). Domains are labeled and colored individually. AMP-PNP is shown in sticks. The binding pocket is formed by the junction of A1 (green) and A2 (yellow) domains.

II.   Entering a structure and a pharmacophore in a Pharmit session

Start a Pharmit session by connecting on https://pharmit.csb.pitt.edu/ . If you were on the **help** page, simply click on **home** at the top right of the window.

We will use the "**start from PDB**" option with PDB identifier 7NN0

Enter 7NN0 in the text field, close to "Start from PDB". The entry is recognized and the ligands appear in the field at the right of the PDB identifier? How many are listed ? Which one will be the important ligand for pharmacophore construction? Ignore water molecules and click on submit.

NB: Alternatively, if you have your own PDB structures for the receptor and a ligand of interest you want to use, you can start by clicking on '**enter pharmit search**' to reach the main interface. The

receptor can be loaded by the button '**Load Receptor**' and the ligand or already processed pharmacophore features can be loaded by the button '**Load Features'**.

Back to our entry (7NN0): take some time to explore the visualization window.

⇨ The structure has appeared. Scroll to the bottom left panel 'Visualization' and set Receptor Surface Opacity to 0 in order to see only the Cα backbone of the receptor. By default, the dispay is Cartoon + Wire for the receptor in order to see the lateral chains of the residues as sticks (very faint on the white background, more visible with black background, try to change using the '**Background color**' buttons).

⇨ The ligand appears as sticks in the active site. Ligand features are represented by spheres on the structure
- White spheres : Hydrogen donor
- Yellow spheres : Hydrogen acceptor
- Green spheres : Hydrophobic group
- Blue spheres : PositiveIon
- Red spheres : NegativeIon
- Purple spheres : Aromatic group

On the left, in the panel Pharmacophore, each feature (= a sphere on the visualization) is listed with several associated functionalities.

You can turn it ON or OFF, you can check the coordinates of the center of the sphere, and change the nature of the feature if you wish.

ATTENTION : do not click on the '**x**' on the right of a feature unless you want to delete it definitively.

FILTERS : In this window you can filter by size or ADME property range desired for the hits, and also fix the maximal number of results you want. For the moment we will not apply any filter.

Concerning our ligand, try to count how many features of each type you have. Note it in the table below

Type of feature	Symbol	Number in AMP-PNP ligand
Hydrogen donor	White spheres	
Hydrogen acceptor	Yellow spheres :	
Hydrophobic group	Green spheres :	
PositiveIon	Blue spheres :	
NegativeIon	Red spheres :	
Aromatic group	Purple spheres :	

Note that the two 'Aromatic' features at the beginning of the list (purple spheres) are superposed with two 'Hydrophobic' features (green spheres) at the end of the list which are by default inactivated (OFF).

The same for the two inactivated 'HydrogenAcceptor' features at the end of the list that are superposed with two 'HydrogenDonor' features.

Finally, at any time you can save your session in a json format and re-load it later in the interface.

III.   Exploring various pharmacophore search

The search space is huge with so many features but the Pharmit server allows to test various combinations of features by inactivating or activating desired features.

The expertise of knowing what feature to activate or inactivate comes with experience and some knowledge in chemistry. However, we can try some exploration in this section. At first, we will keep all features on 'ON' and then we will decrease the number of activated features to guide the search towards similar but distant molecules.

Each group of students can choose to test one of the proposed libraries in the scrolling menu of the Search button at the top of the left panel. Choose either:

- MCULE : ~40M molecules,
- Zinc : ~13 M molecules,
- MolPort : ~5 M molecules,
- CHEMBL32 : ~2 M molecules.

Avoid too large libraries (PubChem, ChemSpace) or not recently updated ones (ChemSpace, ChemDiv…).

You can create an Excel Googledoc document to trace all your attempts. Keep track in the table of the features you keep activated or you inactivate and of the numbers of molecule you retrieve

For example

Library searched	MolPort						
Search Number	Hydrogen donor	Hydrogen acceptor	Hydro-phobic	Positive Ion	Negative Ion	Aromatic	Number of hits
1	All (4)	All (15)	All (2)	-	All (3)	All (2)	0
2	None	None	All (2)	-	All (3)	All (2)	273
3	None	None	None	-	All (3)	All (2)	354
4	None	None	All (2)	-	All(3)	None	378
Etc.							

.

1.   With all ligand features

Click on Search 'your selected library' -> see results on the right.

For example in MolPort: No results found

Conclusion : the full pharmacophore is too restrictive, too many constraints.

2.   Remove all HydrogenDonor and HydrogenAcceptor features

Search MolPort -> 273 hits

- By default, the list is ranked by RMSD, the 3D-distance between each hit and the native ligand AMP-PNP.
- Aliases of the molecule in other libraries is given when passing over the molecule name.

- To visualize the molecule in the active site, click on the corresponding row and look at the visualization window.

3. Remove all HydrogenDonor and HydrogenAcceptor and the 2 hydrophobic features

Search MolPort -> 354 hits

4. Remove all HydrogenDonor and HydrogenAcceptor and the 2 aromatic features

Search MolPort -> 378 hits

Conclusion : Although the Hydrophobic and Aromatic features are superposed, they do not play the same role and having them in the search does not retrieve the same molecules.

5. Use shape constraints

Sometimes you see by visualization that some of the ligands are clearly too big to enter the binding sit. There is a clash with the Ca backbone.

- In the '**Shape**' window on the left, use '**Exclusive shape**' and set it to '**Receptor**', with a Tolerance of '**0.5**'.
- Launch the search again in you favorite library with your favorite combination of features and see the result.
- For instance, with MolPort and combination 2, one gets only 55 hits (instead of 273).
- Continue your table and record your results.

6. Use Hit Screening in the Filter section (left panel)

From the 55 hits retrieved from MolPort, some probably do not have ideal properties to become drugs. Such properties refer to the qualities expected from drugs and called ADME for Absorption, Distribution, Metabolism, Excretion. There are servers that compute physico-chemical parameters of molecules related to ADME such as SWISSADME. The Pharmit server allows you to filter the hits list with a series of appropriate thresholds in the Hit screening Filter menu in the lateral left panel.

Here, we can select a limit for the size. Usually, a druggable compound should not be larger that MW:500 (Lipinski rule) and to be specific enough it should be larger than MW:150. Solubility should be good enough but the molecule should not be too hydrophilic to be able to pass through the membranes. The thresholds values for the hit screening parameters should be taken from expert rules such as the Lipinski rule of 5*[1].

Test with the MolWeight range 150-500 on MolPort -> 8 compounds instead of 55.

Try to limit the LogP parameter (octanol-water partition coefficient) in the range [-10, 5] -> still 8 compounds. No reduction in the number of hits.

---

[1] Lipinski's rule states that, in general, an orally active drug has no more than one violation of the following criteria:[9]

- No more than 5 hydrogen bond donors (the total number of nitrogen–hydrogen and oxygen–hydrogen bonds)
- No more than 10 hydrogen bond acceptors (all nitrogen or oxygen atoms)
- A molecular mass less than 500 daltons
- A calculated octanol-water partition coefficient (Clog *P*) that does not exceed 5

Note that all numbers are multiples of five, which is the origin of the rule's name. As with many other rules of thumb, such as Baldwin's rules for ring closure, there are many *exceptions*.

Visualize some of the compounds by clicking on them. You can print a screenshot to show the superposition with the pharmacophore.

Follow the web link under the molecule name to display the compound entry pages.

Extract the SMILES representation and test in the SWISSadme server (http://www.swissadme.ch/index.php). Check for instance the solubility and many other properties.

7. Relax again pharmacophore constraints (if your list gets too short)

At this stage you can continue to relax the pharmacophoric constraints.

- For example, remove the intermediate NegativeIon (the second one in the list of 3).
- Keep the shape filter and hit filters.
- Run the search on Molport -> 28 hits.

Conclusion: Various sets of hits can be generated depending on the pharmacophore features and on the filtering criteria. It is difficult to determine in advance what the best list will be. In the frame of this tutorial, it is better to limit ourselves to a list of 20-30 molecules.

IV.   Ranking and downloading the results
Ranking can be performed in the Results panel by clicking on the desired column.

1. RMSD ranking

RMSD (Root Mean Square Deviation) represents the 3D shape difference between the hit and the pharmacophore. It is interesting to rank from smallest to largest RMSD. To do that, simply click on **the RMSD column header** ('RMSD'), either once or twice as it changed from smallest to largest and from largest to smallest at each click.

In the last set of 28 hits from MolPort, RMSD ranges from 0.335 to 0.888.

IMPORTANT : Pharmit tests several conformers for each molecule. Thus, two rows from the results table can correspond to the same compound under a different conformation (for example MolPort-051-694-605 appears 4 times in the last set of 28 hits).

2. Minimize ranking

This function minimizes the energy of the ligand bound to the receptor using smina, a fork of AutoDock Vina, using the default Vina scoring function. Click on the **Minimize** button at the bottom of the right panel to launch the Minimize function.

After the Minimize step the hits are presented with two values

- Score : the more negative the better (minimal interaction energy)
- mRMSD : Minimized RMSD

3. Producing the final non redundant list

Before downloading, if you want to retain all the hits with a negative score and a mRMSD below 2A, and without redundancy, enter the following parameters in the box at the bottom of the right panel:

- enter Max Score = 0
- enter Max mRMSD = 2A
- tick the Single conformer box to get a non-redundant list.
- click on **Apply** and then on the **Save** button

In our example, with the indicated threshold and one conformer per molecule we get a list of 13 compounds.

The default name when saved is 'minimized_results.sdf.gz'.

Rename it according to the PDB file used and the number of the attempt. For example : 7NN0_PharmitResult_Attempt01.sdf.gz . Don't forget to note in your Google Doc how this list was produced.

Unzip the file and check the content of the sdf file. Open in a text editor and count the occurrence of the separator $$$$ -> this fives the number of molecules that have been recorded in the file (here 13).

V.   Prepare the results for docking

You can try to use the public Galaxy server at https://Cheminformatics.usegalaxy.eu

- to convert the sdf file in a mol2 file format for GOLD
- or to prepare files for VINA Autodock software (if you are used to it)

To convert from sdf to mol2, select the **Compound Conversion** tool in the **ChemicalToolBox** menu in the left panel of Galaxy.

Upload your sdf file and select the '**Sybyl mol2 format (mol2)**' as output format.

Run without modification of the default parameters.

You see the status of your job on the right panel of Galaxy. When the color is green, you can download the result by clicking on the small eye at the right of your job name. The file is downloaded into the download directory of your computer. This mol2 file can now be uploaded on the GOLD Galaxy interface and tested with GOLD on the PDB:7nn0 or PDB:7nio structures.

CONCLUSION

The interactive Pharmit server is a powerful tool for manual exploration of chemical libraries with pharmacophores. The goal of this tutorial was to give an idea of what is a pharmacophore. Other tools exist for example on the Galaxy server for chemoinformatics. Testing various bioinformatics tools is time consuming but necessary before choosing the right tool for a task of interest.

****