Small-Molecule Library Screening by Docking with PyRx

Article *in* Methods in molecular biology (Clifton, N.J.) · January 2015

DOI: 10.1007/978-1-4939-2269-7_19 · Source: PubMed

2 authors, including:

Sargis Dallakyan
Home
16 PUBLICATIONS **288** CITATIONS

SEE PROFILE

Small Molecule Library Screening by Docking with PyRx

Sargis Dallakyan, and Arthur J. Olson

Department of Integrative Structural and Computational Biology, The Scripps Research Institute. 10550 North Torrey Pines Road La Jolla, CA 92037-1000 USA

Summary

Virtual molecular screening is used to dock small molecule libraries to a macromolecule in order to find lead compounds with desired biological function. This in silico method is well known for its application in computer-aided drug design. In this chapter, we describe how to perform small molecule virtual screening by docking with PyRx, which is open source software with intuitive user interface that runs on all major operating systems (Linux, Windows and Mac OS). We also describe specific steps for using PyRx, as well as considerations for data preparation, docking, and data analysis.

Key words: Virtual molecular screening, Computer-aided drug design, Molecular docking, PubChem, AutoDock, Vina, Open Babel

1. Introduction

SMALL MOLECULE LIBRARY SCREENING BY DOCKING WITH PYRX

Drug discovery is an attractive research area that enables application of cutting-edge biomedical research to improve health of many people ***(1)***. In the past, medicine was derived from natural products, mostly from plant sources. While natural products continue to be used and researched for medicine, it is now possible to synthesize a large number of chemical compounds that are not readily available in nature. The increased number of possible chemical compounds presents both a challenge and opportunity for the pharmaceutical industry. Testing different drug candidates in human clinical trials is a long and expensive process, which is why phenotypic or target based screening is so important in the earlier stages of drug discovery ***(2)***.

In phenotypic screening, different compounds are tested in cells or organisms to see which compound makes intended changes in the phenotype. When molecular causes of the disease are unknown, phenotypic screening is, in many cases, the only available option for finding life-saving drugs. For diseases that are well studied and understood at the molecular level, altering a single macromolecule can lead to a desired outcome. We will discuss an example of such a macromolecular target for the common flu virus shortly. In target based screening, compounds are tested with purified macromolecules (usually a protein) to find lead compounds that make intended macromolecular changes. For a lead compound to become a drug, it needs to be able to reach a site of action in the body, bind to its target macromolecule, and elicit the desired biological effect.

Compared to large biological molecule therapeutics, such as insulin or antibodies, which are administered through injection, small molecules can be taken orally and are better at reaching different sites in the body. This is why the majority of approved and experimental drugs are small molecules. Small molecules are also better suited for virtual molecular screening, which is the main subject of this chapter. With virtual screening, we dock different compounds from a small molecule library to a target macromolecule (usually a protein) to find compounds with the best binding affinity ***(3)***. Note that virtual screening is not limited to drug targets and it can be used to screen against herbicides, pesticides, or any other target of interest ***(4)***.

In all cases, finding the right target is very important for virtual screening campaign to succeed. When the 3D (three-dimensional) structure of a target is available, through X-ray crystallography, NMR spectroscopy, or any other means, we can apply docking algorithms to search for the best binding mode between target macromolecule and ligand.

In this chapter, we outline methods for performing virtual screening experiments with PyRx open source software. We use the 3D structure of the influenza virus neuraminidase ***(5)*** as an example to show how to prepare an input file for the target macromolecule. Influenza virus neuraminidase cleaves sialic acid from the infected cell surface to release newly created viruses. Neuraminidase inhibitors bind to neuraminidase and prevent them from binding to sialic acid. This leaves the influenza virus stuck on the surface of infected cells, so that the influenza virus cannot infect nearby healthy cells ***(6)***. Here, we describe steps to prepare input structures for zanamivir (a neuraminidase inhibitor), sialic acid, and sucrose (table sugar). We then use these small molecules to run virtual screening against influenza virus neuraminidase.

2. Materials

PyRx is written in Python programming language and it can run on nearly any modern computer, from PC (personal computer) to supercomputer. Below we provide details of the Windows PC used in the Methods section, although similar methods also work on Linux and Mac OS as well.

2.1 Hardware and Software

1. Dell Studio 540S with Intel Core 2 Duo CPU @ 2.53 GHz, 4GB memory (RAM), ATI Radeon HD 3400 series graphics card, and 32-bit Windows Vista operating system.
2. Binary distribution of PyRx version 0.8 for Windows available free from http://pyrx.sourceforge.net.

2.2 Input Files

To start with structure-based virtual screening, we need structures of the target macromolecule and small molecules as input files. There are a number of publicly available websites where users can download these input files. Here we are using: DrugBank ***(7)*** to get the structure of zanamivir, PubChem ***(8)*** for 3D structure of sucrose, and Protein Data Bank ***(9)*** to get 3D structures of influenza virus neuraminidase and sialic acid.

1. Open your web browser and visit http://www.drugbank.ca/drugs/DB00558, click on SDF link next to Download and save that page as DB00558.sdf.
2. Go to http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=5988, click SDF icon on top right and select 3D SDF: Save.
3. Visit http://www.pdb.org/pdb/explore/explore.do?structureId=2BAT, click on Download Files and select PDB File (Text).

The reason for choosing these particular molecules is that they are familiar to most of the readers and our computations can be run relatively quickly on a PC. To apply the protocol described in the Methods section to other binding target and ligands, users would need to obtain input files corresponding to their specific binding target and ligands. Selection of the binding target depends on the biological problem of interests, and we assume that 3D structure of the target is available in PDB format through Protein Data Bank (http://pdb.org) or other sources (*see* **Note 1**). Selection of ligands depends whether virtual screening is used for lead discovery or lead optimization. For lead discovery, it is advised to include as many ligands with diverse shapes, sizes, and composition as possible. Since individual docking computations are independent from each other, users are practically only limited by computational power available at their disposal. For lead optimization, on the other hand, ligands are selected to closely match the lead compound ***(10)***. One of the advantages of the virtual screening is that we are not limited to commercially available compounds; we can also use a ligand file for a novel compound not found in any of the databases (*see* **Note 2**).

3. Methods

3.1 Prepare Input Files for Docking

Before we can use input files for virtual screening, we need to convert them to PDBQT file format suitable for docking with AutoDock Vina ***(11)***.

1. Start by double-clicking on PyRx icon on the Desktop.
2. Select Open Babel tab under Controls panel and click on the first icon on its toolbar with plus (+) sign on it. Navigate to Downloads folder and select CID_5988.sdf (sucrose from PubChem).
3. Click on the first icon on Open Babel toolbar again, locate and open DB00558.sdf (zanamivir from DrugBank). 558 is the accession number of zanamivir in DrugBank and it is listed under Title column in Open Babel table. If you have other molecules to include in virtual screening, you can use Open Babel widget to convert them to PDBQT file format (*see* **Note 3**).
4. Select row corresponding to zanamivir with Title 558, right-click and use Minimize Selected option. Click OK and wait for energy minimization to complete. Notice that the title of this molecule has changed to 558_uff_E=197.68. The _uff part corresponds to the force field used for energy minimization, which, by default, is the Universal Force Field ***(12)*** as implemented in Open Babel software package ***(13)***. The _E=197.68 part corresponds to the energy of the minimized molecule. The precise value for this energy is not important here. However, this notation is helpful to capture changes made to this molecule before we convert them AutoDock ligand file in the next step.
5. Right-click on any of the rows in Open Babel table and use Convert All to AutoDock Ligand (pdbqt). This will create two pdbqt files corresponding to sucrose and zanamivir molecules under Ligands folder.
6. Select Documents tab under View panel, click on Open icon (second from the left) and open 2BAT.pdb file. 2BAT is the PDB ID for the structure of the complex between influenza virus neuraminidase and sialic acid ***(5)***. The following steps are specific to this structure. To apply this method to targets which have no ligand attached, please go directly to step 10 and replace 2BAT with the name of your target macromolecule.

7. Next we need to select lines corresponding to sialic acid from 2BAT.pdb. Scroll down, use Ctrl-F or Find icon on the toolbar to search for SIA residues and select lines with HETATM 3216 to 3236. Use Ctrl-C or right-click Copy, click on the New icon and paste these lines (Ctrl-V or right-click Paste) in a new file. Save this file as SialicAcid.pdb using Save icon (3rd from the left) on the Documents panel. If you are working with another target that contains a ligand that you want to re-dock, you can use Documents panel in PyRx or any other text editor of your choice (such as Notepad or WordPad) to extract HETATM records corresponding to your ligand of interest. The web page for 2BAT (http://www.rcsb.org/pdb/explore.do?structureId=2BAT) also lists different ligands bound to neuraminidase, including sialic acid, which is listed under Ligand Identifier column as SIA. This web page also offers the possibility to download ligand SDF file for sialic acid.

8. Click on 2BAT.pdb tab under Documents panel, scroll up and left-click at the beginning of the line starting with TER 3023. The TER record indicates the end of a list of ATOM records for a chain according to PDB file format specification. In our case, we want to keep neuraminidase atom records only and delete all other records that correspond to different ligand and water molecules cocrystalized with this structure of neuraminidase. With the left mouse button pressed, scroll down till the end of the file and click Delete. Save this modified 2BAT.pdb file using Save icon again.

9. From the menu bar, use File → Load Molecule menu and open SialicAcid.pdb. Right-click on SialicAcid under Molecules panel and select AutoDock → Make Ligand.
10. Use File → Load Molecule menu again and open 2BAT.pdb. Right-click on 2BAT under Molecules panel and select AutoDock → Make Macromolecule.

3.2 Run Virtual Screening Using Vina Wizard

1. Select Vina Wizard tab under Controls panel and click on Start button.

2. Select 558_uff_E=197.68.pdbqt, 5988.pdbqt and SialicAcid.pdbqt under Ligands folder (use the Shift key for selecting multiple ligands).
3. Select 2BAT under Macromolecules folder and click on Forward button on Vina Wizard.
4. Click on Maximize button under Vina Search Space and then click on Forward button. This starts AutoDock Vina and docks each ligand, one by one, to neuraminidase (2BAT). It takes less than 20 minutes to complete this virtual screening on a PC mentioned in **Subheading 2.1** (*see* **Note 4**).
5. After virtual screening is completed, PyRx automatically advances to Analyze Results page, where you can see results of virtual screening computation. AutoDock Vina, by default, outputs 10 best binding modes for each docking run (*see* **Note 5**). Left-click on Binding Affinity (kcal/mol) table header cell under Analyze Results tab to sort this table by predicted binding affinity (*see* **Note 6**).

4. Notes

1. During docking runs, the 3D structure of the target is fixed while ligand is moved and rotated to find the best binding modes. While it is possible to make some of the side chains flexible during the docking, incorporating full flexibility of the target is still a subject of active research (***14***).
2. There is a variety of desktop or Web-based molecular editors available that be used to generate a ligand file for a novel compound not found in any of the databases. The Web-based molecular editors allow users to sketch molecules in 2D, while desktop tools such as Avogadro ***(15)*** can draw molecules in 3D.
3. SDF (Structure-Data File) format is commonly used to store multiple structures in a single file. It allows storing arbitrary data together with coordinates and atom types. Oftentimes, small molecules stored in SDF are flat (2D) and we do energy minimization to get 3D structures with proper bond length between different atoms.
4. The main results from virtual screening runs are the best predicted binding modes and corresponding binding affinity. The negative values for binding affinity (or binding free energy)

indicate that the ligand is predicted to bind to a target macromolecule. The more negative the numerical values for the binding affinity, the better is the predicted binding between a ligand and a macromolecule. In this particular case of screening neuraminidase with zanamivir, sialic acid and sucrose we can see from **Fig 1** that zanamivir (2BAT_558_uff_E=197.68) and sialic acid (2BAT_SalicAcid) are both predicted to have the best binding affinity of -7.3 kcal/mol, whereas the best binding mode for sucrose (2BAT_5988) is predicted to have binding affinity of -7.0 kcal/mol. In other words, zanamivir and sialic acid are predicted to have better binding affinity to neuraminidase than sucrose. The fact that both zanamivir and sialic acid have the same predicted binding affinity indicates that zanamivir can bind to neuraminidase and inhibit it from binding to sialic acid.

5. PyRx users can also export virtual screening results as CSV (Comma Separated Values) or SDF files. This is useful for further analysis, filtering or reranking of virtual screening results with third-party packages.
6. There are a number of approximations used to model protein-ligand interactions ***(16)*** and there are a number of unknowns when it comes to comparing virtual screening results with experiments ***(17)***, not the least of which is that we are docking a single protein with a single ligand. In practice, even with purified samples, it's hard to predict if proteins or small molecule ligands would aggregate and idealistic prediction of binding affinity with single protein-ligand docking applies to diluted samples. Nevertheless, small molecule virtual screening by docking is very valuable in silico method that can rank small molecules according to their predicted binding affinity to a target macromolecule. The cost of running virtual screening experiments is minuscule compared to real screening experiments. Virtual screening is also very good tool for hypothesis generation with which we can test modified versions of existing compounds or custom compounds that are not commercially available. With advances in computer software and hardware, and with the increasing number of publicly available bioassay data, virtual screening will continue to remain a vibrant research field.

Acknowledgements

This work was supported by National Institute of Health (NIH) grant R01GM069832. The authors are very grateful to all those who contributed software used in this chapter. We also would like to thank Alex L. Perryman, Oleg Trott, Ruth Huey, and all members of our laboratories for helpful comments and discussions. This is manuscript # 25035 from The Scripps Research Institute.

References

1. Ng, R. (2008) Drugs: From Discovery to Approval, 2nd edn. Wiley-Blackwell, Hoboken, New Jersey.
2. Swinney, D.C., and Anthony, J. (2011) How were new medicines discovered? *Nat Rev Drug Discov.* **10**, 507-19.
3. Jacob, R.B., Andersen, T., and McDougal, O.M. (2012) Accessible high-throughput virtual screening molecular docking software for students and educators. *PLoS Comput Biol.* doi:10.1371/journal.pcbi.1002499.
4. Lindell, S.D., Pattenden, L.C., and Shannon, J. (2009) Combinatorial chemistry in the agrosciences. *Bioorg Med Chem.* **17**, 4035-46.
5. Varghese, J.N., et al. (1992) The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins.* **14**, 327-32.
6. McKimm-Breschkin, J.L. (2013) Influenza neuraminidase inhibitors: antiviral action and mechanisms of resistance. *Influenza Other Respi Viruses.* **7**, 25-36.
7. Knox, C., et al. (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **39**, D1035-41.
8. Li, Q., et al. (2010) PubChem as a public resource for drug discovery. *Drug Discov Today.* **15**, 1052-7.

9. Berman, H.M., et al. (2002) The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr.* **58**, 899-907.
10. Scior, T., et al. (2012) Recognizing Pitfalls in Virtual Screening: A Critical Review. *J Chem Inf Model.* **52**, 867-81.
11. Trott, O., and Olson, A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* **31**, 455-61.
12. Rappe, A.K., et al. (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10024–10035.
13. O'Boyle, N.M., et al. (2011) Open Babel: An open chemical toolbox. *J Cheminform.* doi:10.1186/1758-2946-3-33.
14. Perryman, A.L., et al. (2010) A dynamic model of HIV integrase inhibition and drug resistance. *J Mol Biol.* **397**, 600-15.
15. Hanwell, M.D., et al. (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform.* doi:10.1186/1758-2946-4-17.
16. Muchmore, S.W., et al. (2010) Cheminformatic tools for medicinal chemists. *J Med Chem.* **53**, 4830-41.
17. Nicholls A. (2008) What do we know and when do we know it? *J Comput Aided Mol Des.* **22**, 239-55.

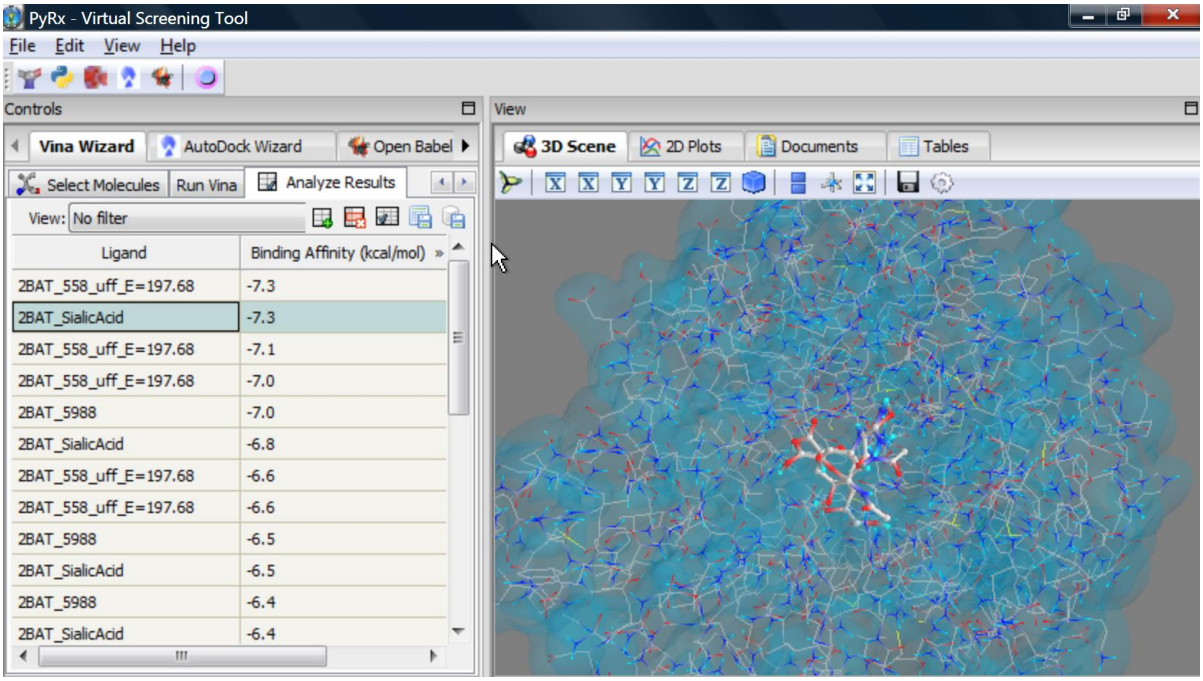

Figure Caption

Fig. 1. A screenshot of PyRx virtual screening tool. The table on the left lists predicted binding affinity of zanamivir (2BAT_558_uff_E=197.68), sialic acid (2BAT_SalicAcid), and sucrose (2BAT_5988) for influenza virus neuraminidase (2BAT). 3D scene on the right shows line drawing and transparent molecular surface of neuraminidase. Ball and stick models for zanamivir and sialic acid are also shown on this 3D scene.