SI100 Project: Virtual Screening

Instructor: Fang Bai

Spring, 2023

Due time: 11:59 pm, June 4, 2023

Introduction

Virtual screening (VS) is a computational technique used in drug discovery to search libraries of small molecules in order to identify those structures which are most likely to bind to a drug target, typically a protein receptor or enzyme.

VS has several advantages over traditional screening methods, such as higher hit rate, lower cost, faster speed, and wider chemical diversity. it can also be used to discover new targets or potential side effects of drugs by reversing the roles of ligands and targets. VS is an important tool for modern drug discovery programs. Virtual screening can help design and optimize targeted libraries, enrich libraries of available compounds, and select the most promising compounds for further experimental testing.

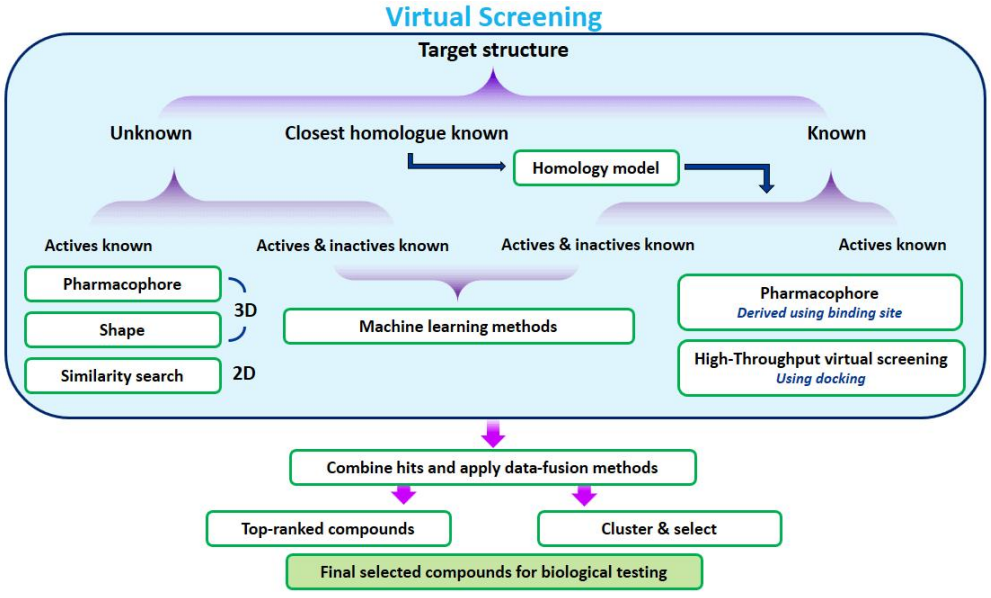

Figure 1. Virtual screening process

What we want you to accomplish in this project is a simple virtual screening task. Our virtual screening task will be based on the DUD-E database. DUD-E (Directory of Useful Decoys, Enhanced) is a benchmarking dataset for virtual screening of small molecules against protein targets. It contains 102 protein targets and 22,886 active compounds from ChEMBL database. For each active compound, 50 decoy compounds with similar physicochemical properties but different topology are selected from ZINC database. The decoy compounds are also matched by net charge and filtered by structural diversity. The purpose of DUD-E is to provide challenging decoys for testing docking algorithms and evaluating their performance.

Molecular fingerprint is a way of representing the structure of a molecule as a mathematical object, such as a vector or a hash. It can encode various features of the molecule, such as its atoms, bonds, functional groups, substructures, or properties and it can be used for various purposes, such as similarity search, virtual screening, clustering, classification, or chemical space mapping. By comparing the fingerprints of different molecules, one can measure their similarity or dissimilarity based on their structural features or properties. We hope you can master how to find similar molecules through molecular fingerprints and identify possible relationships between similar molecules and their binding complex.

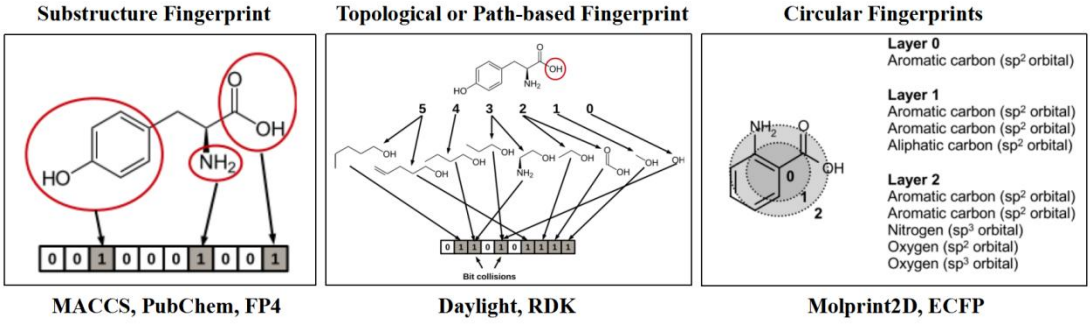

Figure 2. Molecular fingerprint

Background

fabp4 is a fatty acid binding protein that is mainly present in adipocytes, macrophages and endothelial cells. It can regulate lipid metabolism and inflammatory response, and is associated with various diseases such as diabetes, obesity, atherosclerosis and cancer. In ischemic stroke, fabp4 can exacerbate the damage of the blood-brain barrier, leading to neuronal death. In septic acute kidney injury, fabp4 can promote inflammation and apoptosis of renal tubular epithelial cells through TLR4/c-Jun signaling pathway, resulting in kidney dysfunction and tissue injury6. Therefore, fabp4 may be an effective therapeutic target, and inhibiting its activity or expression can protect tissues and organs from ischemic or infectious damage.

Task

1. Different types of molecular fingerprints can be classified based on different fingerprint generation algorithms and the different molecular properties considered. Please investigate the molecular fingerprints of your interest through the literature and provide a brief introduction for each one (include at least three types of molecular fingerprints and provide references at the end, submitted file name: "T1_FP.docx"). (10')

2. Please compare the molecular fingerprint similarity of the active molecule, the inactive molecule and the crystal ligand separately (any molecular fingerprint you are interested in) based on the documents provided in the supplementary material and export the results as two csv files (submitted files name: "T2_active_FP_result.csv" and "T2_inactive_FP_result.csv"). (30')

3. Select the five active molecules with the highest molecular fingerprint similarity to the crystal ligand and the five inactive molecules with the lowest molecular fingerprint similarity to the crystal ligand, and use molecular docking software to dock these ten molecules to the fabp4 receptor (docking results are placed in a folder named "T3_results"). (20')

4. Based on the docking results and knowledge of biology, try to explain why the active molecule plays an inhibiting role on fabp4 receptor, but the inactive molecule does not (no standard answers, submitted file name: "T4.docx"). (10')

5. After your group completes the project, please report during the 13-16 week class time (Wednesday/Friday). Note: please contact TA in advance to confirm the report time. You should demo your code and problem-solving process to TA in presentation and answer questions from TA. (30')

- Please submit all the code your group used to solve the above tasks (if you do this).
- Please package all the files and name them "SI100F-GroupName-P4".
- Failing to submit the project package before the due date will receive 50% of the score. No submission will be no score for the project.
- If you have questions about the topic (not coding problems), please contact the teaching assistant. lizq1 or wangshh12022@shanghaitech.edu.cn

Reference

https://en.wikipedia.org/wiki/Virtual_screening

https://www.sciencedirect.com/topics/medicine-and-dentistry/virtual-screening

https://dud.docking.org/

https://www.youtube.com/watch?v=kBk8HbjWwCw

https://www.nature.com/articles/s41419-022-04794-w