Project Proposal: Performing virtual screening by using multiple protein conformations

Aleksandra Olshanova

Introduction

Virtual screening (VS) is a powerful technique for identifying hit molecules as starting points for many industrial and academic drug discovery projects. Indeed, the number of methods and softwares which use different types of VS approaches is increasing at a rapid pace [1]. One type of them is structure - based virtual screening which plays a key role in the early stage of drug discovery. This process includes different methods to filter a chemical compound library using the structure of the molecular target of interest [2].

Statement of problem

Available experimentally - determined 3D structures of the proteins are stored in the Protein Data Bank (PDB) archive, and the number of experimentally solved structures is increasing rapidly [3]. In order to obtain accurate docking results, it is important to have a protein structure from PDB with a co-crystallized ligand which is chemically similar to target compounds available for screening. And based on the chosen protein structure we can obtain various results for structure-based virtual screening. But even if we consider pairs of proteins with high sequence identity, they might have geometric differences over regions that are well-aligned in sequence [4]. At the same time identical protein structures might be crystallized with particular ligands which define different properties of the binding pocket. As a consequence this will lead to docking results which have significant differences for different conformations of the same protein.

Objectives

The present project is focused on an assessment of docking performances across different protein structures of the acriflavine resistance B (AcrB) protein, downloaded from the PDB. AcrB is the inner membrane protein of the efflux complex and is responsible for the recognition and binding of compounds before their transportation out of the cell [5].

In this project I want pursue the following goals:

- Perform structure-based virtual screening with the Glide program
- Evaluate the influence of protein structure on docking performance by producing ROC curve

The stages involved in this project

1) Retrieve from the PDB a high resolution structures of protein – ligand complexes of AcrB inner membrane protein.
2) Run a SiteMap search for all possible existing binding sites in each protein structure to define the most promising pocket.

3) Align all protein structures of AcrB based on the best found binding site.
4) According to received RMSD values perform hierarchical clustering on these aligned structures.
5) Retrieve from the CheMBL database a set of active compounds against AcrB protein.
6) Retrieve from the DUD-E database a set of appropriate decoy compounds.
7) Test Glide program for its ability to reproduce the experimentally determined protein-ligand structure for each cluster.
8) Dock the library of active + decoy compounds against a selected protein target from each cluster.
9) For each selected protein, produce a ROC curve of docking results for active and decoy compounds.
10) Rank the different clusters based on the resulting ROC Area Under the Curve (AUC) score and analyse the differences between each cluster.

References

1) Lavecchia A., Di Giovanni C. Virtual screening strategies in drug discovery: a critical review. // Current medicinal chemistry. 2013. 20(23). P. 2839–2860.
2) Kitchen D. B., Decornez H., Furr J. R., Bajorath J. Docking and scoring in virtual screening for drug discovery: Methods and Applications. // Nat. Rev. Drug Discovery. 2004. V. 3. P. 935– 949.
3) Burley, S.K. et. al. Protein Data Bank: A comprehensive review of 3D structure holdings and worldwide utilization by researchers, educators, and students. // Biomolecules. 2022. V. 12. P. 1425.
4) Kosloff M., Kolodny R. Sequence-similar, structure-dissimilar protein pairs in the PDB. // Proteins. 2008. 71(2). P. 891-902.
5) Rajapaksha P., Pandeya A., Wei Y. Probing the dynamics of AcrB through disulfide bond formation // ACS Omega. 2020. 5 (34). P. 21844-21852.