HIGH-THROUGHPUT DOCKING IN MODERN DRUG DESIGN

Khomovych A.P., Bilyk O.A., Klenina O.V., Chaban T.I.
Danylo Halytsky Lviv National Medical University, Lviv, Ukraine
homovica1601@gmail.com

One of the most widely-used approaches in modern drug design is structure-based virtual screening (SBVS) methodology based on molecular docking performance protocols. In molecular docking, the pose of a drug-like small molecule is optimized within the receptor binding site, and then the best (or the few lowest-energy poses) are estimated quantitatively with a scoring function which is a measure of the likelihood of binding.

Obvious advantages of molecular docking tool can be summarized as: (f) there is a decrease in the time and cost involved in the screening of millions of small molecules; (b) there is no need for the physical existence of the molecule, so it can be tested computationally even before being synthesized; (c) several virtual screening tools are also available to assist in a structure-based approach like fingerprints pharmacophore modeling.

Virtual screening protocols has some disadvantages also which can be highlighted as the following: (a) it is difficult to accurately predict the correct binding position and classification of compounds due to the difficulty of parameterizing the complexity of ligand-receptor binding interactions; (b) both false positive and false negative poses may be generated.

In structure-based virtual screening, the docking process is performed in a high-throughput fashion (high-throughput docking, HTD) using chemical libraries of small molecules, with the aim of prioritizing a hit-list enriched with potential ligands. Top-ranking compounds are then synthesized (or purchased), and biologically evaluated in binding and/or functional assays.

A schematic workflow of HTD protocol commonly includes four main steps:

1. Target protein identification and its structure availability: an obviously necessary component in HTD is the 3*D* structure of the target receptor. There are approximately 140 000 macromolecular structures deposited in the Protein Data Bank (PDB, www.rcsb.org), and this number continuously increases every year. Of those structures, about 90% have been solved by x-ray crystallography, and the rest 10% by NMR, a technique somehow limited by protein molecular weight and solubility. In cases where no experimental structure is available, protein homology modeling (HM) may provide structural models with different degrees of accuracy to be used in HTD Consideration regarding the inclusion or not of crystallographic water molecules should be taken at this stage. Also the protein binding site must be selected and delimited. This step can be done manually by specifying the coordinates, or automatically using the coordinates of any bound ligand. Additionally, some programs allow for the calculation of cavities or probable binding sites.

2. Preparation of ligands database. Ligand preparation involves similar considerations. In virtual screening, ligands may come from sources different from PDB (for example public repositories like PubChem or ZINC15 database, organic synthesis or virtual compounds libraries). In such cases, the procedure may vary, but it involves the construction of ligands molecules from their Simplified Molecular Input Line Entry (SMILES) format or sketching molecules and save their structures in a SDF/MOL file, which serves as input. Later on,

following molecule construction several optimizations may follow: geometry and/or charge assignment using ab initio or semiempirical methods, energy minimization or conformational/tautomer search.

3. Following the ligand and protein preparation, the type of docking must be selected. In the case of flexible docking, residues are then selected as additional terms for energy calculation. This may be done manually (Autodock, Vina) or automatically (ICM, MOE). This usually means the potential is "softened" to allow residue flexibility and its interaction with the ligand. In the rigid molecule docking both the target and ligand are related to as rigid molecules that cannot change their spatial shape during the docking process.

4. Scoring functions are then calculated and differentiate true binding modes from all the other parallel modes evaluating a broad range of properties including, but not limited to, intermolecular interactions, desolvation, electrostatic, and entropic effects. Molecular docking software uses scoring functions to estimate the force of non-covalent interactions between a ligand and molecular target using mathematical methods. A scoring function is primarily responsible for predicting the binding affinity between a target and its ligand candidate. Scoring functions can be classified as force field-based, empirical, and knowledge-based. The force field scoring functions are based on the intermolecular interactions between the ligand and target atoms, such as the van der Waals, electrostatic and bond stretching/bending/torsional force interactions, obtained from experimental data and in accordance with the principles of molecular mechanics. Some published force-field scoring functions include Goldscore and Sybyl/D-Score. Empirical scoring functions estimate the binding free energy based on weighted structural parameters by adjusting the scoring functions to experimentally determine the binding constants of a set of complexes. To create an empirical scoring function, a set of data from protein-binding complexes whose affinities are known is initially used for training. A linear regression is then performed as a way of predicting the values of some variables. The weight constants generated by the empirical function are used as coefficients to adjust the equation terms. Each term of the function describes a type of physical event involved in the formation of the ligand-receptor complex. Thus, hydrogen bonding, ionic bonding, non-polar interactions, desolvation and entropic effects are considered. Some popular empirical, scoring functions include Glide-Score, Sybyl-X/F-score and DOCK 6 empirical force field. In the knowledge-based scoring functions, the binding affinity is calculated by summing the binding interactions of the atoms of a protein and the molecular target. These functions consider statistical observations performed on large databases. The method uses pairwise energy potentials extracted from known ligand-receptor complexes to obtain a general scoring function. These methods assume that intermolecular interactions occurring near certain types of atoms or functional groups that occur more frequently are more likely to contribute favorably to the binding affinity. The final score is given as a sum of the score of all individual interactions. One example of software that uses a knowledge-based scoring function is ParaDockS.

To verify the quality of a docking approach, some methods are used to evaluate generated complexes and to verify if the protein generated by the docking can reproduce the experimental data results of the ligand-receptor complex. The most common evaluation methods are root mean square deviation (RMSD), receiver operating characteristic (ROC), area under the curve ROC (AUC-ROC), enrichment factors (EFs) and Boltzmann-enhanced discrimination of ROC (BEDROC).

Identification of active drug candidates is the final result of docking performing. After hit identification lead optimization should be fulfilled under which biologically active compounds are transformed into appropriate drugs by improving their physicochemical properties. Finally these optimized leads will undergo preclinical and clinical trials to ultimately be approved by regulatory bodies.

Examples of drugs that came to the market with the assistance of virtua; screening include captopril (antihypertensive drug), saquinavir, ritonavir, and indinavir (three drugs for the treatment of human immunodeficiency virus), tirofiban (fibrinogen antagonist), dorzolamide (used to treat glaucoma), zanamivir (a selective antiviral for influenza virus), aliskiren (antihypertensive drug), boceprevir (protease inhibitor used for the treatment of hepatitis C).