ISSN: 2321-6743

Review Article

International Journal of Pharmacy and Natural Medicines

www.pharmaresearchlibrary.com/ijpnm

A Breif Review on Virtual Screening in Drug Discovery

Jahnawi.CH*, V. Vinodini, S. Ananda Kumar

Vellore Institute of Technology, School of advanced sciences, School of computing science and engineering, Vellore, 632014, TamilNadu

Received: 02 May 2014, Accepted: 27 September 2014, Published Online: 15 December 2014

Abstract

There is rapid increase in rate of identification of novel therapeutic targets and characterization of compounds 3Dstructures. There is continuous development in computational screening methods and enhanced as acceptable and complementary choices to high-throughput biochemical compound screening (HTS). Currently majority of drug candidates had headway have been found through HTS methods. Virtual Screening, High-throughput docking and pharmacophore based searching algorithms are acquiring acceptance and evolving into appropriate tools for major source of lead molecules in drug discovery. In the pharmaceutical industry, receptor-based virtual screening has become a viable source of novel leads. In contrast to high-throughput screening, in virtual ligand screening (VS), compounds are identified using computer programs to predict their affinity to a target receptor. It is necessary to have proper understanding about the spatial and energetic criteria accounting for protein–ligand binding. The concepts and conditions to carry out VS are discussed in this paper. Target selection, analysis and preparation are explained, and also determinants about the compilation of candidate ligand libraries.

Keywords: Virtual screening- Target and Database preparation- Scoring- Post analysis

Contents

1. Introduction 156
2. Molecular docking. 158
3. Virtual Screening Tools. 159
4. References 160

***Corresponding author**
Jahnawi.CH
School of advanced sciences, School of computing science and engineering,
Vellore, 632014, TamilNadu
Manuscript ID: IJPNM2100

PAPER-QR CODE

1. Introduction

It's been a challenge for pharmaceutical industries to bring about new drugs in to the market for their survival and success. It is obvious that it is a tedious process of more than 14 years to identify a compound as a hit and introducing it into market as a approved drug[1]. This process involves enormous cost and investments which may or may not in the end give fruitful results. In conventional methods for developing a therapeutic drug from lead

compound requires a High throughput screening of compounds available in limited collection of compounds known to researcher knowledge.

High-throughput screening (HTS) helps in identifying the compounds with desired activity. 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. The term Virtual screening was coined in the late 1990's. However, VS technique is more efficient than HTS in the fact that it employs computer-based searches for selecting potent drugs. VS make use of fast and accurate algorithms. VS is now a standard tool in medicinal chemistry. The evolution of virtual screening methods expands the feasibilities of various molecules for a researcher which is not substantial but easily procured or can be synthesized. Moreover, Virtual screening has evolved into a productive process for steering compound as a hit molecule to lead optimization. VS enable chemists to reduce a large virtual library to a manageable size. HTS assays yield an oversupply of hits which need further confirmation.

VS identify the false positive hits. Many drug candidates fail the clinical trials due to various reasons. Thus VS evaluates drug likeness based on the following criteria:

a. Simple counting method
b. Functional group filter
c. Topological filter
d. Pharmacophore filter

In virtual screening compounds are selected by predicting their binding to a macromolecular target using computer programs. The compound of interest need not be substantial and their testing need not require any costly materials. On other hand the limitations encountered during experimental conditions related to its physical properties like limited solubility, formation of aggregates or during its assay can be ruled out in VS. It requires profound knowledge about the energetic and spatial criteria needed for the binding of specific candidate ligand to the receptor under investigation. The approach involved in VS in newer than the techniques involved in it. The general templates for VS are to discover particular ligand by docking or through mapping to pharmacophore models.

Success rate in identification of ligand molecule in drug development process is enhanced by Virtual screening. This computational analysis can be classified into two broad categories as ligand based and receptor based. For ligand-based methods, the strategy is to use information provided by a compound or set of compounds that are known to bind to the desired target and to use this to identify other compounds in the corporate database or external databases with similar properties. This can be done by a variety of methods, including similarity and substructure searching [2], pharmacophore matching [3] or 3D shape matching [4].

Receptor based methods can be employed when the structure of the target protein is known. These involve explicit molecular docking of each ligand into the binding site of the target, producing a predicted binding mode for each database compound, together with a measure of the quality of the fit of the compound in the target binding site. This information is then used to rank the compounds with a view to selecting and experimentally testing a small subset for bio-logical activity. Structure-based virtual screening involves a range of sequential computational phases, including target and database preparation, docking and post-docking analysis, and prioritization of compounds for testing.

Database preparation

In the beginning stages of a virtual screening project it is compulsory to construct the database of compounds to be screened. The source of the database typically be in agreement with corporate collection of physically available compounds, or a directory of compounds available externally from chemical vendors [e.g. the MDL® Available Chemicals Directory (ACD) from MDL Information Systems; http://www.mdli. com)]. A supplementary reference that is fitfully considered for virtual screening is an *in silico* virtual library corresponding to compounds constructed from a list of reagents and a database of known chemistries. The constituents of such an *in silico* library are easily synthesizable and could readily be fabricated once they have been chosen from a virtual screen. In the event of a corporate or external collection, the beginning database is concised in size by employing several physical and chemical filters, in an attempt to have a collection of compounds that have physical properties and chemical functionality similar to the majority of known drugs. For *in silico* libraries, these factors should already have been taken into consideration during the intial stages of design of the libraries. Lipinski's rule-of-five [5], which is an empirical set of rules based on molecular weight, lipophilicity and hydrophobicity, that provides a simple profile for orally bioavailable compounds is common filtering protocol for drug-likeness that are applied to databases. Limit on the number of rotatable bonds in the molecules or the polar surface area include the other physical filters [6].

Additional filters are employed to rule out compounds containing specific chemical substructures associated with poor chemical stability or toxicity [7]. All of these filtering protocols are computationally inexpensive and can be applied rapidly to a large database (of the order of 10^5–10^6 compounds).

Target preparation

The development of the active site is dependent on the docking tools being used. Some methods demand the addition of hydrogens and care should be taken to fortify that this is done sensibly, circumventing atomic clashes. The proper protonation states of ionizable residues in the active site need to be decided and the correct tautomer for histidines should be assigned. It is suggested that after the addition of hydrogen atoms to the protein, the positions of the hydrogens are relaxed by energy minimization to prevent any steric clashes that have been added.

2. Molecular docking

Docking each molecule in the database into the binding site of the target is the next phase of the screening process. Sampling the co-ordinate space of the binding site and scoring each possible ligand pose, which is then taken as the predicted binding mode for that compound is what the docking process consists of. A variety of docking softwares are present for use in virtual screening and they vary in the sampling algorithms, the handling of ligand and protein flexibility, the scoring functions they involve, and the cpu time required to dock a molecule to a given target. In Table 1, several commonly used docking softwares are listed along with the types of sampling and scoring algorithms used, and an indication of their suit-ability for large-scale virtual screening.

Table 1: Common docking tools for virtual screening

Method	Ligand flexibility sampling	Scoring function	Suitability for large-scale virtual screening
Dock [29]	Incremental build	Force field or contact score	High
FlexX [28]	Incremental build	Empirical score	High
Slide [30]	Conformational ensembles	Empirical score	High
Fred (Openeye Software)	Conformational ensembles	Gaussian score or empirical scores	High
Gold [25]	Genetic algorithm	Empirical score	Low
Glide (Schrodinger)	Exhaustive search	Empirical score	Low
AutoDock [26]	Genetic algorithm	Force field	Low
LigandFit (Accelrys)	Monte Carlo	Empirical score	Low
ICM [27]	Pseudo-Brownian sampling and local minimization	Mixed force field and empirical Score	Low
QXP [23]	Monte Carlo	Force field	Low

Qualitative assessments of the suitability of the methods to screen 100K compounds on 8 processors within few days. Methods designated as having low suitability for 100Ks of compounds should be considered suitable for 10K compounds under similar conditions, or would require considerably more computer power to be applied to 100Ks of compounds. Openeye Software, http://www.eyesopen.com; Schrodinger, http://www.schrodinger.com; Accelrys, http://www.accelrys.com.

Ligand flexibility

The conformation of a compound bound to a target differs from the conformation of the unbound form in solution. In general, the free state in solution can be thought of as an ensemble of conformations of which a small subset is pertinent to the bound form. It is therefore important to review the conformational flexibility of the compounds during the docking process. This can be attained by pre-computing a database that contains of several conformers of each compound to be screened. There are numerous software packages [Confort, Omega (Openeye Scientific Software; http://www.eyesopen.com) and Catalyst (Accelrys)] that are accessible to attain large conformational data-bases rapidly, and then each conformer can be rigidly docked into the target binding site. Examples of methods that work with configurational unities [8,9] are Slide, Dock and Fred. Alternatively, as the calculation proceeds, through a variety of docking algorithms, the majority of docking programs explore the conformational flexibility of the compounds. Number of rotatable bonds present in the ligand and the size of conformational space for a ligand

are directly related, and efficient algorithms need to be conceived to control the combinatorics of this issue. Numerous computational strategies that are used by current docking methods, and although it is not the intention to detail all the methods available, the major approaches are discussed with some common docking methods mentioned as examples (other specific methods are reviewed elsewhere in the literature [10]).

Target flexibility

An characteristic of the docking problem that has not considered is the flexibility of the protein target. The majority of docking tools currently make the assumption that the protein target is held fixed in its crystal structure conformation. This is usually an approximation but is normally a necessary one because of the increased complexity, and the computational cost, that is necessary to accurately sample the flexibility of the binding site. There are some programs that attempt to take into account protein flexibility to a certain degree, such as Slide [11], which enables the motion and relaxation of binding-site side chains in response to the presence of a docked ligand.

The efforts to include protein flexibility try to make use of an ensemble of protein structures, which can be obtained from several high-resolution structures of the same target bound to different ligands, or in the apo form. An alternative source of entities of target structures could be from an NMR structural study of the target. The conformational ensemble could be derived from a molecular dynamics simulation in the absence of experimental information.

Scoring

A pose has been deduced for a compound in the binding site and it needs to be scored to rank the quality of the pose in comparison to other poses for the compound, and with respect to other compounds within the database. There are huge choice of scoring functions available [12], and they can be classified as being physical-based (force-field), empirical or knowledge-based. Physical-based scoring functions like Amber [13] or CHARMM [14] are based on atomic force fields. Free energy perturbation (FEP) or thermodynamic integration (TI) methods [15], when employed with force-fields are normally unambiguous at determining binding free energies, Because of the vast amount of sampling required and the inclusion of explicit solvent makes the computational time associated with FEP and TI methods is prohibitive for the purposes of molecular docking.

3. Virtual Screening Tools

Py Rx

It is a virtual screening software used by medicinal chemists worldwide to screen libraries of compounds with pharmacological interests. It can be run from any platform. The important feature of this software includes docking wizard which enables Computer-aided drug design. Also, rational drug design is achieved by chemical spreadsheet and efficient visualization engine.

Py Rx is using large body of established open source software such as:

- AutoDock 4 and AutoDockVina are used as a docking software.
- AutoDockTools, used to generate input files.
- Python as a programming/scripting language.
- wxPython for cross-platform GUI.
- The Visualization ToolKit (VTK) by Kitware, Inc.
- Enthought Tool Suite, including Traits, for application building blocks.
- Opal Toolkit for running AutoDock remotely using web services.
- Open Babel for importing SDF files, removing salts and energy minimization.
- matplotlib for 2D plotting

A potent anticancer target of danshensu was identified by inverse docking with the help of this VS tool. PharmMapper and idTarget servers were used as tools, and the results were checked with the molecular docking program autodockvina in PyRx 0.8. The docking studies confirmed that the target drug –complex matched pretty well with respect to structure, H bonds and hydrophobicity. Dual inverse docking proved efficient identification of potent anticancer targets in Dhanshensu. Another example of this software application is the Molecular Docking studies of Aromatase Inhibitors. Aromatase is an enzyme plays a vital role in developing estrogen receptor positive breast cancer. It brings about the aromatization of androstenedione to estrone. Cytochrome P450 19A1 is the scientific name for Aromatase .The undersirable effects of Aromatase Inhibitors(AIs) demand the discovery of novel AIs with less toxicity and other pharmacological side effects. The docking protocol followed in this study enabled successful interaction of aromatase with its substrate at 1.350 Amstrong. The docking studies also confirmed that polar and nonpolar residue were essential for interaction with the AIs.

Lead Finder

This is yet another commonly used docking tools for Computer aided drug design. It screens the libraries of chemical compounds to find potent binders for the target protein. The basic principle lies in the fact that each ligand from a library is docked to a target protein. Ligands are then ranked according to their binding potencies. This software makes use of a special type of scoring function called VS score. This function helps in differentiating active compounds from the inactive ones. The docking algorithm is called screening mode. This increases the efficiency of the software to multifold when compared to the default docking mode. Leading Finder is also employed to predict the 3D structure of covalent and noncovalently bound protein –ligand complex. Free energy of protein-ligand complex has been reported successfully. This feature is unique to Lead Finder.

a. Lead Finder enables preparation of protein structure models for docking by following ways :
b. Addition of hydrogen atoms to protein's heavy atoms according to optimal ionization states of protein residues at a given pH;
c. Optimization of polar hydrogen positions with respect to the ligand, substrate and cofactor present in the structure;
d. Optimization of side chain orientations of His, Asn and Gln residues for which X-ray analysis
e. can return flipped orientations due to apparent symmetry.

Mol Ware:

This software was discovered in 2004. It was designed for Pattern Recognition and Data Analysis. It also finds applications in green chemistry. In a Virtual Screening, a set of molecular structures is subjected to a computational analysis to reveal potential bioactive compounds. Virtual Screening requires, as a fundamental tool, a theoretical model able to evaluate the possibility of pharmacological activity of the compounds represented in a virtual library. It results in a filtered set of chemical structures ready to be tested in vitro or in vivo with high probability of success. Virtual Screening can be done over a catalogue of chemical structures or a client library, to search potential activities of the structures in the set. The result is a labelled library. The Virtual Screening concept is also applied to other fields such as catalysts discovery

4. References

1. Myers, S. and Baker, A. Drug discovery – an operating model for a new era. Nat. Biotechnol. **2001**, 19, pp.727–730.
2. Mestres, J. and Knegtel, R.M.A. Similarity versus docking in 3D virtual screening. Perspect. Drug Des. Discovery, **2000**, 20, pp. 191–207
3. Mason, J.S. et al. 3-D pharmacophores in drug discovery. Curr. Pharm. Des., **2001**, 7: 567–597
4. Srinivasan, J. et al. Evaluation of a novel shape-based computational filter for lead evolution: Application to thrombin inhibitors. J. Med. Chem., **2002**, 45: 2494–2500
5. Halperin, I. et al. Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins Struct. Funct. Genet., **2002**, 47, pp. 409–443.
6. Lipinski, C.A. et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Discov. Res., **1997**, 23: 3–25
7. Palm, K. et al. Correlation of drug absorption with molecular surface properties. J. Pharm. Sci., **1996**, 85, 32–39.
8. Hendlich, M. Databases for protein–ligand complexes. Acta Crystallogr. D54, **1998**, pp.1178–1182
9. Hooft, R.W.W. et al.Errors in protein structures. Nature, **1996**, 381, pp.272
10. Kearsley, S.K. et al. Flexibases – a way to enhance the use of molecular docking methods. J. Comput.-Aided Mol. Design, **1994**, 8: 565–582
11. Lomber, D.M. and Shoichet, B.K. Flexible ligand docking using configurational ensembles. Proteins Struct. Funct. Genet. **1998**, 7: 938–950
12. Schnecke, V. and Kuhn, L.A. Virtual screening with solvation and ligand-induced complementarity. Perspect. Drug Des. Discov. **2000**, 20:171–190
13. Ajay, A. and Murcko, M.A. Computational methods to predict binding free energy in ligand-receptor complexes. J. Med. Chem., **1995**, 38: 4953–4967
14. Cornell, W.D. et al. A second generation force-field for the simulation of proteins, nucleic acids and organic molecules. J. Am. Chem. Soc., **1995**, 117: 5179–5197
15. MacKerell, A.D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B, **1998**, 102, 3586–3616
16. Kollman, P.A. Free energy calculations: application to chemical and biochemical phenomena. Chem. Rev., **1993**, 93: 2395–2417.
17. Paul D. Lyne, Structure based Virtual Screening: an Overview; DDT. **October 2002**, 7(20).