Methods 131 (2017) 10–21

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Function and structure-based screening of compounds, peptides and proteins to identify drug candidates

Vidhi Malik, Jaspreet Kaur Dhanjal, Anjani Kumari, Navaneethan Radhakrishnan, Kamya Singh, Durai Sundar *

Department of Biochemical Engineering and Biotechnology, DBT-AIST International Laboratory for Advanced Biomedicine (DAILAB), Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

ARTICLE INFO

Article history:
Received 14 June 2017
Received in revised form 21 August 2017
Accepted 21 August 2017
Available online 24 August 2017

Keywords:
High throughput virtual screening
Inverse virtual screening
Molecular docking
Molecular dynamic simulations
Pharmacophore
QSAR

ABSTRACT

Drug discovery in simple words is all about finding small molecular compounds that possess the potential to interact with specific bio-macromolecules, mainly proteins, thereby bringing a desired effect in the functioning of the target molecules. Virtual screening of large compound libraries using computational approaches has come up as a great alternative to cost and labor-intensive high-throughput screening carried out in laboratories. Virtual high-throughput screening enormously reduces the number of compounds for systematic analysis using biochemical assays before entering the clinical trials. Here, we first give a brief overview of the rationale behind virtual screening, types of virtual screening – structure-based, ligand-based and inverse virtual screening, and challenges that need to be addressed to improve the existing strategies. Subsequently, we describe the methodology adopted for virtual screening of small molecules, peptides and proteins. Finally, we use few case studies to provide a better insight to the application of computer-aided high-throughput screening.

1. Introduction

Introduction of a new therapeutic intervention into the market has remained an expensive and time-consuming process. A typical drug discovery cycle starting from the identification of the active molecule, structural optimization to get a lead and its entrance into clinical trials is estimated to take around 14 years with an expenditure of around 800 million US dollars [1,2]. In the early 1990s, advancement in the field of combinatorial chemistry and advent of high-throughput screening (HTS) technologies improved the conventional process of drug design and discovery by expediting and automating the screening process of thousands of compounds simultaneously. Although, there was a drastic decrease in the human efforts associated with the screening processes, HTS is still an expensive approach requiring a lot of resources that include the cost of synthesis of targets and ligands being screened. Also, the hit rates are often low and not many of these identified hits are able to develop into actual leads and enter into preclinical studies [3]. So, the pharmaceutical companies are now focusing on alternative strategies that help in avoiding the wastage of resources on ligands with no possible future success. Computer-aided drug discovery tool like virtual screening is therefore gaining popularity as a complementary approach to HTS in the pharmaceutical industry and academic research [4].

Virtual screening is a computational approach with a basic aim to reduce the enormous virtual chemical space of small organic compounds, peptides or protein molecules to be synthesized or experimentally screened against a specific molecular target. The huge number of molecules thus gets limited to a few for systematic assessment, synthesis and testing; all having a high probability of leading to a potential drug candidate. It therefore requires digital repositories for the study of chemical interaction relationships [5]. One can also explore the derivatives of known compounds that may not exist physically but possess interesting physicochemical characteristics, which when synthesized can improve the therapeutic activity of the active compound [3].

Virtual high-throughput screening not only reduces the costs associated with drug discovery by selecting the best possible lead molecules from millions of compounds that finally enter animal studies, but also helps in bringing down the time taken for a drug

Abbreviations: CoMFA, Comparative Molecular Field Analysis; DAIM, Decomposition and Identification of Molecules; HTS, high-throughput screening; QSAR, Quantitative structure–activity relationship; VDA, Vascular-disrupting agents.

* Corresponding author.
E-mail addresses: bez148267@iitd.ac.in (V. Malik), bez138509@iitd.ac.in (J.K. Dhanjal), bey157502@iitd.ac.in (A. Kumari), ird11463@iitd.ac.in (N. Radhakrishnan), ird11270@iitd.ac.in (K. Singh), sundar@dbeb.iitd.ac.in (D. Sundar).

http://dx.doi.org/10.1016/j.ymeth.2017.08.010
1046-2023/

to reach the market. The following example shows how computer aided virtual screening acts as a 'virtual shortcut' in the drug discovery pipeline. In 2003, two research groups at Eli Lilly and Biogen Idec were in search for novel transforming growth factor-b1 receptor kinase inhibitors. The group at Eli Lilly used the regular HTS and identified a lead compound, which was further optimized and improved using structure-activity relationship assays. On the other hand, the group at Biogen Idec used virtual HTS to shortlist 87 hits that showed structural interactions with transforming growth factor b1 receptor kinase *in silico*. It was observed that the best hit obtained by Biogen Idec was structurally identical to the lead compound discovered at Eli Lilly through the traditional HTS approach [6]. So, computer aided drug designing was able to yield the same results at much lower cost and workload.

2. Concepts and challenges

2.1. Concepts of virtual screening

Virtual screening is broadly categorized into structure-based and ligand-based virtual screening. A newer approach of computational high-throughput screening called inverse virtual screening or reverse docking has also been designed by the researchers to overcome some of the limitations associated with *in silico* prediction of drug targets.

2.1.1. Structure-based virtual screening

This method of virtual screening is used when the structure of the macromolecule to be targeted is available. The structure for the targets solved experimentally by X-ray crystallography or Nuclear magnetic resonance can be obtained from various structural databases like Protein Data Bank [7]. In case the experimentally solved structure is not available, it is possible to model a three dimensional structure for the protein of interest *in silico* using 3-D coordinates of other homologous proteins as template. With the rationale that the functional activity of the proteins reside within a specific pocket, knowing the structure makes it possible to use structure-based high-throughput virtual screening or direct docking methods to screen the ligands, small molecules or peptides that fit into the binding cavity within the target proteins. The resulting molecules are ranked based on their affinity for the target molecule evaluated in terms of binding free energies. The hits obtained are then tested *in vitro* to get an estimate of their biological activity. The structure of the obtained lead molecule can further be optimized by substituting various functional groups for altering the affinity and improving the biological response of the candidate molecule [4]. The use of natural peptides for therapeutic purposes is not very common because of the problems like instability due to proteolytic cleavage and poor bioavailability. Attempt has been made to overcome these difficulties by screening for peptidomimetics. These are the compounds in which the basic chemical skeleton mimics the natural peptide or protein in 3D space. These compounds can interact with the target molecule in the way similar to the natural peptides and also produce the same effect [8]. Knowing the structure of the biological molecule to be targeted, virtual screening can easily help in selection of potential peptidomimetics from a huge library of known compounds.

2.1.2. Ligand-based virtual screening

When the experimentally solved three dimensional structure of the target macromolecule is not available and prediction of the structure through computational methods like homology modeling or *ab initio* structure prediction is also challenging, ligand-based virtual screening becomes a good alternative to structure-based approach [9,10]. This approach relies on the knowledge of a set of small molecular compounds that actively bind to the target site within the protein of interest. Molecular structural similarity approaches, screening against compound libraries using pharmacophores and Quantitative structure–activity relationship (QSAR) models are some of the popular approaches that come under ligand-based virtual screening [11]. In molecular similarity methods, the molecular fingerprint of ligands known to bind a specific target is generated and screened against virtual libraries for finding new molecules with similar molecular fingerprint. The new hits obtained generally have a structural scaffold similar to the already known molecules. Pharmacophore is a virtual ensemble of steric and electronic features derived from a set of active ligands, contributing to the interactions between the small molecules and the target protein. Molecules with a pattern of functional groups possessing the same physiochemical properties as in the pharmacophore, are screened out from millions of compounds and analyzed further. The advantage of using pharmacophore modeling approach is that the molecules obtained are diverse in structure but similar in function. QSAR models can help in drawing a relationship between structural features of ligands that bind to a target and their corresponding biological activity. This approach is quite useful for optimization of lead molecules to get active compounds with biological activity within nanomolar range [4].

2.1.3. Inverse virtual screening

Recently, a new method for virtual screening has been designed that screens a set of compounds against a database of macromolecules, mainly proteins. It is a reverse setup compared to normal virtual screening. This approach of virtual screening is called *inverse virtual screening* [12]. It is useful for drug repositioning or repurposing [13]. New therapeutic potential of already known drugs can be explored. It can be used to study the effect of a single molecule on various protein targets involved in a particular biochemical pathway. It can also be used to answer the questions related to side effects associated with the usage of a compound. A single drug can also be designed against multiple targets to have synergistic therapeutic effect [12]. For example, using the approach of inverse virtual screening against a drug target database, *H. pylori* Peptide deformylase was reported to be an antibiosis target of an active natural product. This result was further validated by obtaining a crystal structure [14]. Till 1980's, the drug called Tofisopam was used to treat anxiety. Using the reverse docking software, *SELNERGY*, Tofisopam was found to inhibit phosphodiesterase-4 as well [15]. This example illustrates the idea of drug repurposing. In another study, researchers were able to predict 83% of the experimentally known toxicity and side effects associated with 8 clinical agents, namely aspirin, gentamicin, ibuprofen, indinavir, neomycin, penicillin G, 4H-tamoxifen, and vitamin-C using inverse virtual screening [16].

2.2. Scoring functions

Virtual screening of number of structurally diverse compounds will result in large number of docked complexes. Hence, assessment and ranking of these predicted molecular interactions become a crucial step for drug-target identification. This can be performed using scoring functions. Scoring functions are heuristic mathematical methods developed for evaluation of protein–ligand interactions. Using scoring functions, a numerical value is derived for each protein–ligand binding conformation that aids in comparison of binding affinities of different ligands and their different poses in the active site of target macromolecule. This helps in relative ranking of ligands with reference to the target and thereby identification of the potential drug candidates. Scoring functions are used in virtual screening for hit identification as well as for lead optimization [17,18].