Methods 71 (2015) 26–37

Contents lists available at ScienceDirect

Methods

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

Hierarchical virtual screening approaches in small molecule drug discovery

Ashutosh Kumar, Kam Y.J. Zhang *

Structural Bioinformatics Team, Center for Life Science Technologies, RIKEN, 1-7-22 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan

ARTICLE INFO

Article history:
Received 29 May 2014
Revised 16 July 2014
Accepted 17 July 2014
Available online 27 July 2014

Keywords:
Hierarchical virtual screening
Molecular docking
Similarity search
Pharmacophore modeling

ABSTRACT

Virtual screening has played a significant role in the discovery of small molecule inhibitors of therapeutic targets in last two decades. Various ligand and structure-based virtual screening approaches are employed to identify small molecule ligands for proteins of interest. These approaches are often combined in either hierarchical or parallel manner to take advantage of the strength and avoid the limitations associated with individual methods. Hierarchical combination of ligand and structure-based virtual screening approaches has received noteworthy success in numerous drug discovery campaigns. In hierarchical virtual screening, several filters using ligand and structure-based approaches are sequentially applied to reduce a large screening library to a number small enough for experimental testing. In this review, we focus on different hierarchical virtual screening strategies and their application in the discovery of small molecule modulators of important drug targets. Several virtual screening studies are discussed to demonstrate the successful application of hierarchical virtual screening in small molecule drug discovery.

1. Introduction

Modern drug discovery process starts with the identification of initial hits that are further optimized to improve the potency, selectivity, metabolic stability and oral bioavailability. Among many ways of identifying initial hits in drug discovery, high-throughput screening (HTS) and virtual screening (VS) are most common. The VS was originally developed to bring down the cost of discovering new molecules using HTS. In the last two decades, advances in computational programs and processing power have made VS an important tool to identify starting points, inhibitors and chemical probes in various drug discovery campaigns [1–4].

In light of the immense potential of VS methodologies in the identification of initial hits, various structure and ligand-based approaches were developed [1,5–8]. Structure-based methods rely on the structural information of the protein target and typically include methods such as molecular docking [9,10], structure-based pharmacophores [11,12] and *de novo* design [13,14]. Ligand-based methods can work in the absence of structural information for the protein target. These methods include two or three-dimensional (2D or 3D) similarity searches [15,16], ligand-based pharmacophore screenings [5,17], machine learning approaches [18,19], quantitative structure activity relationships (QSAR) [5,20] among others. Ligand-based methods, however, require the availability of at least one known active molecule. Although utilizing these structure and ligand-based methods individually have demonstrated immense potential in retrieving initial hits, these methods are unable to fulfill all the practical requirements of drug discovery alone. Furthermore, with ever-increasing screening library size [21] and computational cost associated with some VS approaches especially flexible molecular docking [22–25], it is indispensable to integrate different VS approaches to filter compounds. Ligand and structure-based methods can be combined in a sequential or parallel manner (Fig. 1). The most common way of combining these methods is to use them in a sequential funnel like manner commonly known as hierarchical VS (HLVS). In HLVS, a large small molecule library is reduced to a number of compounds that is small enough for biological assay by applying a series of filters (generally two or three) sequentially (Fig. 1A). In contrast to HLVS, there is parallel virtual screening (PVS) where several complementary methods are run in parallel and the best hits ranked according to each method are selected for biological testing (Fig. 1B). Although the retrospective analysis of literature data has shown the successful application of PVS [26–30], only a few applications in real world scenario could be found [27,31,32].

In this paper, we review the current status of commonly used HLVS approaches and try to understand their utility in drug

* Corresponding author. Fax: +81 45 503 9559.
E-mail address: kamzhang@riken.jp (K.Y.J. Zhang).

http://dx.doi.org/10.1016/j.ymeth.2014.07.007
1046-2023/

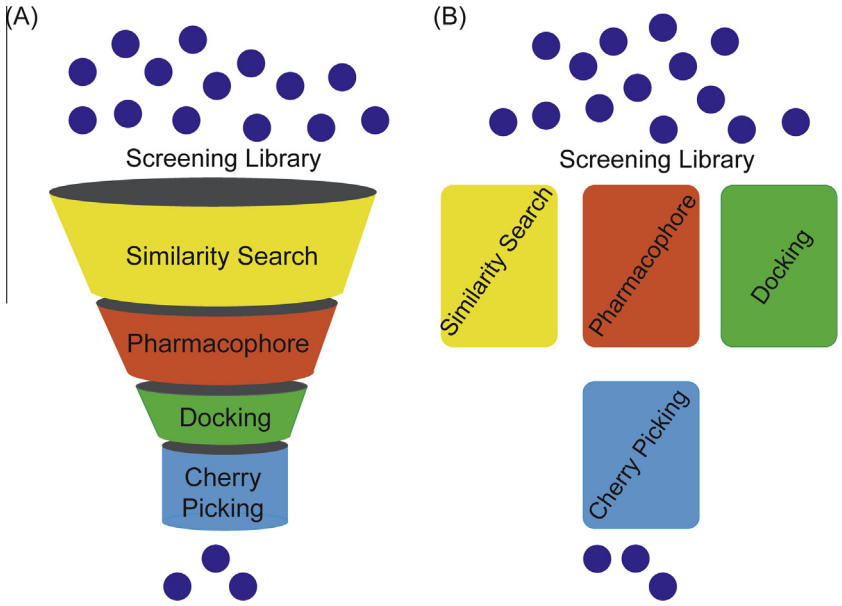

Fig. 1. Integration of ligand and structure-based approaches. (A) Hierarchical virtual screening (HLVS): series of filters (here similarity search, pharmacophore and molecular docking) are sequentially applied to bring down the number of compounds to be cherry-picked for biological assay. (B) Parallel virtual screening (PVS): ligand and structure-based filters are performed independently on the same or similar number of compounds.

discovery campaigns of important therapeutic targets. Although the scientific literature is inundated by hierarchical computational approaches and protocols, we have restricted our review on only those studies that were validated by experimental assays. In the following sections we will outline the types of HLVS approaches utilized in various small molecule discovery campaigns. Later, we will discuss recent cases of successful hit identification utilizing HLVS protocols. Finally, we will describe the usefulness of HLVS in discovering inhibitors of important drug targets.

2. Hierarchical combination of VS methods

Hierarchical combination of ligand and structure-based VS approaches generally involves sequential execution of dissimilar VS methods. Mostly, computationally inexpensive ligand based approaches such as similarity search and pharmacophore screening are used during initial steps of an HLVS protocol. Methods demanding comparatively high computational resources such as molecular docking and molecular dynamics (MD) simulation are used once the number of compounds to be screened decreases to a reasonable number. The final step in a majority of HLVS campaigns incorporates the visual selection of compounds by expert researcher commonly known as "cherry picking". In this step, ranking from VS methods is combined with expert chemical intuition and with literature-based knowledge. The HLVS can be classified in three categories based on the combination of VS methods: ligand-based HLVS, structure-based HLVS and hybrid HLVS, which will be described in detail below. A few successful cases of small molecule discovery using these three classes of HLVS are summarized in Table 1.

2.1. Ligand based HLVS (LB-HLVS)

LB-HLVS sequentially combines methods based on similarity search or compound classification. Similarity based techniques include methods accessing the similarity of one or a few experimentally identified hits with molecules in a large library in terms of their physicochemical properties [33], structural fingerprints [34], 3D-shape [35], electrostatic potential [36] and pharmacophore features [37] etc. Compound classification techniques include clustering [38,39] or machine learning based methods such as Bayesian methods and support vector machines [18,40]. Although most of the ligand-based VS methods are used either standalone or in combination with structure-based VS methods, only ligand-based methods were reported to have been effectively combined in prospective applications. Yao et al. [41] reported an efficient multi-step ligand-based VS protocol that included physicochemical property filtering, pharmacophore-based screening, protein–ligand interaction fingerprint similarity analysis and 2D-fingerprint structural similarity search. Their protocol significantly improved the hit rate when compared with individual methods. Among the prospective applications, LB-HLVS that included a combination of shape-based VS and pharmacophore modeling has been used by Temml et al. [42] to identify two agonist of liver X receptor. The reported agonist activated both subtypes of liver X receptor (LXR α and β). Shape similarity has also been combined with electrostatic potential matching in the discovery of melanin-concentrating hormone receptor 1 antagonist [43], *Francisella tularensis* enoyl-reductase inhibitors [44] and chemical probe for nicotinic acid adenine dinucleotide phosphate (NAADP) [45]. In another application [46], ZINC database [47] subset enriched with quinoxaline scaffold was filtered based on pharmacokinetic properties. The resulting molecules were again investigated for pharmacophore fingerprint similarity with known quinoxaline based inhibitors of folate cycle proteins. Associated biological assay resulted in three compounds, which interfered with dihydrofolate reductase and thymidylate synthase and reduced their levels. Levit et al. [48] integrated 1D molecular descriptors, 2D fingerprint-based molecular similarity, ligand-based pharmacophore models and a shape-based VS method to identify activators of human bitter taste receptor TAS2R14. Bayesian analysis has been used with pharmacophore modeling to identify inhibitors of breast cancer resistance protein (BCRP) [49]. VS was carried out against 2000 FDA-approved drugs and 19 drugs were found to exhibit significant effect on BCRP transport function. Another machine learning technique, support vector machine (SVM) in combination with fingerprint similarity search was also