European Journal of Medicinal Chemistry 83 (2014) 226–235

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Ligand efficiency based approach for efficient virtual screening of compound libraries

Yi-Yu Ke [a, *], Mohane Selvaraj Coumar [b], Hui-Yi Shiao [a], Wen-Chieh Wang [a], Chieh-Wen Chen [a], Jen-Shin Song [a], Chun-Hwa Chen [a], Wen-Hsing Lin [a], Szu-Huei Wu [a], John T.A. Hsu [a], Chung-Ming Chang [a], Hsing-Pang Hsieh [a, *]

[a] Institute of Biotechnology and Pharmaceutical Research, National Health Research Institutes, 35 Keyan Road, Zhunan, Miaoli County 350, Taiwan, ROC
[b] Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Kalapet, Puducherry 605014, India

ARTICLE INFO

Article history:
Received 25 February 2014
Received in revised form
4 June 2014
Accepted 13 June 2014
Available online 14 June 2014

Keywords:
Ligand efficiency
Fit quality
Virtual screening
Pharmacophore model
Aurora kinase inhibitor

ABSTRACT

Here we report for the first time the use of fit quality (FQ), a ligand efficiency (LE) based measure for virtual screening (VS) of compound libraries. The LE based VS protocol was used to screen an in-house database of 125,000 compounds to identify aurora kinase A inhibitors. First, 20 known aurora kinase inhibitors were docked to aurora kinase A crystal structure (PDB ID: 2W1C); and the conformations of docked ligand were used to create a pharmacophore (PH) model. The PH model was used to screen the database compounds, and rank (PH rank) them based on the predicted IC_{50} values. Next, LE_Scale, a weight-dependant LE function, was derived from 294 known aurora kinase inhibitors. Using the fit quality (FQ = LE/LE_Scale) score derived from the LE_Scale function, the database compounds were reranked (PH_FQ rank) and the top 151 (0.12% of database) compounds were assessed for aurora kinase A inhibition biochemically. This VS protocol led to the identification of 7 novel hits, with compound **5** showing aurora kinase A IC_{50} = 1.29 μM. Furthermore, testing of **5** against a panel of 31 kinase reveals that it is selective toward aurora kinase A & B, with <50% inhibition for other kinases at 10 μM concentrations and is a suitable candidate for further development. Incorporation of FQ score in the VS protocol not only helped identify a novel aurora kinase inhibitor, **5**, but also increased the hit rate of the VS protocol by improving the enrichment factor (EF) for FQ based screening (EF = 828), compared to PH based screening (EF = 237) alone. The LE based VS protocol disclosed here could be applied to other targets for hit identification in an efficient manner.

1. Introduction

Using the advancements in parallel synthesis technology of the late 1980s and early 1990s, several pharmaceutical companies set up high-throughput screening (HTS) of large compound libraries in order to identify leads for drug development. One of the main disadvantages of HTS is the need to screen a large collection of compounds, which is both time and resource intensive. Recently, virtual screening (VS) is routinely used as a supplement to or as a replacement for HTS in lead identification, due to virtual screening's savings in time and money. VS is a computational method used to score and prioritize HTS library compounds for further biochemical testing [1,2]. Use of VS can increase the hit rates of lead identification, with some studies reporting hit rates ten times greater or more for VS compared to HTS [3].

Several success stories of lead identification using VS are reported in the literature [3,4], using either structure-based virtual screening (SBVS) [5] or ligand-based virtual screening (LBVS) [6]. In SBVS, molecular docking of the compounds with the target protein is performed and a score is assigned based on the interaction of the target protein and the compounds; the score of the compounds is used to rank and prioritize them for testing. However, current docking methods have deficiencies in both correctly identifying the crystallographic conformation of the ligands and also accurately

Abbreviations: 3D-QSAR, three dimensional quantitative structure–activity relationship; EF, enrichment factor; FBDD, fragment-based drug discovery; FQ, fit quality; HA, number of heavy atoms; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; HTS, high through-put screening; HY, hydrophobic; LBVS, ligand-based virtual screening; LE, ligand efficiency; PH, pharmacophore model; ROC, receiver operating characteristic; SBVS, structure-based virtual screening; VS, virtual screening.

* Corresponding authors.
E-mail addresses: yiyuke@nhri.org.tw (Y.-Y. Ke), hphsieh@nhri.org.tw (H.-P. Hsieh).

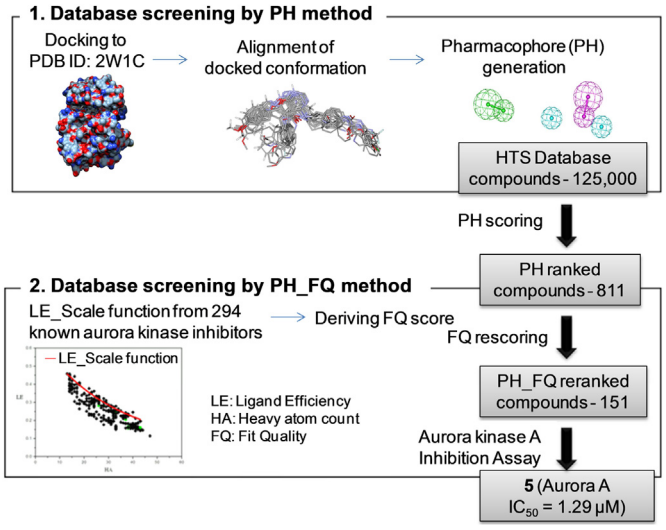
http://dx.doi.org/10.1016/j.ejmech.2014.06.029
0223-5234/

scoring the docked poses [7–9]. More critically, availability of target protein information (either 3D X-ray crystallographic/NMR solved structure or amino acid sequence of the target protein) is essential for SBVS. Hence, in the absence of such protein information, LBVS (an alternative method for VS) is used to identify leads for drug-development projects.

In LBVS, ligand similarities based on either 2D or 3D information of known ligands are used to identify new leads [6]. Particularly, 3D-QSAR (quantitative structure–activity relationship) based techniques — such as CoMFA (comparative molecular field analysis) [10,11], CoMSIA (comparative molecular similarity indices) [12], and pharmacophore models [13] — are currently much used. A pharmacophore model represents the relative orientation of the functional groups of a molecule in three-dimensional space, which are necessary for maintaining the activity. A typical pharmacophore model constitutes various pharmacophoric elements such as the hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic (HY) group, and excluded volume (EV); the relative positions of the pharmacophoric elements in the model are defined by the distances and angles between them. Pharmacophore models can be generated from a set of known active ligands, by superimposition of the bioactive conformation of the ligands. Recently, several groups have reported the use of pharmacophore models for carrying out VS to identify hits [13].

Hits identified from VS, either from SBVS or LBVS methods, are used as a starting point for lead optimization in drug discovery programs. Retrospective analysis of the starting lead – drug pairs from several successful drug discovery programs reveals that the molecular properties (molecular weight and lipophilicity) inflation is very common during lead optimization; hence there is a need to keep tight control over them during the process of drug evolution [14]. Moreover, the success of developing a drug depends heavily on the nature of the starting hits chosen for lead optimization. Simple criteria such as activity alone are insufficient for selecting an appropriate hit for optimization; additional criteria such as Ligand efficiency (LE) need to be considered to select the right hit candidates for further optimization [15,16]. LE can be defined as the binding affinity per heavy atom of the molecule and can be computed from experimental or calculated binding affinity divided by the number of heavy atoms. LE was first introduced in the field of fragment-based drug discovery (FBDD) to select the best of two fragments and is now routinely used to choose the optimized leads during hit to drug evolution in drug discovery programs [17]. It should also be noted that the application of LE has its own limitation as it considers all heavy atoms the same even though oxygen, nitrogen, and other important heavy atoms present in drugs have different physico-chemical and binding properties. This limitation of LE is recognized in the field of drug design; accordingly, the use of the LE function in conjunction with other parameters such as LLE (lipophilic ligand efficiency) during hit selection/optimization in drug discovery programs is recommended [18–20].

The aim of this work is to develop an efficient VS protocol for the improved identification of hits from an HTS library; the VS protocol is exemplified by aurora kinase inhibitor identification. For this purpose, 125,000 compounds in an in-house HTS library were prioritized for biochemical screening using a two-step procedure (Scheme 1). First, HTS compounds were screened by a pharmacophore model developed using two sets of aurora kinase inhibitors (pyrazoles & furanopyrimidines). The model predicts the IC_{50} of the HTS compounds by matching them to the pharmacophore features of the model and ranks (PH_rank) them based on their predicted activity. In the second step, compounds identified using the pharmacophore model were reranked (PH_FQ rank) using Fit quality (FQ) score, a LE based function. Finally, the top ranked compounds were subjected to biochemical testing for aurora kinase A inhibition. The VS protocol reported here has led to the identification of a potent aurora kinase A inhibitor with an $IC_{50} = 1.29$ μM, by testing only 151 compounds from a HTS library of 125,000 compounds. Moreover, incorporation of a LE based scoring function into the VS method has improved the hit rate; such a protocol could be applied to other targets as well.



Scheme 1. Flow chart depicting the Ligand Efficiency (LE) based approach for virtual screening of HTS database for aurora kinase inhibitor identification.

2. Chemistry

The aurora kinase inhibitor **5** identified in this study was synthesized from commercially available methyl-5-formyl-2-hydroxybenzoate (**8**) and *p*-nitro-aniline (**14**) using a convergent synthesis protocol as shown in Scheme 2. For this purpose, aldehyde **8** was subjected to bromination under acid condition, as reported earlier, to give **9** in quantitative yield [21]. Treatment of **9** with the Grignard reagent – 2-chlorophenylmagnesium bromide in THF at 0 °C provided the secondary alcohol **10** in 65% yield. Acid mediated dehydroxylation of **10** followed by LAH reduction resulted in the primary alcohol **12** in 83% yields, over two steps. The key aldehyde intermediate **13** was prepared from **12** by MnO_2 oxidation, in 77% yield. Next, the aniline intermediate **16** was synthesized from the aniline **14**, by first coupling with 2-furoyl chloride under basic condition to give **15** in 74% yield, followed by hydrogenolysis under H_2 atmosphere over 10% Pd/C in 95% yield. Finally, both intermediates, **13** and **16**, were condensed under reflux conditions to give the desired compound **5** in 52% yield.

3. Results and discussion

3.1. Pharmacophore model generation and validation

As a first step, we set out to develop an LBVS platform using pharmacophore model for screening the in-house HTS library to identify aurora kinase inhibitors. It is known that 3D pharmacophore based VS is much faster than SBVS. Moreover, pharmacophore-based VS results in the identification of much more diverse chemotypes and is more useful in scaffold-hopping [13]. Albeit these advantages of LBVS, use of SBVS is often much higher than the use of LBVS (pharmacophore) methods for hit identification, due to certain limitations [4]. For instance, the accuracy of the pharmacophore model generated from a set of known