Journal of Molecular Graphics and Modelling 28 (2010) 775–787

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

PhDD: A new pharmacophore-based *de novo* design method of drug-like molecules combined with assessment of synthetic accessibility

Qi Huang, Lin-Li Li, Sheng-Yong Yang *

State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, West China Medical School, Sichuan University, #1 Keyuan Road 4, Chengdu, Sichuan 610041, China

ARTICLE INFO

Article history:
Received 2 August 2009
Received in revised form 25 January 2010
Accepted 7 February 2010
Available online 11 February 2010

Keywords:
Pharmacophore model
De novo design
Assessment of synthetic accessibility
Drug-like molecule

ABSTRACT

This account describes a new pharmacophore-based *de novo* design method of drug-like molecules (PhDD). The method PhDD first generates a set of new molecules that completely conform to the requirements of a given pharmacophore model, followed by a series of assessments to the generated molecules, including assessments of drug-likeness, bioactivity, and synthetic accessibility. PhDD is tested on three typical examples, namely, pharmacophore hypotheses of histone deacetylase (HDAC), cyclin-dependent kinase 2 (CDK2) and HIV-1 integrase (IN) inhibitors. The test results demonstrate that PhDD is able to generate molecules with novel structures but having similar biological functions with existing inhibitors. The validity of PhDD together with its ability of assessing synthetic accessibility makes it a useful tool in rational drug design.

1. Introduction

De novo design methods of drug molecules have increasingly attracted much attention in recent years since they can produce completely novel chemical structures with desired pharmacological properties. So far a considerable number of *de novo* design methods have been developed, including HSITE/2D Skeletons [1], 3D Skeletons [2], LEGEND [3], LUDI [4], NEWLEAD [5], CONCEPTS [6], SPROUT [7], MCSS&HOOK [8], SMoG [9], CONCERTS [10], LEA [11], LigBuilder [12], TOPAS [13], F-DycoBlock [14], ADAPT [15], SYNOPSIS [16], CoG [17], BREED [18], etc. (for a review, see Ref. [19]). These methods have already been used in drug discovery and some of them have shown a good performance. However, all of these methods except NEWLEAD, which will be mentioned later, adopt receptor-based strategy. This strategy is difficult to process the cases in which the receptors or their structures are unknown. On the other hand, very few methods consider the synthetic accessibility of the designed compounds, which has been thought as the most challenging problem in *de novo* molecule design. Other problems which still have no optimal solution in the known methods include (1) how to sample the molecular search space effectively, and (2) how to evaluate the potency of designed molecules [19].

NEWLEAD, proposed by Tschinke and Cohen [5], is the first pharmacophore-based *de novo* design method, which has ever played an important role in promoting the development of *de novo* design methods. It uses as input a set of disconnected molecule fragments that are consistent with a pharmacophore model. Then, it tries to link the disconnected fragments with spacers (such as atoms, chains, or ring moieties). Actually the NEWLEAD can only process the cases that the pharmacophore features are concrete functional groups (not abstract chemical features, like hydrogen bond acceptor, hydrogen bond donor, hydrophobic feature). Additionally, sterically forbidden regions of the receptor binding site are not considered in NEWLEAD.

In order to overcome problems mentioned above, we developed a new pharmacophore-based *de novo* design method of drug-like molecules, called PhDD (a *Ph*armacophore-based *De Novo D*esign *M*ethod). The method PhDD has the following characteristics that are distinct from the commonly used receptor-based *de novo* design methods and NEWLEAD: (1) PhDD works with abstract pharmacophore models. The pharmacophore models might be established by using receptor-based methods, such as LigandScout [20], Pocket [21], or ligand based methods, such as Catalyst (Accelrys Inc., USA), Galahad [22], GASP [23], and DISCO [24]. Further, PhDD also works with pharmacophore models with excluded volumes involved. (2) PhDD incorporates with the assessment of synthetic accessibility of the designed molecules. (3) Fragments as well as linkers which are used to link the different fragments were obtained by splitting drug molecules that are clinically used or in clinical trials. This would help to reduce the molecular search space effectively and make the generated molecules more drug-like. (4) The bioactivity of designed molecules is estimated by using a fit value, which describes how well a ligand is aligned with a pharmacophore model.

The rest of this paper is organized as follows: the second part presents a detailed description of the algorithms used in PhDD. In

* Corresponding author. Tel.: +86 28 85164063; fax: +86 28 85164060.
E-mail address: yangsy@scu.edu.cn (S.-Y. Yang).

1093-3263/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2010.02.002

the third part, PhDD is tested on three typical examples. The fourth part is a short discussion about the performance of PhDD. Conclusions will be offered in the final part.

2. Methodology

2.1. An overview of PhDD

The purpose of PhDD is to generate drug-like molecules which completely conform to the requirements of a given pharmacophore model. An overall flow chart for PhDD is presented in Fig. 1. The working process can be briefly described as follows. Given a pharmacophore hypothesis, in the first step, PhDD chooses proper fragments from different fragment databases that match chemical features of the pharmacophore hypothesis, followed by installing them to the 3D framework of the pharmacophore model. In the second step, PhDD tries to link all the disconnected fragments together by suitable linkers to form a complete molecule. In the third step, PhDD performs a series of assessments to the generated molecules, including assessments of drug-likeness, bioactivity, and synthetic accessibility. Details for all the algorithms used are given as follows.

2.2. Fragment and linker databases

Eight types of fragment databases and one linker database were established in advance. The eight types of fragment databases correspond to eight popular pharmacophore features respectively, including hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), positive ionizable (PI), negative ionizable (NI), ring aromatic (RA), hydrophobic (H), hydrophobic aromatic (HAR), and hydrophobic aliphatic (HAL) features. All the fragments and linkers were obtained by splitting the molecules in the databases of MDDR (MDL Drug Data Report) and CMC (Comprehensive Medicinal Chemistry), which might be helpful to make the created molecules more drug-like. The splitting operation was accomplished by using a module named Generate Fragments in Pipeline Pilot (Accelrys Inc., USA). The pharmacophore feature category for each fragment was identified by Catalyst (Accelrys Inc., USA), and confirmed visually. The molecular weight of each fragment and linker was restricted to be less than 250 and 200 Da, respectively. All the fragments and linkers were minimized by using the CHARMm force field [25] and stored in MOL2 format. The number of fragments is 385, 749, 43, 52, 530, 609, 436, and 173 for HBD, HBA, PI, NI, RA, H, HAR, and HAL, respectively. The linker database contains 1974 fragments. Users are allowed to modify each library to meet their specific purposes, including adding or deleting fragments/linkers from corresponding libraries.

2.3. Installation of fragments in the 3D framework of pharmacophore model

PhDD starts its work from a given pharmacophore model, which is the only input. The default format of pharmacophore model input file used here is the CHM format since it has been widely used and has become a *de facto* standard of pharmacophore model. Firstly, PhDD reads the pharmacophore feature information from the input, such as the number of pharmacophore features, coordinates of centers of pharmacophore features, tolerance of each pharmacophore feature, etc. For each pharmacophore feature, PhDD randomly chooses a fragment from its corresponding fragment database. Subsequently the chosen fragments are properly positioned in the 3D framework defined by the pharmacophore model. The placement of fragment should satisfy the following conditions: (1) the center of fragment should be superposed with that of its corresponding pharmacophore feature. For example, given a fragment 1-methyl-1H-indole and a ring aromatic feature (see Fig. 2a), PhDD places the center of the benzene ring over that of the ring aromatic feature (since 1-methyl-1H-indole has two ring aromatic centers, one benzene ring and one pyrrole ring, the larger one is automatically chosen); (2) if the pharmacophore feature is HBA, HBD or RA, a proper orientational adjustment of the fragment is needed to make it

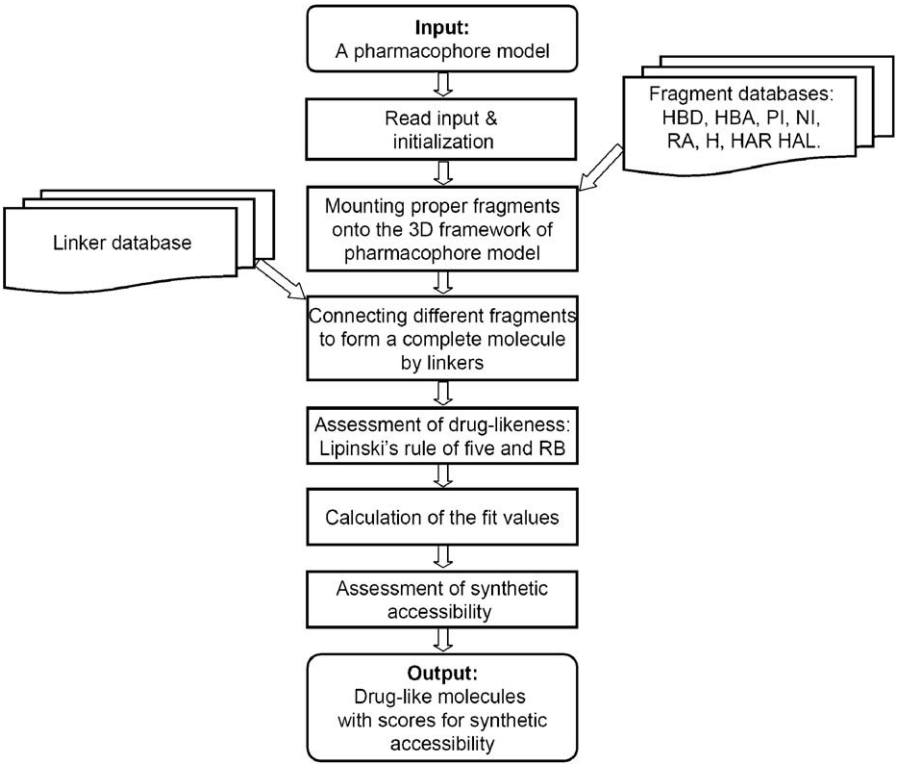

Fig. 1. Overall flow chart of PhDD.