Journal of Molecular Graphics and Modelling 44 (2013) 278–285

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

A combined molecular docking-based and pharmacophore-based target prediction strategy with a probabilistic fusion method for target ranking

Guo-Bo Li[a], Ling-Ling Yang[a,b], Yong Xu[a], Wen-Jing Wang[a], Lin-Li Li[a,c], Sheng-Yong Yang[a,*]

[a] State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, West China Medical School, Sichuan University, Sichuan 610041, China
[b] College of Chemical Engineering, Sichuan University, Sichuan 610041, China
[c] West China School of Pharmacy, Sichuan University, Sichuan 610041, China

ARTICLE INFO

Article history:
Accepted 12 July 2013
Available online 23 July 2013

Keywords:
Drug target prediction
Molecular docking
Pharmacophore
A probabilistic fusion method

ABSTRACT

Herein, a combined molecular docking-based and pharmacophore-based target prediction strategy is presented, in which a probabilistic fusion method is suggested for target ranking. Establishment and validation of the combined strategy are described. A target database, termed TargetDB, was firstly constructed, which contains 1105 drug targets. Based on TargetDB, the molecular docking-based target prediction and pharmacophore-based target prediction protocols were established. A probabilistic fusion method was then developed by constructing probability assignment curves (PACs) against a set of selected targets. Finally the workflow for the combined molecular docking-based and pharmacophore-based target prediction strategy was established. Evaluations of the performance of the combined strategy were carried out against a set of structurally different single-target compounds and a well-known multi-target drug, 4H-tamoxifen, which results showed that the combined strategy consistently outperformed the sole use of docking-based and pharmacophore-based methods. Overall, this investigation provides a possible way for improving the accuracy of in silico target prediction and a method for target ranking.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Drug target identification is extremely important not only for determining mechanism of action of active agents but also for anticipating their side effects or exploring possible new therapeutic indications of old drugs [1–5]. The most direct methods for the target identification correspond to those based on chemical biology [6–8]. However, these methods often require many expensive and time consuming wet experiments. In order to reduce the cost and save time, various computational methods [9], which are generally much cheaper and faster, have been involved in this kind of task. Because the predicted targets by computational methods still need further confirmation by wet experiments, a hybrid mode of target identification has been widely adopted at present, in which computational methods are first used to predict the potential targets, followed by validation by wet experiments. In this mode, the target prediction ability of computational methods is fairly important for the final success of target identification [9,10].

Currently a number of sophisticated computational methods have been established for the target prediction, which mainly include molecular docking-based, pharmacophore-based, molecular similarity-based, and others. A molecular docking-based method tries to dock a query compound to a panel of known target proteins to determine which one is the most likely interaction partner according to the scoring function. The representative examples of this method are INVDOCK [11] and TarFisDock [12]. A pharmacophore-based method finds the best mapping poses of the query molecule against a set of predefined pharmacophore models, in which each one corresponds to a target, and outputs the top best-fitted hits as the target candidates. PharmMapper is one of the typical representatives [13]. A molecular similarity-based method simply compares a query compound with a database of compounds whose targets are known. If the query compound is similar in structure with some compounds in the database, the targets of these compounds are considered as the target candidates of the query compound. This method is relatively simple and has more applications in recent years [5,14]. Other methods such as machine learning-based [15,16] and biochemical network-based [17–19] have also been developed recently.

Though each method has its own inherent advantages and disadvantages, which have been discussed in literature [10,20,21],

* 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 © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2013.07.005

these methods have some common problems. Of which the biggest problem for all of these methods is the poor target prediction ability. In finding a solution to this problem, we thought of a combined strategy of these methods, which has been used successfully in virtual screening by us [22,23] as well as other groups [24–26]. We thus, in this investigation, proposed a combined molecular docking-based and pharmacophore-based target prediction strategy. Here we chose the combination of the molecular docking-based and pharmacophore-based methods mainly because the two methods are apparently complementary. For example, the scoring function and the protein flexibility problems are obsessions in the docking-based method [20], whereas they are not a problem anymore in the pharmacophore-based method. The pharmacophore-based method often lacks consideration of receptor structural information [21], while it is a strong point of docking-based method. Even so, there is still a problem when using the combined strategy in target prediction, namely, how to sort the targets predicted by these methods. Here, we adopted a probabilistic fusion method for target ranking, which is based on Belief Theory (also known as Dempster–Schafer Theory) [27–29].

2. Methods

2.1. The target database

To construct a comprehensive potential target database (TargetDB), we first collected potential drug targets as many as possible from several public databases, including Therapeutic Target Database (TTD) [30], Potential Drug Target Database (PDTD) [31], DrugBank [32], and RSCB Protein Data Bank (PDB) [33]. Only those protein targets whose protein–ligand complex structures are known were selected. A total of 1105 different targets were deposited in TargetDB. Meanwhile, we also noticed that many of these targets have two or more crystal structures in the PDB database (see Supplementary Fig. S1). Thus, for some targets, several crystal structures are included; these structures have a relatively large difference. The finally formed TargetDB contains 1481 crystal structures covering the selected 1105 drug targets. These targets were annotated with biochemical type, therapeutic disease and development state.

2.2. The binding site database and the pharmacophore database

Based on TargetDB, we further constructed a binding site database and a pharmacophore database. Before the compilation of these databases, all the structures in TargetDB were prepared by utilizing DS 3.1 (Discovery Studio 3.1, Accelrys, Inc., San Diego, CA) software package. Operations for the preparation included: (i) removing water molecules and buffers, but preserving pivotal enzyme cofactor and metal cations; (ii) assigning CHARMM force filed [34]; (iii) for the structures with homopolymers, only one monomer was reserved; (iv) for the structures determined by NMR with multiple conformations, only the first conformation was remained.

The commercial molecular docking program GOLD [35] (CCDC, Cambridge CB2 1EZ, UK) was used in the docking-based target prediction; GOLD was chosen since it is one of the most widely used docking programs and has shown a better performance in virtual screening. Accordingly, the binding site database was created using GOLD, in which a binding site was defined as a sphere that contains all the residues around the ligand in the complex structure. A configuration file (gold.cfg) for each crystal structure including the absolute path of the corresponding protein target file and the 3-D coordinates of the binding site center was also recorded and saved for later use.

The pharmacophore database, which will be used in the pharmacophore-based target prediction method, was constructed using the module 'Receptor–Ligand Phamacophore Generation' implemented in the DS 3.1 software package. Six pharmacophore features, including hydrogen-bonding acceptor, hydrogen-bonding donor, aromatic ring, hydrophobic feature, positive charge center, and negative charge center, were considered in the model building process. Other parameters for the program were set as default. The program generated ten pharmacophore models for each complex, and the model with the highest score was selected to stay in the pharmacophore database. Overall, we finally obtained a binding site database containing 1481 binding sites and a pharmacophore database comprising 1481 pharmacophore models.

2.3. The docking-based and pharmacophore-based target prediction protocols

The GOLD program was taken as the docking engine in the docking-based target prediction method. The protocol or workflow for the docking-based target prediction method can be briefly described as follows: (i) preparing the query compound; (ii) docking the query compound to each binding site in the binding site database using GOLD, and calculating two scoring functions: Chemscore (empirical) [36] and Goldscore (force field-based) [37]; (iii) preserving the best docking pose for each target, and extracting the corresponding scoring values; (iv) prioritizing the targets according to the scoring values of Chemscore and Goldscore, respectively. The top-ranking targets are supposed to be the most potential targets of the query compound.

The Catalyst program [38] implemented in DS 3.1 software package was used in the pharmacophore-based target prediction method. The protocol or workflow for the pharmacophore-based target prediction method can be simply described as follows: (i) generating conformers of the query compound using the 'fast conformer generation' approach with 20 kcal/mol being set as the energy cutoff and 250 as the maximum number of conformers; (ii) mapping the generated conformers onto each pharmacophore model in the pharmacophore database using a grid-fitting method; (iii) calculating the fitness value, which is used to define how well a given compound is mapped to a pharmacophore model, according to the following formula (Eq. (1)) [39]:

$$Fitness = \frac{\sum_n [1 - \sum (d/t)^2]}{n} \quad (1)$$

where n denotes the number of pharmacophore features, d represents the displacement of the feature from the center of the location constraint, t is the radius of the location constraint sphere for the feature (tolerance); (iv) prioritizing all the pharmacophore models (actually they correspond to targets) in the pharmacophore database according to the fitness values. The top best-fitted hits are considered as the target candidates of the query compound.

2.4. The probabilistic fusion method

To provide a reasonable ranking order for the targets in the combined docking-based and pharmacophore-based target prediction method, we introduced a probabilistic fusion method, which is based on Belief Theory (also known as Dempster–Schafer Theory) [27]. The basic requirement of Belief Theory is that quantifiable probabilities of an event being true can be obtained. For satisfying this requirement, we created a training set to construct probability assignment curves (PACs), which are empirically derived functions that can translate a measure (e.g. Chemscore) into a probability of true prediction by this measure. The training set contains 20 protein targets, which cover a variety of biochemical types (see Supplementary Table S1). For each target, 200 known ligands or actives

Download English Version:

https://daneshyari.com/en/article/6877788

Download Persian Version:

https://daneshyari.com/article/6877788

Daneshyari.com