Computational Biology and Chemistry 73 (2018) 139–146

Contents lists available at ScienceDirect

Computational Biology and Chemistry

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

Research Article

Multiple grid arrangement improves ligand docking with unknown binding sites: Application to the inverse docking problem

Tomohiro Ban[a,b], Masahito Ohue[a,c], Yutaka Akiyama[a,b,c,*]

[a] School of Computing, Tokyo Institute of Technology, 2-12-1 W8-76 Ookayama, Meguro-ku, Tokyo 152-8550, Japan
[b] Education Academy of Computational Life Sciences, Tokyo Institute of Technology, 2-12-1 W8-93 Ookayama, Meguro-ku, Tokyo 152-8550, Japan
[c] Advanced Computational Drug Discovery Unit, Institute of Innovative Research, Tokyo Institute of Technology, 4259 Nagatsutacho, Midori-ku, Yokohama, Kanagawa 226-8501, Japan

ARTICLE INFO

Article history:
Received 5 June 2017
Received in revised form 5 February 2018
Accepted 10 February 2018
Available online 15 February 2018

Keywords:
Computational ligand docking
Conformational search space
Drug–target interactions
Inverse docking
Multiple grid arrangement
Structure-based drug design
Scoring function
Virtual screening

ABSTRACT

The identification of comprehensive drug–target interactions is important in drug discovery. Although numerous computational methods have been developed over the years, a gold standard technique has not been established. Computational ligand docking and structure-based drug design allow researchers to predict the binding affinity between a compound and a target protein, and thus, they are often used to virtually screen compound libraries. In addition, docking techniques have also been applied to the virtual screening of target proteins (inverse docking) to predict target proteins of a drug candidate. Nevertheless, a more accurate docking method is currently required. In this study, we proposed a method in which a predicted ligand-binding site is covered by multiple grids, termed multiple grid arrangement. Notably, multiple grid arrangement facilitates the conformational search for a grid-based ligand docking software and can be applied to the state-of-the-art commercial docking software Glide (Schrödinger, LLC). We validated the proposed method by re-docking with the Astex diverse benchmark dataset and blind binding site situations, which improved the correct prediction rate of the top scoring docking pose from 27.1% to 34.1%; however, only a slight improvement in target prediction accuracy was observed with inverse docking scenarios. These findings highlight the limitations and challenges of current scoring functions and the need for more accurate docking methods. The proposed multiple grid arrangement method was implemented in Glide by modifying a cross-docking script for Glide, xglide.py. The script of our method is freely available online at http://www.bi.cs.titech.ac.jp/mga_glide/.

1. Introduction

Computational protein–ligand docking is a computational tool used to predict the optimal binding mode and affinity of the complex between structures of proteins and compounds. Ligand docking enables the visualization of an optimal complex that can be predicted from a target protein structure and a candidate drug compound. Moreover, it enables the rapid screening of potential hit compounds against a target protein (called virtual screening) (Halgren et al., 2004; Cheng et al., 2012). These advantages have encouraged the use of computational ligand docking in drug discovery for exploring lead compounds for target proteins. In addition, ligand docking has been used for inverse docking (Chen and Zhi, 2001; Chen and Ung, 2001; Li et al., 2006; Schomburg et al., 2014) to identify targets of candidate drug compounds. This screening can be used to avoid off-target protein binding and side effects that can halt further development (Paul et al., 2010; Arrowsmith and Miller, 2013), as well as to uncover new therapeutic effects, which is known as drug repositioning (Schomburg et al., 2014; Ashburn and Thor, 2004; Ekins et al., 2011).

Numerous versions of computational ligand docking software have been developed since the launch of DOCK in 1982 (Kuntz et al., 1982). Two of the most important elements of ligand docking are a conformational search algorithm to generate sample conformation and a scoring function to evaluate hits. Numerous studies have proposed improvements for docking accuracy (Goodsell and Olson, 1990; Goodsell et al., 1996; Trott and Olson, 2010; Eldridge et al., 1997; Korb et al., 2009; Spitzer and Jain, 2012; Jones et al., 1997; Ruiz-Carmona et al., 2014; Friesner et al., 2004).

* Corresponding author at: School of Computing, Tokyo Institute of Technology, 2-12-1 W8-76 Ookayama, Meguro-ku, Tokyo 152-8550, Japan.
E-mail addresses: ban@bi.c.titech.ac.jp (T. Ban), ohue@c.titech.ac.jp (M. Ohue), akiyama@c.titech.ac.jp (Y. Akiyama).

https://doi.org/10.1016/j.compbiolchem.2018.02.008

Current scoring functions include the physicochemical score function implemented in AutoDock (Goodsell and Olson, 1990; Goodsell et al., 1996) and AutoDock Vina (Trott and Olson, 2010), and empirical scoring functions such as the Chemscore (Eldridge et al., 1997), ChemPLP (Korb et al., 2009), and Surflex-Dock's scoring functions (Spitzer and Jain, 2012). Furthermore, conformational search algorithms such as the genetic algorithm-based algorithm implemented in GOLD (Jones et al., 1997) and rDock (Ruiz-Carmona et al., 2014) can be used with hierarchical comprehensive search programs such as Glide (Friesner et al., 2004). These kinds of studies enable the improvement of the accuracy of ligand docking (Kellenberger et al., 2004). Furthermore, improving the capability of the scoring function, conformational search, or both is expected to improve ligand docking, virtual screening, and inverse docking.

We focused on the conformational search process of ligand docking in this study. Although various algorithms have been developed to improve the capability of the conformational search, the receptor grid, which determines its conformational search space, is also an important factor for improving the docking accuracy in grid-based ligand docking algorithms such as Glide, AutoDock, and AutoDock Vina. However, the question of how to arrange grids (called "grid arrangement") has not been adequately addressed in previous studies. It was eventually reported that the docking accuracy depends on the size of the receptor grid (Feinstein and Brylinski, 2015). Therefore, it is necessary to discuss the modalities for generating receptor grids with an appropriate size and arrangement, as well as designing the score function and the conformational search algorithm. This necessity is particularly obvious with the inverse docking problem involving the unknown binding site of target proteins. This is because there are numerous ways to generate the receptor grids to cover the ligand-binding site candidates. Thus, the development of the optimal grid arrangement algorithm is important.

In this study, we proposed a multiple grid arrangement method for grid-based ligand docking to enhance the conformational search capability (Fig. 1). We implemented the technique with the state-of-the-art commercial docking software Glide (Friesner et al., 2004). The proposed method was validated and compared with the Glide standard method in two computational experiments—the re-docking (Morris et al., 2009) and inverse docking (Schomburg et al., 2014) tests—to confirm its ability to discover the correct docking pose and binding target protein of a given compound, respectively.

2. Materials and methods

2.1. Datasets

Performances of the grid arrangement methods were evaluated on the Astex diverse dataset (Hartshorn et al., 2007), consisting of 85 complexes, along with all complex structures in the Protein Data Bank (Berman et al., 2000) (PDB). All structures had a resolution of 2.5 Å or better. This dataset had 84 total ligands (the ligands of "1of6" and "1x8x" are the same) for all 85 proteins. The ligands were distinct and corresponded to each PDB id. This was because the experiment was aimed at predicting the co-crystallized target for each of the 85 ligands as a true partner and ranking the true partner to the best one of the 85 proteins, similar to a previous study (Schomburg et al., 2014).

2.2. Re-docking and inverse docking

The performance of a docking calculation is evaluated by the methods of the re-docking test and inversed docking test. The re-docking test is an experimental method to evaluate the reproducibility of a complex structure by docking calculation. If the root-mean-square deviation (RMSD) for the ligand between the experimentally determined complex and the predicted complex was $<$2.0 Å when their protein structures were superimposed, the complex was predicted correctly. In several studies, the re-docking test only docks to the known binding site, to which the ligand was co-crystalized (Kellenberger et al., 2004; Erickson et al., 2004; Repasky et al., 2012). In this study, however, the assumption was that the binding site of the ligand is unknown. Since the docking calculation is assumed, the ligand binding site is predicted, and the docking calculation is performed for the entire predicted site.

In contrast, the inverse docking test evaluates the ability to screen true target proteins against specific ligands. The evaluation method was carried out in accordance with the preceding study

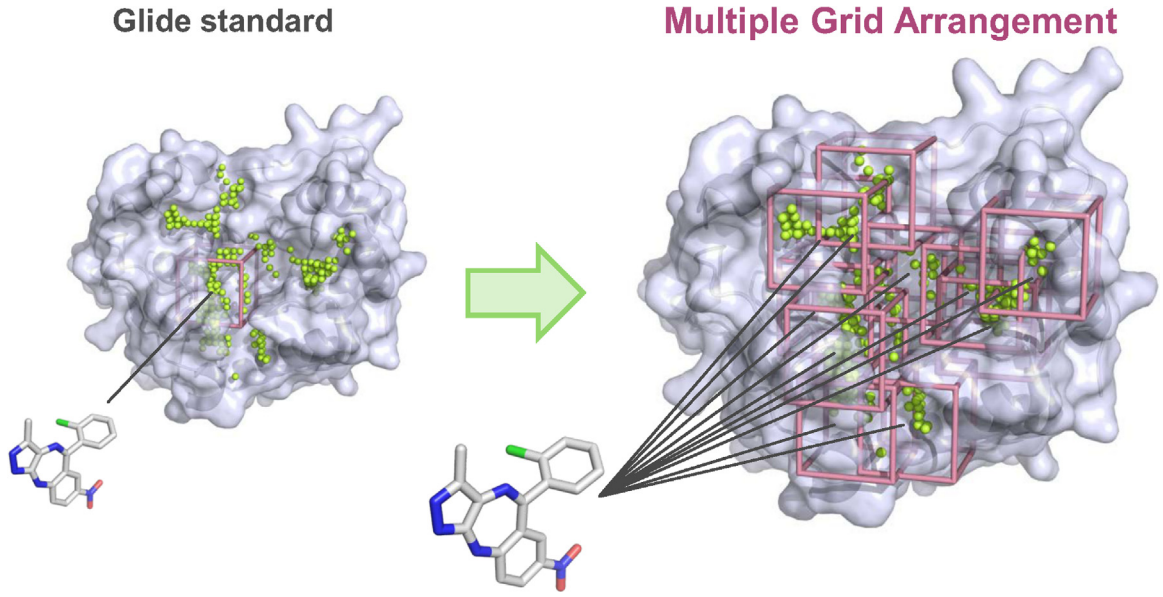

Fig. 1. Illustration of the multiple grid arrangement method. Green dots on the protein indicate a predicted ligand-binding site, and pink boxes show receptor grids for ligand docking. The multiple grid arrangement method can cover an entire space of any predicted ligand-binding site automatically. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)