Review

In silico target identification of biologically active compounds using an inverse docking simulation

Youngjin Choi

BioChip Research Center and Inflammatory Disease Research Center, Hoseo University, 165 Sechul-ri, Baebang-up, Asan, Chungnam, 336-795, Republic of Korea

ABSTRACT

Identification of target protein is an important procedure in the course of drug discovery. Because of complexity, action mechanisms of herbal medicine are rather obscure, unlike small-molecular drugs. Inverse docking simulation is a reverse use of molecular docking involving multiple target searches for known chemical structure. This methodology can be applied in the field of target fishing and toxicity prediction for herbal compounds as well as known drug molecules. The aim of this review is to introduce a series of *in silico* works for predicting potential drug targets and side-effects based on inverse docking simulations.

Keywords inverse docking, herbal medicine, target prediction, computer simulation

INTRODUCTION

Rapid advances in computer-aided molecular design (CAMD) have led to the ever-increasing interests in structure-based drug discovery for obtaining pharmaceutical candidates having suitable efficacy and safety (Sinko et al., 2013). Computational tools including pharmacophore searching, molecular docking, data mining, and dynamic simulations are widely used in virtual drug screening process (Ekins et al., 2007). In particular, protein-ligand docking has been used as a gold-standard approach for both the business and academic institutions when three-dimensional structures of target receptors are available (Kitchen et al., 2004; Moitessier et al., 2008; Shoichet et al., 2002; Taylor et al., 2002). Multiple small molecules in the chemical libraries are concurrently docked to single receptor site in this approach. Therefore, the ligand-docking program is suitable for high-throughput evaluation of binding affinities of small molecules upon druggable target site to effective lead-generation. Various research groups have been developed different flexible docking algorithms, for example simulated annealing, shape-matching, genetic or evolutionary programming, grid-based, incremental construction, and Monte-Carlo method (Kitchen et al., 2004; Moitessier et al., 2008). These molecular docking approaches are practically being used in the field of target-oriented drug development by means of appointed action mechanism.

By contrast, orphan compounds refer to chemicals intended for modulating cellular functions but having unknown target proteins (Jenkins et al., 2006). Many natural products and herbal medicine are extensively used as therapeutics in clinic, but their mechanism-of-action (MOA) is not yet fully understood (Barlow et al., 2012). The subsequent target identification is therefore encouraged for the well-defined

medicinal use of herbal compounds. A series of *in silico* target fishing methods such as data mining, similarity searching, machine learning, molecular docking, and bioactivity spectra has been introduced to make realize the prediction of biological target of the compounds (Jenkins et al., 2006; Nonell-Canals et al., 2011; Rognan, 2010). In the present review, ligand-protein inverse docking simulation is mainly described as a way to predict target protein *in silico*.

Molecular docking and inverse docking simulations

Identifying the three-dimensional structure of druggable protein is a starting point in the current target-based drug discovery process (Canduri and Azevedo, 2008; Chrysina et al., 2011). As more and more protein structures become accessible, protein-ligand docking recognized to be important virtual drug screening tool (Kitchen et al., 2004). Protein-ligand docking is usually performed with computing facility to reach high-throughput, whereby large chemical libraries are automatically applied to single protein target. Over 60 docking programs have been developed by industries or academies in the world (Moitessier et al., 2008). Each docking algorithm relies on a sampling method to explore conformational space and a scoring function to evaluate binding affinity of docked poses.

For the conformational sampling methods (Taylor et al., 2002), multi-conformer docking (DOCK, FDock, LIGIN, and SANDOCK) use multi-conformational rigid-body ligand docking with shape complementary or interaction site matching algorithm. An incremental construction algorithm (DOCK, FlexX, FLOG, Hammerhead, and Surflex) builds up the ligand in the active site by connecting the fragments. Stochastic approaches often use genetic algorithm (AutoDock, DARWIN, GOLD, and PSI-DOCK) or Monte Carlo search (ICM, MCDOCK, and SLIDE).

Scoring is a critical step to decide docking accuracy for the binding free energy or compounds' rank-order in the virtual screening process. The scoring functions are classified into three categories; empirical, force-field, and knowledge-based functions (Kitchen et al., 2004). Empirical scoring function is

designed by approximation that binding energies are expressed as a sum of individual contributing terms. The coefficients are taken from structural information determined by previous experiments. ChemScore, LUDI, and X-Score are representative scoring methods in the type of empirical functions (Böhm, 1992; Eldridge et al., 1997). Force-field based scoring function estimates internal ligand energy and intermolecular energy between ligand and protein. The ligand-protein interaction is usually described by the combination of electrostatic and van der Waals energy terms. The Amber, CHARMM, and Tripos are typical force-field used in this type of scoring functions. (Cramer, 1989; Vanommeslaeghe et al., 2010; Wang et al., 2004). In the case of knowledge-based function, docking score has been developed from statistical analysis of ligand-protein structure. These analyses include converting process from ligand-protien atom typing into pair-wise potential. The DrugScore, SMoG, and PMF are well-known knowledge-based scoring functions suitable to reproduce ligand-protein complex structure rather than binding affinity (Muegge et al., 1999; Velec et al., 2005).

Despite some disappointments, docking tools are useful in guiding experimental works in the drug discovery process (Ekins et al., 2007). Moreover docking simulations have been gradually applied to serial screening of target proteins with small molecules as many protein structures were available from public database. This type of job was named 'ligand-protein docking' or 'inverse docking' process (Chen and Zhi, 2001). The inverse docking approach deals with a single ligand onto multiple protein while general docking handles one protein target with a great deal of small-molecule ligands. Therefore the inverse docking jobs require as many as protein target structures of known three-dimensional coordinates; at this moment (March 2013 release) there are 82,376 protein 3D structures in Protein Data Bank (Sussman et al., 1998). The inverse docking (Chen and Ung, 2001; Chen and Zhi, 2001) can be useful in the field of drug discovery as follows: (1) therapeutic target fishing of small-molecular drug, (2) toxicity and side-effect predicting for the chemicals, (3) mechanism probing of biologically-active natural products or herbal medicine in the molecular level. As following topics, practical applications of the inverse docking simulations were introduced in brief.

Case studies in applying inverse docking simulations
Protein-ligand docking is a high priority in the process of conventional virtual screening in which multiple small-molecules are ranked by specific docking score for a corresponding protein target. This method is known to be capable of predicting hit molecules and bound conformations close to experimentally determined structures. Ligand-protein inverse docking is a reverse use of the protein-ligand docking simulation. Inverse docking is an attempt to find multiple putative targets to which a small molecule of interest can bind or interact. Chen and Zhi (2001) have been developing inverse docking program called INVDOCK for the automated search of the protein targets of a small molecules. The INVDOCK process is composed of four steps: ligand aligning within binding sites, conformation optimization by a torsion space sampling, energy minimization of ligand complexed with protein binding sites, and finally scoring. Principal basis of the INVDOCK algorithm is similar to flexible ligand docking software, DOCK (Oshiro et al., 1995) developed throughout pioneering works in Kuntz's lab. An anticancer drug, tamoxifen was evaluated with the INVDOCK to find additional protein target of the drug. Most of the known targets for tamoxifen, such as estrogen receptor, alcohol dehydrogenase, protein kinase C, prostaglandin synthetase, and collagenase were

re-confirmed on the INVDOCK search. Moreover two potential target proteins including glutathione transferase and 3α-hydroxysteroid dehydrogenase were identified by the INVDOCK work. For the target of vitamin E, the INVDOCK did successfully reproduce seven known targets and predict eleven potential targets. The computational studies performed by Chen and coworkers showed the potential of the INVDOCK in the research field of virtual target fishing for a lead compound or drug candidate.

Paul et al. (2004) recently attempted to recover true target of drug-like ligands from Protein Data Bank (PDB) using inverse-docking procedure. They constructed an in-house screening protein library (sc-PDB) of 2148 entries from original PDB database. The sc-PDB is an annotated binding site DB for protein targets, and currently contains 2380 different proteins on September 2012. The inverse-docking simulations performed with GOLD program (Verdonk et al., 2003) offered potential targets for four investigated ligands containing biotin, tamoxifen, 6-hydroxyl-1,6-dihydropurine ribonucleoside, and methotrexate. This approach was fast enough compared to INVDOCK, and was able to define target binding sites accurately.

Considering a safety issue for lead compounds or even for herbal medicine is an important aspect in the process of drug discovery. Chen and Ung (2001) have made an effort to predict potential toxicity and side effects caused by ligand binding on the protein off-targets. The automated inverse-docking procedure was applied in this work using limited target entities known to be related with the toxicity or side effect. The most well known drugs including aspirin, gentamicin, ibuprofen, indinavir, neomycin, penicillin G, 4H-tamoxifen, and vitamin C were evaluated using INVDOCK procedure. For the confirmed off-targets of aspirin, alcohol dehydrogenase, antithrombin, and L-3-hydroxyacyl-CoA dehydrogenase were identified by the INVDOCK process. Each off-target of aspirin was involved with toxicity/side effects such as blood alcohol increasing, thrombolysis, and Reye's syndrome, respectively. Different off-targets were also finely predicted by subsequent INVDOCK jobs for other therapeutic agents. In total, the INVDOCK predicted 38 targets within 68 experimentally determined targets for the regarding 8 drugs. It missed five targets and found extra 29 targets without experimental finding.

TarFisDock is another inverse docking tool for the searching of protein target with small molecule query in automatic mode (Li et al., 2006). This program is available *via* web server, in which potential drug target database (PDTD) is incorporated as a repository for the structure of target proteins. PDTD is a web-accessible drug target database containing 1,207 entries with 841 known 3D structures from PDB (Gao et al., 2008). TarFisDock automatically searches for possible combination of target protein for the given small molecule input file by using inverse docking engine. The ligand-protein interaction is ranked by binding energy terms originated from DOCK program. A plausible protein target list may be selected from the top 2, 5 or 10% ranking output for subsequent biological evaluations. Cai et al. (2006) conducted inverse docking study to discover unknown drug targets for *Helicobacter pylori* throughout the TarFisDock server. They drew possible 15 target proteins from the automatic inverse docking by biologically active natural compound. Using homology search and enzymatic assay, a peptide deformylase was finally suggested as a potential target for anti-*Helicobacter* drugs.

Recent reverse docking approach by clustering the docking profile has suggested the predictability of the inverse docking simulation for the druggability of protein targets. Lee and Kim (2012) have performed large-scale reverse docking simulations

for 10,886 human proteins with thirty-five drug compounds. They found previous inverse docking approaches were insufficient to search wide target space because pre-existing target structure DBs were usually inherent only non-redundant protein conformation. The *in silico* target fishing of hydrocortisone with PDTD did fail to detect authentic glucocorticoid receptor as a target. The PDTD just included two glucocorticoid receptor structures out of nine in the PDB. The large scale reverse docking approach using non-redundant target structure DB made a success to find glucocorticoid receptor as the top-ranking protein target of hydrocortisone. This study recommends that large-scale inverse docking simulations with as many as protein structures is preferred way in finding potential drug targets or their mode of action.

Traditional herbal medicine database for the inverse docking approach
For general traditional medicine attempt, it is often hard to obtain compound information from herbal medication. As chemical structure is available, target prediction is also possible using by inverse docking simulations. Therefore, obtaining a compound's information is the functional requirements for target fishing process of natural products or herbal medicine *in silico*. General databases for this purpose are currently available on the web-service. Oriental Medicine Advanced Searching Integrated System (OASIS) provides information on the identified chemical ingredients from well-known 25 oriental prescriptions. Users are able to browse a group of chemical structures belong to the type of herbal medicine. TCM database@Taiwan is non-commercial traditional Chinese medicine DB based on scientific texts and publications (Chen, 2011). It contains more than 20,000 compound's information from oriental medicine or herbal drugs. These databases are valuable resource for predicting target proteins of some herbal products because they supply structural information of single compound in the complex oriental medications.

CONCLUSION

Inverse docking simulation is useful *in silico* methodology to identify potential drug targets or side effects concerning the compounds of herbal prescription. This approach can be used for predicting action mechanism or toxicity of traditional medicine as well as drug candidates. Various molecular docking algorithms have been incorporated and modified to use them as suitable simulation engines in the inverse docking program.

ACKNOWLEDGEMENTS

CONFICT OF INTERESTS

The author states no conflict of interest.

REFERENCES

Barlow DJ, Buriani A, Ehrman T, Bosisio E, Eberini I, Hylands PJ. In-silico studies in Chinese herbal medicines' research: evaluation of in-silico methodologies and phytochemical data sources, and a review of research to date. J Ethnopharmacol. 2012;140:526-534.

Böhm HJ. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. J Comput Aided Mol Des. 1992;6:593-606.

Cai J, Han C, Hu T, Zhang J, Wu D, Wang F, Liu Y, Ding J, Chen K, Yue J, Shen X, Jiang H. Peptide deformylase is a potential target for anti-*Helicobacter pylori* drugs: reverse docking, enzymatic assay, and x-ray crystallography validation. Protein Sci. 2006;15:2071-2081.

Canduri F, de Azevedo WF. Protein crystallography in drug discovery. Curr Drug Targets. 2008;9:1048-1053.

Chen CYC. TCM database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. PloS one. 2011;6:e15939.

Chen YZ, Ung CY. Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand-protein inverse docking approach. J Mol Graph Model. 2001;20:199-218.

Chen YZ, Zhi DG. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. Proteins. 2001;43:217-226.

Chrysina ED, Chajistamatiou A, Chegkazi M. From structure--based to knowledge--based drug design through x-ray protein crystallography: sketching glycogen phosphorylase binding sites. Curr Med Chem. 2011;18:2620-2629.

Cramer RD 3rd, Patterson DE, Bunce JD. Recent advances in comparative molecular field analysis (CoMFA). Prog Clin Biol Res. 1989;291:161-165.

Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. Br J Pharmacol. 2007;152:9-20.

Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des. 1997;11:425-445.

Gao Z, Li H, Zhang H, Liu X, Kang L, Luo X, Zhu W, Chen K, Wang X, Jiang H. PDTD: a web-accessible protein database for drug target identification. BMC Bioinformatics. 2008;9:104-110.

Jenkins JL, Bender A, Davies JW. In silico target fishing: predicting biological targets from chemical structure. Drug Discov Today: Technologies. 2006;3:413-421.

Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. 2004;3:935-949.

Lee M, Kim D. Large-scale reverse docking profiles and their applications. BMC Bioinformatics. 2012;13:S6.

Li H, Gao Z, Kang L, Zhang H, Yang K, Yu K, Luo X, Zhu W, Chen K, Shen J, Wang X, Jiang H. TarFisDock: a web server for identifying drug targets with docking approach. Nucl Acids Res. 2006;34:W219-224.

Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR. Towards the development of universal, fast and highly accurate docking/scoring methods: a long ways to go. Br J Pharmacol. 2008;153:S7-S26.

Muegge I, Martin YC. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem. 1999;42:791-804.

Nonell-Canals A, Mestres J. In silico target profiling of one billion molecules. Mol Inf. 2011;30:405-409.

Oshiro CM, Kuntz ID, Dixon JS. Flexible ligand docking using a genetic algorithm. J Comput Aided Mol Des. 1995;9:113-130.

Paul N, Kellenberger E, Bret G, Muller P, Rognan D. Recovering the true targets of specific ligands by virtual screening of the protein data bank. Proteins. 2004;54:671-680.

Rognan D. Structure-based approaches to target fishing and ligand profiling. Mol Inf. 2010;29:176-187.

Shoichet BK, McGovern S, Wei B, Irwin JJ. Lead discovery using molecular docking. Curr Opin Chem Biol. 2002;6:439-446.

Sinko W, Lindert S, McCammon JA. Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design. Chem Biol Drug Des. 2013;81:41-49.

Sussman JL, Lin D, Jiang J, Manning NO, Prilusky J, Ritter O, Abola EE. Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. Acta Crystallogr D Biol Crystallogr. 1998;1:1078-1084.

Taylor RD, Jewsbury PJ, Essex JW. A review of protein-small molecule docking methods. J Comput Aided Mol Des. 2002;16:151-166.

Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem. 2010;31:671-690.

Velec HF, Gohlke H, Klebe G. DrugScore(CSD) -knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. J Med Chem. 2005;48:6296-6303.

Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins. 2003;52:609-623.

Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J Comput Chem. 2004;25:1157-1174.