ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES
Knowledge to Innovation

Vol 8, Issue 1, 2015

ISSN - 0974-2441

Research Article

A STATISTICAL ANALYSIS TO FIND OUT AN APPROPRIATE DOCKING METHOD

JUN M KALITA*, SURAJIT K GHOSH, SUPRIYA SAHU, MAYURAKHI DUTTA

Department of Pharmaceutical Sciences, Dibrugarh University, Dibrugarh, Assam, India. Email: pjmk84@gmail.com

Received: 26 August 2014, Revised and Accepted: 25 October 2014

ABSTRACT

Objective: The objective of the study was to determine a suitable and reliable docking protocol based on a statistical study.

Methods: Statistical analysis was done to find out the co-relation between *in-silico* and *in-vitro* results.

Results: All the docking method shown acceptable root mean square deviation (RMSD) value those were found to be less than that of 2 Å. Coefficient of co-relation value was also quite satisfactory with highest of r = 0.6574 and r_s=0.8322 for LigandFit.

Conclusion: Among the three docking method used for the study, LigandFit was found to be more appropriate as it shown low RMSD and high co-relation coefficient.

Keywords: Dihydrofolate reductase, 1J3I, LigandFit, CDOCKER, Gold, Karl Pearson's coefficient of correlation, Spearman's rank correlation (r_s).

INTRODUCTION

Docking, the important tool in drug design and discovery involves search of the best possible conformations and orientations of a ligand in the target binding site [1,2]. Virtual screening tools are an important part of the modern era of drug design and discovery. It is a tedious task to find out the most suitable virtual screening tool mainly in case of protein ligand docking [3-7]. LigandFit is a shape based docking of a flexible ligand at the active site of the protein by cavity detection method, various spaces inside the target protein active site are searched as correspond to the shape of the ligands. A combination of shape comparison filter and Monte Carlo technique is used to search various ligand poses and then these poses are minimized within the active site to get protein ligand interaction energy [8-11]. CDOCKER is a CHARMm-based docking engine where flexible ligands are docked at protein active site. Several random conformations of the ligand are generated at the protein active site which is then followed by MD-based simulated annealing composed of several heating and cooling steps. The final score is obtained by energy minimization process [12,13]. Whereas GOLD docks flexible ligands at the protein active site by employing a genetic algorithm (GA) [14,15].

The strength of association of two sets of variables can be measured by the use of different coefficients. Among them the three most widely used are Pearson's coefficient of correlation (r), spearman's rank correlation (r_s), and Kendall's tau coefficient (T) [16]. Here in the present work Pearson's coefficient of correlation and Spearman's rank correlation are been used to select a best docking method on the basis of correlation between dock score and reported inhibitory concentration 50 values of some anti-folate molecules.

Although difficult, it is important to select the best docking tool that can be suitable for a particular study. In the present work comparative, statistical analysis is been carried out to select the most suitable docking method out of three selected methods [17].

METHODS

In silico study

Preparation of protein

X-ray crystal structure of wild type *Pf*-DHFR-TS complex was obtained from protein data bank using Accelrys' Discovery studio version 2.5 (PDB entry code: 1J3I). Water molecules, co-crystallized ligand WR99210 were removed and cofactors Nicotinamide adenine dinucleotide phosphate, dUMP was allowed to retain. Finally, receptor was prepared according to the requirements of the docking protocols LigandFit, Gold and CDOCKER.

Preparation of ligand

Reference ligands (Fig. 1) were prepared by Marvin sketch tool as supported by Sanjeevani online program. The 3D structures of the ligands were imported to Discovery Studio workplace and energy minimization was done by applying CharmM forcefield.

Docking

Validation of all the docking protocol was done by the calculating root mean square deviation (RMSD) value of the docked ligand with respect to the co-crystallized ligand (Fig. 2).

LigandFit

In order to simulate the biological environment, CharmM forcefield was applied. Sites were generated from the receptor protein, and site 1 was selected for docking. All the molecules were docked keeping dreiding energy grid with an extension of 5.0 Å, dielectric constant as 1.0, and parallel processing was kept as false. Binding energy of the docked molecules was calculated using calculate binding energy protocol.

Gold

Receptor protein was prepared according to the requirements of the protocol. A sphere was constructed around active site 1 and molecules were docked at the active site. This docking protocol employs GA to dock the molecules at the assigned site. GA was speeded up by 7-8 times, and GA automatic search efficiency was kept as 100.0. Finally, binding energy was calculated using calculate binding energy protocol.

CDOCKER

This is a CHARMm forcefield based docking method where a sphere was constructed around the active site of the receptor protein. Molecules were then docked at the active site keeping simulated annealing as true and parallel processing was kept false. Finally, ligand-protein binding energy was calculated for all the docked ligands.

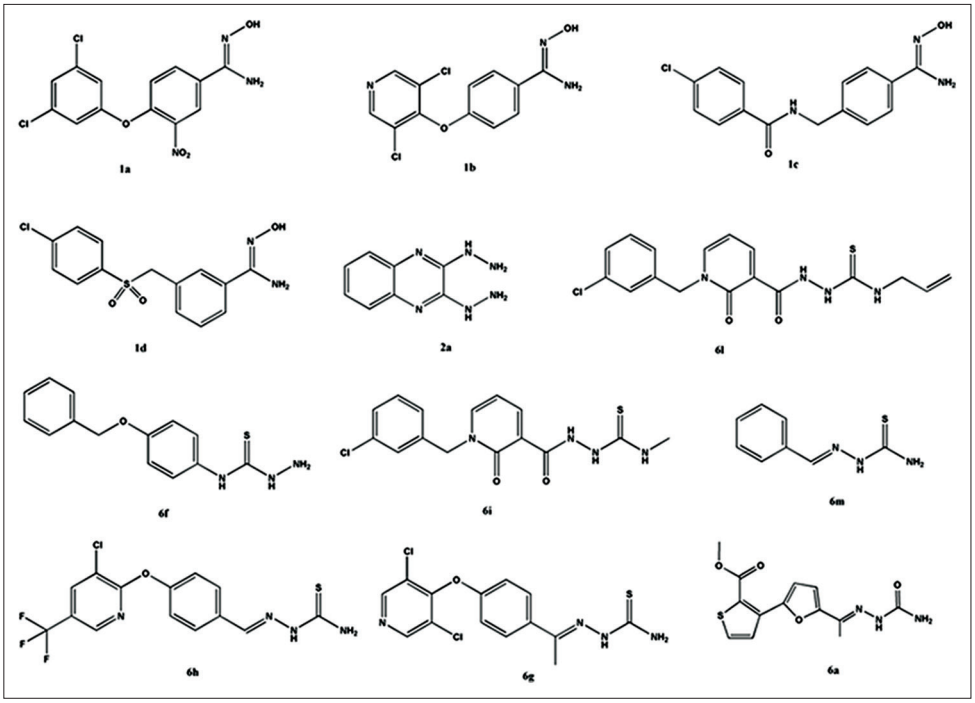

Fig. 1: Structure of the reference ligands used for the study

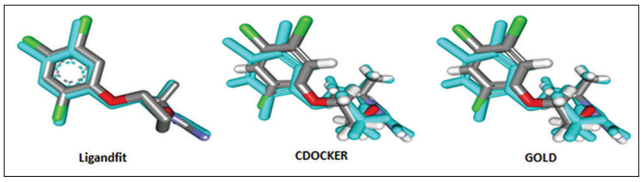

Fig. 2: Root mean square deviation poses for the docking protocols

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\sum x^2 - \left(\sum x^2\right)\right]\left[n\sum y^2 - \left(\sum y^2\right)\right]}} \qquad r_S = 1 - \frac{6\left(\sum d^2\right)}{n\left(n^2 - 1\right)}$$

Fig. 3: Formula used for the calculation of co-relation coefficient

Statistical analysis

K_i values for the reference ligand were obtained from Ref [8] then the data were organized and Karl Pearson's coefficient of correlation (r) and spearman's rank correlation (r_s) were calculated using the standard formulas as shown in Fig. 3.

RESULTS AND DISCUSSIONS

Calculated RMSD value reveals that docking protocol LigandFit closely reproduces the co crystallized ligand with an RMSD value of 0.248 Å whereas RMSD value for CDOCKER and Gold was found to be 0.538 Å and 0.6187 Å (Fig. 2). Although an RMSD value <2 Å is considered as valid, here docking protocol LigandFit can be said as most accurate because of of its lowest RMSD value.

Calculated Karl Pearson coefficient of correlation (r) and rank correlation (r_s) signifies that docking protocol LigandFit is more accurate than that of other docking protocols. As in the case of LigandFit both Karl Pearson coefficient of correlation (r) and rank correlation (r_s) are greater and more near to one (Table 1). This signifies that the docking value obtained from LigandFit fits more accurately to the practical K_i. Finally, it can be concluded that the accuracy of docking protocol follows the following order LigandFit > CDOCKER > Gold, where LigandFit is more accurate, and Gold is the least.

Table 1: Estimated binding energy and K_i values

Molecule*	K_i*	(-) Binding energy LigandFit	(-) Binding energy CDOCKER	(-) Binding energy Gold
1a	2.4	133.038	57.568	60.028
1b	25.2	47.195	115.170	42.770
1c	77.2	16.096	43.722	45.628
1d	102	12.389	20.464	12.802
2a	3.6	71.560	45.576	41.101
6a	9.7	70.968	136.419	61.520
6f	13.7	70.231	85.708	60.617
6g	0.9	121.749	137.807	129.082
6h	5.2	60.205	126.214	165.403
6i	11.3	38.629	60.156	78.402
6l	15.2	15.687	43.880	58.592
6m	23.6	68.969	30.926	33.067

*K_i values were obtained from ref no. [18], r=0.6574; r=0.5172; r=0.5324, r_s=0.8322; r_s=0.6294; r_s=0.6364, RMSD=0.248; RMSD=0.538; RMSD=0.6187, Coefficient of correlation (r), rank correlation (r_s), RMSD: Root mean square deviation

REFERENCES

1. 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(11):935-49.
2. Kumar GP, Sharmila JS, Murugan S. Docking of ctx-m-9 group of enzymes with drugs and inhibitors and their evolutionary relationship. Asian J Pharm Clin Res 2014;7(1):237-42.
3. Kirchmair J, Markt P, Distinto S, Wolber G, Langer T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection – What can we learn from earlier mistakes? J Comput Aided Mol Des 2008;22(3-4):213-28.
4. Daisy P, Niveda RP, Bakiya RH. *In silico* drug designing approach for biotin protein ligase of *Mycobacterium tuberculosis*. Asian J Pharm Clin Res 2013;6(1):103-7.
5. McGaughey GB, Sheridan RP, Bayly CI, Culberson JC, Kreatsoulas C,

Lindsley S, *et al*. Comparison of topological, shape, and docking methods in virtual screening. J Chem Inf Model 2007;47(4):1504-19.
6. Sutherland JJ, Nandigam RK, Erickson JA, Vieth M. Lessons in molecular recognition 2. Assessing and improving cross-docking accuracy. J Chem Inf Model 2007;47(6):2293-302.
7. Yuriev E, Agostino M, Ramsland PA. Challenges and advances in computational docking: 2009 in review. J Mol Recognit 2011;24(2):149-64.
8. Venkatachalam CM, Jiang X, Oldfield T, Waldman M. LigandFit: A novel method for the shape-directed rapid docking of ligands to protein active sites. J Mol Graph Model 2003;21(4):289-307.
9. Vidhya VG, Bhaskar A, Vijayasri S. Molecular docking analysis of sorbitol dehydrogenase using Ligandfit algorithm. Indian J Bioinform Biotechnol 2013;2(5):95-9.
10. Neelakantan S. Structure based drug designing of p38 map kinase inhibitors for the treatment of osteoarthritis. Rasayan J Chem 2011;4(1):210-6.
11. Thangapandian S, Krishnamoorthy N, John S, Sakkiah S, Lazar P, Lee Y, *et al*. Pharmacophore modeling, virtual screening and molecular docking studies for identification of new inverse agonists of human histamine H1 receptor. Bull Korean Chem Soc 2010;31(1):52-8.
12. Taufer M, Crowley M, Price D, Chien AA, Brooks CL. Study of a highly accurate and fast protein-ligand docking algorithm based on molecular dynamics. Concurr Comput 2005;17(14):1627-41.
13. Akdo˘gan ED, Erman B, Yelekc K. *In silico* design of novel and highly selective lysine-specific histone demethylase inhibitors. Turk J Chem 2011;35:523-42.
14. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267(3):727-48.
15. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins 2003;52(4):609-23.
16. Hauke J, Kossowski T. Comparison of values of Pearson's and spearman's correlation coefficients on the same sets of data. Quaestiones Geogr 2011;30(2):87-93.
17. Swann SL, Brown SP, Muchmore SW, Patel H, Merta P, Locklear J, *et al*. A unified, probabilistic framework for structure- and ligand-based virtual screening. J Med Chem 2011;54(5):1223-32.
18. Rastelli G, Pacchioni S, Sirawaraporn W, Sirawaraporn R, Parenti MD, Ferrari AM. Docking and database screening reveal new classes of *Plasmodium falciparum* dihydrofolate reductase inhibitors. J Med Chem 2003;46(14):2834-45.