

A tool for the post data analysis of screened compounds derived from computer-aided docking scores

Gali Nageswara Rao¹, Allam Appa Rao², Peri Srinivasa Rao¹ & Naresh Babu Muppalaneni^{2*}

¹Department of CS & SE, Andhra University College of Engineering; ²C R Rao Advanced Institute of Mathematics, Statistics & Computer Science, Prof. C R Rao Road, Gachibowli, Hyderabad-500046, INDIA; Naresh Babu Muppalaneni - Email: nareshmuppalaneni@gmail.com; *Corresponding author

Received December 17, 2012; Revised December 28, 2012; Accepted January 03, 2013; Published February 21, 2013

Abstract:

A method is described for the analysis of the results obtained from the docking studies applied on a protein target and small molecules chemical compounds as ligands from various sources using different docking tools. We show the use of Dempster Shafer Theory (DST) to select the high ranking top compounds for further analysis and consideration.

Availability: Application is freely available at <http://allamapparao.org/dst/>

Keywords: Docking, Dempster shafer theory, Theory of evidence, Virtual screening.

Background:

The number of compounds available for high throughput screening (HTS) has dramatically increased. However, large-scale random combinatorial libraries have made limited contributions in the identification of leads in drug discovery projects. Therefore, the concept of 'drug-likeness' of compound selections has become a focus in recent years. The low success rate of converting lead compounds into drugs often due to unfavorable pharmacokinetic parameters has sparked a renewed interest in understanding more clearly what makes a compound drug-like [1].

Docking predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules using scoring functions [2]. We show the use of Dempster Shafer Theory (DST) [3] on docking results using an example structure with PDB ID: 3CL2 (crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants).

In DST, evidence can be associated with multiple possible events, e.g., sets of events. As a result, evidence in DST can be meaningful at a higher level of abstraction without having to resort to assumptions about the events within the evidential set. Where the evidence is sufficient enough to permit the assignment of probabilities to single events, the Dempster-Shafer model collapses to the traditional probabilistic formulation [4]. Combination rules are the special types of aggregation methods for data obtained from multiple sources.

These multiple sources provide different assessments for the same frame of discernment and Dempster-Shafer theory is based on the assumption that these sources are independent. The requirement for establishing the independence of sources is an important philosophical question. Lianwen Zhang [5] also provides an alternative combination rule to Dempster's rule. With respect to the rule of combination, Zhang points out that Dempster's rule fails to consider how focal elements intersect [5]. To define an alternative rule of combination, he introduces a measure of the intersection of two sets A and B assuming finite

sets. This is defined as the ratio of the cardinality of the intersection of two sets divided by the product of the cardinality of the individual sets.

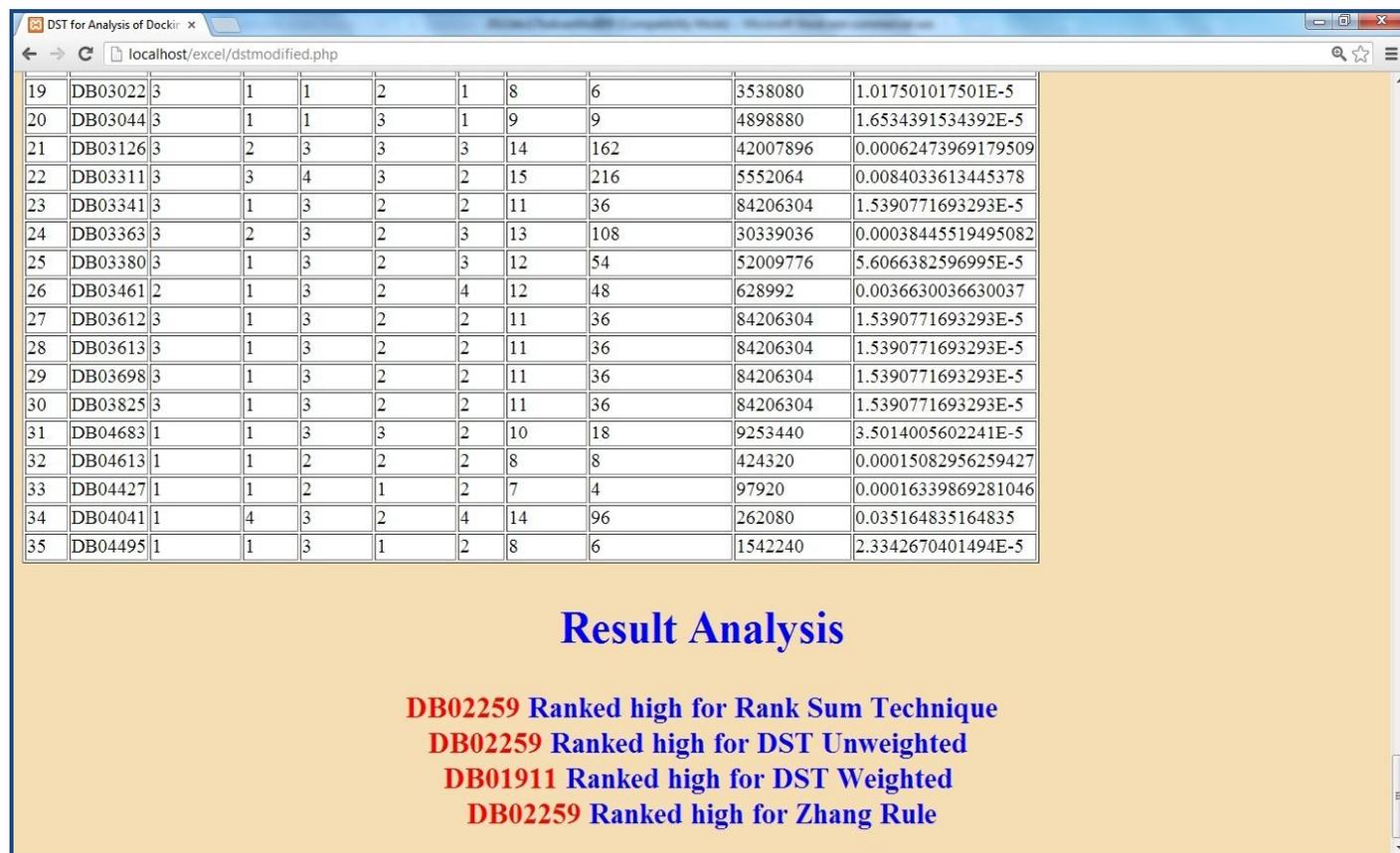


Figure 1: Selection of top 35 compounds using Dempster Shafer Theory (DST) data analysis for the target with PDB ID: 3CL2 (crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants) from an original set of 4,347 compounds from Drug Bank [http://www.drugbank.ca/] post docking and scoring. This procedure thus helps to identify molecules with ID numbers DB02259 and DB01911 with good binding features with the target for further consideration and analysis.

Methodology:

Docking, visualization, scoring and selection

It is important to visualize the docked poses of high-scoring compounds because many ligands are docked in different orientations and may often miss interactions that are known to be important for the target receptor. This sort of study becomes more difficult as the size of the dataset increases. Therefore, an alternative approach is to eliminate unpromising compounds before docking by restricting the dataset to drug-like compounds; by filtering the dataset based on appropriate property and sub-structural features and by performing diversity analysis [6]. Consensus scoring combines information from different scores to balance errors in single scores and improve the probability of identifying 'true' ligands [7]. In our study, we have tested five different scoring functions as used in tools such as: (i) GOLD [8]; (ii) Patchdock [9]; (iii) Molegro[10]; (iv) MEDock [11]; (v) AutodockVina [12]. The input for this application is Spread Sheet with an extension of .xls. The spread sheet consists of docking results of various compounds from various docking tools. The uploaded file is parsed and the data is stored in 2D array. We then use DST in a 5 step procedure. The steps used are: (1) Divide the data into 4 classes; (2) get results from Rank Sum Technique; (3) get results from DST unweighted; (4) get results from DST weighted; (5) get results from Zhang Rule. Compounds selected by steps 2 to 5 in the

above procedure will be considered for further analysis and investigation in the discovery pipeline.

Output data and analysis:

We used 4,347 experimentally determined ligand structures for this study. The target protein structure with PDB ID: 3CL2 was screened against these ligand structures using Molegro to select the optimal binding small molecules using binding scores. The top 35 compounds were further docked using Patchdock, GOLD, MEDock, AutodockVina. Analysis using Dempster Shafer Theory selects compounds having drug bank ID numbers DB02259 and DB01911 as best binding candidates with the protein target and for consideration in further investigations.

Future developments:

We plan to include other combination rules such as Yager's Rule and Inagaki's Rule in future developments.

Acknowledgement:

Dr. AllamAppaRao and Dr. M. NareshBabu would like to thank the Department of Science and Technology (DST) for their financial support (DST-CMS GoI Project No.SR/S4/MS:516/07 Dated 21.04.2008).

Reference:

- [1] Muegge I, *Med Res Rev.* 2003 **23**: 302 [PMID: 12647312]
[2] Lengauer T & Rarey M, *Curr Opin Struct Biol.* 1996 **6**: 402 [PMID: 8804827]
[3] Stuart C Shapiro, *Encyclopedia of Artificial Intelligence*, Second Edition, Wiley. 1992 pp: 330-331
[4] Glenn Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
[5] Ronald R Yager *et al.* *Advances in the Dempster-Shafer Theory of Evidence*. Wiley. 1994
[6] Waszkowycz B, *Drug Discov Today.* 2008 **13**: 219 [PMID: 18342797]
[7] Charifson PS *et al.* *J Med Chem.* 1999 **42**: 5100 [PMID: 10602695]
[8] http://www.ccdc.cam.ac.uk/products/life_sciences/gold/
[9] <http://bioinfo3d.cs.tau.ac.il/PatchDock/>
[10] <http://www.molegro.com/>
[11] <http://medock.csie.ntu.edu.tw/>
[12] <http://autodock.scripps.edu/>

Edited by P Kanguane

Citation: Rao *et al.* *Bioinformation* 9(4): 207-209 (2013)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited