Bioorganic & Medicinal Chemistry Letters xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

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

Performance evaluation of structure based and ligand based virtual screening methods on ten selected anti-cancer targets

Thilagavathi Ramasamy [a], Chelliah Selvam [b,*]

[a] Department of Biotechnology, Faculty of Engineering, Karpagam University, Coimbatore, India
[b] Department of Pharmaceutical Sciences, College of Pharmacy and Health Sciences, Texas Southern University, Houston, TX 77004, USA

ARTICLE INFO

Article history:
Received 4 June 2015
Revised 10 August 2015
Accepted 14 August 2015
Available online xxxx

Keywords:
Structure based
Ligand based
Virtual screening
Docking
DEKOIS
Anti-cancer targets

ABSTRACT

Virtual screening has become an important tool in drug discovery process. Structure based and ligand based approaches are generally used in virtual screening process. To date, several benchmark sets for evaluating the performance of the virtual screening tool are available. In this study, our aim is to compare the performance of both structure based and ligand based virtual screening methods. Ten anti-cancer targets and their corresponding benchmark sets from 'Demanding Evaluation Kits for Objective In silico Screening' (DEKOIS) library were selected. X-ray crystal structures of protein–ligand complexes were selected based on their resolution. Openeye tools such as FRED, vROCS were used and the results were carefully analyzed. At EF1%, vROCS produced better results but at EF5% and EF10%, both FRED and ROCS produced almost similar results. It was noticed that the enrichment factor values were decreased while going from EF1% to EF5% and EF10% in many cases.

Anti-cancer drug discovery is the main focus of many pharmaceutical industries. Several new biomolecular targets are being discovered due to increasing insights in molecular biology and genetics.[1,2] Among the targets identified, kinases are popular anti-cancer targets because they are druggable. According to the recent report,[3] the world market for anti-cancer kinase inhibitors will reach $18.5 billion in 2014. Despite the continuous efforts in the discovery and development of novel drug molecule, cancer is still a highly challenging disease.

Virtual screening and other computational methods play an important role in drug discovery processes.[4–6] Virtual screening methods are inexpensive because they do not use the chemicals and other experimental procedures which are involved in high throughput screening processes in drug discovery. From the collection of large library of compounds, it is possible to select a limited set of compounds. In the literature, there are impressive numbers of successful applications of such methods reported.[7,8]

Numerous software tools have been developed for the purpose of virtual screening. Virtual screening tools are often evaluated for their ability to enrich the fraction of the active ligands from the set of both active and decoys. The benchmark sets usually consist of known actives and for each actives a set of small decoys or inactive. To date, many benchmark sets are made available publically. One of the well-known benchmark set is Directory of Useful Decoys (DUD), a publically available data set of about 100 000 compounds distributed over 40 protein targets. The DUD set has the ligand decoys ratio of 1:36. Decoys are physically similar but topologically different to that of each active ligand.[9]

Maximum Unbiased Validation (MUV) data set[10] is another benchmark set which includes PubChem experimental data. Very recently 'Demanding Evaluation Kits for Objective In silico Screening' (DEKOIS) library[11,12] are made available. In this report, we present a comparative study of performances of both structure-based and ligand-based virtual screening approaches using openeye tools such as FRED and vROCS.

Ten anti-cancer targets were selected form DEKOIS library. Among them, seven targets belong to kinase family. They are RAC-alpha serine/threonine-protein kinase (PKB), Aurora A kinase, B-Raf, PI3-kinase gamma, pim-1, Rho-associated protein kinase-1 and vascular endothelial growth factor rec.2. Two targets belong to histone deacetylases and the other target was p53-binding protein MDM2.

Analysis and the selection of X-ray structures of protein–ligand complexes (Table 1) for the selected targets were done. Receptor grid was generated using the highest resolution structure. For the above mentioned targets, both active ligands and decoys were obtained from the web page.[12] Active ligands and decoys were mixed together[13] and were subjected to conformational analysis[14] using Omega2 program. Crystal structures of highest resolution were selected to generate receptor grids. Multiple conformers of

* Corresponding author. Tel.: +1 713 313 7552.
E-mail address: chelliahs@tsu.edu (C. Selvam).

http://dx.doi.org/10.1016/j.bmcl.2015.08.040
0960-894X/

Table 1
Cancer targets and pdbs selected for the study

No	Target	No of pdb selected	Pdb codes of selected X-ray structures	Selected pdbs for docking (resolution)
1	RAC-alpha serine/threonine-protein kinase (PKB)	7	4GV1, 4EJN, 3QKM, 3QKL, 3QKK, 3CQW, 3CQU	4GV1 (1.49)
2	Aurora A kinase	12	3FDN, 2NP8, 3UP2, 3K5U, 3MYG, 3UO4, 4DEA, 3EFW, 3H10, 3O50, 3UOL, 3W2C	3FDN (1.90)
3	Serine/threonine-protein kinase B-Raf	4	3II5, 3D4Q, 2FB8, 1UWH	3II5 (2.79)
4	Histone Deacetylase 2	3	4LY1, 4LXZ, 3MAX	4LY1 (1.57)
5	Histone Deacetylase 8	7	1T64, 2V5X, 1VKG, 1T69, 3SFH, 1T67, 3SFF	1T64 (1.90)
6	p53-Binding protein MDM2	12	4OGN, 4OGT, 3TU1, 4ODE, 4MDN, 3LBL, 3W69, 4ERE, 4JVR, 4OAS, 4OBA, 4OCC	4OGN (1.38)
7	Phosphoinositid-3-kinase gamma	8	4ANV, 4ANW, 4HVB, 3ENE, 4FUL, 2CHX, 2V4L, 3ZVV	4ANV (2.13)
8	Serine/threonine-protein kinase pim-1	13	3R04, 4DTK, 2C3I, 2J2I, 3F2A, 4BZN, 4ALW, 3R02, 4K0Y, 1YXV, 1YXX, 3BGQ, 3C4E	3R04 (1.70)
9	Rho-associated protein kinase-1	8	3V8S, 3TWJ, 3TV7, 3NDM, 3NCZ, 2ETR, 2ETK, 2ESM	3V8S (2.29)
10	Vascular endothelial growth factor rec.2	17	2XIR, 3VO3, 3VHE, 3EWH, 3VNT, 1YWN, 3BE2, 2P2H, 4AG8, 4AGC, 4ASD, 2OH4, 1Y6A, 1Y6B, 3C7Q, 3CJF, 2QU6	2XIR (1.50)

each compound were then subjected to docking and the resulted hits were analyzed. The percentage enrichment factors were calculated and the results are given in the Table 2.

vROCS program was used to perform ligand based virtual screening. ROCS stands for Rapid Overlay of Chemical Structures. Successful applications of ROCS are reported and it is considered as the industry standard for ligand shape-based virtual screening.[15–20] Screening performance of ROCS is highly dependent on the selection of query molecule.[18,21] Ligand query models were generated using a set of bound conformations of the ligands after superposing the protein–ligand complexes and extraction of bound ligands. The superposed structures for some selected targets are shown in Figure 1a–g.

From the experimental conformations of the bound ligands supplied, vROCS produced possible ligand query models by choosing best three ligands. The ligand query models obtained by vROCS are shown in Figure 2a–h. The query showed the shape and important features. Green sphere represents the ring structure, red sphere represents the acceptor feature, yellow represents the hydrophobic feature and the blue sphere represents the donor feature (Fig. 2a–h).

vROCS was used to screen the dataset. The ligand query models generated for each targets were used for the virtual screening. Based on the Tanimoto similarity index, the ligands were scored. The top scored ligands were visualized and the number of actives found was used in the calculation of enrichment studies.

Comparison of the virtual screening results obtained from vROCS and FRED[22] revealed that at EF1% (Enrichment Factor[23] at 1%) vROCS produced better results for six out of ten targets.

Both FRED and ROCS did not show any hits for Histone deacetylase 8 at EF1%. However, similar enrichment factor values were obtained out of both FRED and vROCS method for the targets p53-binding protein MDM2 and Rho-associated protein kinase-1. Although at EF1%, the vROCS appeared to produce better results but at EF5% (Enrichment Factor at 5%) and EF10% (Enrichment Factor at 10%) the results were almost comparable in most of the cases. It was observed that the enrichment factor values were decreased while going from EF1% to EF5% and EF10% in many cases.

In conclusion, we have evaluated two openeye tools such as FRED, and vROCS, for their performance in virtual screening studies by taking 10 anticancer targets and the recently developed benchmark set 'DEKOIS'. Most of the selected targets belong to kinases including RAC-alpha serine/threonine-protein kinase (PKB), Aurora A kinase, B-Raf, PI3-kinase gamma, pim-1, Rho-associated protein kinase-1 and vascular endothelial growth factor rec.2. Two targets belong to histone deacetylases and the other target was p53-binding protein MDM2.

The comparison between ligand-based (vROCS) and structure-based method (FRED) demonstrated that the ligand based methods performed superior and thereby it has yielded higher enrichment during early retrieval of active compounds at EF1% for 6 out of 7 kinases selected. Both the methods performed equally well for the Rho-associated protein kinase at EF1%. Among the 7 kinases, vROCS yielded a very good enrichment factor, that is, 28.4 for B-Raf. Structure based method performed well for B-Raf, p53-binding protein MDM2, and VEGFR2 targets at 1% enrichment level. However comparison of both methods revealed that ligand based virtual screening provided excellent enrichment factor values at EF1%. Moreover, the results revealed the superiority of the ligand-based method vROCS in terms of both speed and hit enrichment. We also observed a trend that the enrichment factor values were decreased while going from EF1% to EF5% and EF10% in many cases.

DEKOIS includes several new targets that allow us to choose important anti-cancer targets including seven kinase targets. Furthermore, the results demonstrated the usefulness of the ligand

Table 2
Comparison of structure based and ligand based methods: Enrichment factors calculated at 1, 5 and 10 percentage levels

Target	EF1%		EF5%		EF10%	
	Structure based	Ligand based	Structure based	Ligand based	Structure based	Ligand based
RAC-alpha serine/threonine-protein kinase(PKB)	7.8	10.3	5.0	4.0	2.8	2.8
Aurora A kinase	7.8	23.3	4.5	7.0	4.0	4.0
Serine/threonine-protein kinase B-Raf	10.3	28.4	3.0	9.6	2.3	7.3
Histone Deacetylase 2	5.2	10.3	6.0	2.5	5.0	2.3
Histone Deacetylase 8	0	0	2.5	1.0	2.8	1.0
p53-Binding Protein MDM2	10.3	10.3	13.0	12.5	8.3	8.5
Phosphoinositid-3-kinase gamma	7.8	12.9	5.0	8.0	3.8	5.3
Serine/threonine-protein kinase pim-1	7.8	10.3	6.5	3.0	4.8	2.5
Rho-associated protein kinase-1	5.2	5.2	3.5	3.5	1.8	3.8
Vascular endothelial growth factor rec.2	10.3	15.5	5.5	5.0	3.5	3.8

Please cite this article in press as: Ramasamy, T.; Selvam, C. *Bioorg. Med. Chem. Lett.* (2015), http://dx.doi.org/10.1016/j.bmcl.2015.08.040