Life Science Informatics Publications
Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences
Journal Home page http://www.rjlbpcs.com/
RJLBPCS
ISSN 2454-6348

Original Research Article **DOI - 10.26479/2018.0403.18**

COMPARATIVE ANALYSIS OF BLIND DOCKING REPRODUCIBILITY

Swati Sharma, Iti Garg, Bhuvnesh Kumar, Mohammad Zahid Ashraf*

Defence Institute of Physiology and Allied Sciences, Delhi, India

ABSTRACT: Virtual screening methods of computer aided drug discovery (CADD) has revolutionised the drug discovery process. It has not only paced the discovery of new drugs but also reduced the cost needed for it. Cost effectiveness and time saving attributes has made CADD a popular choice in lead identification. But these methods are also prone to certain errors. Blind docking (BD) is a popular choice for analysis binding of ligand with the protein. But can blind docking be relied upon blindly. The present work is aimed at giving insight about the effectiveness of BD approach and its reproducibility. AutodockVina was used for blind docking of three natural compounds viz. poncirin, quercitrin and squalene with factor Xa. Each compound was docked five times to assess the reproducibility of protocol in terms of interacting residues of proteins and binding affinity. BD being influenced by exhaustiveness of ligand was also analysed by using two values of exhaustiveness i.e. 8 and 32 in an otherwise same protocol. Overall, this study indicates that it is easy to generate in silico data but it is required to validate and integrate data in the manner that generate reproducible information.

KEYWORDS: Blind Docking, Computational biology, Computer Aided Drug Discovery, Exhaustiveness

Corresponding Author: Dr. Mohammad Zahid Ashraf Ph.D.
Defence Institute of Physiology and Allied Sciences, Delhi, India
Presently at Jamia Milia Islamia, Central University, Delhi
Email Address: mohammadzashraf@gmail.com

1.INTRODUCTION

The in silico process of molecular docking has paced the drug designing process, therefore is widely used in discovery and development of new lead compounds [1]. It is also termed as computer aided drug discovery (CADD) and has emerged as a tool for understanding the intricacies of interaction between a molecular target (usually a protein) and a ligand [2,3]. As compared to conventional drug discovery process, CADD is not only reliable and cost effective, but saves considerable time needed for lead compound identification [4,5,6]. In last decade CADD has gained popularity in terms of use as evident by the quote "The future is bright, the future is virtual" given by a review [7]. This area of research has experienced fast expansion mainly due to the progress in sophistication of computational power and increasingly available dataset of ligands and protein targets from public repositories. All the steps of drug discovery process from identifying hits (probable drug candidate) to selecting leads (most apt hit for further evaluation) and optimizing leads can be performed by CADD. Lead optimization involves improving the physicochemical, pharmaceutical, ADMET/PK (pharmacokinetic) properties of the lead identified so that it can be used as suitable drug [8] The virtual screening has significantly decreased the need of chemical synthesis and biological testing thus minimizing the resource requirements and utilisation [8]. The mentioned attributes has made CADD an integral part of drug discovery process by pharmaceutical companies and research labs. Structure based drug discovery is easiest approach used in molecular docking as it screens the compounds (ligands) listed in chemical library by "docking" them against the target protein of known structure providing the details of binding affinity and conformation of ligands [9]. The docking of ligands can be targeted focusing only on the predicted binding sites of the target protein or it can be BD that covers the entire area of a protein. BD is considered to be unbiased as it scans the entire structure of the protein for finding out the putative binding site of ligand. Several computational tools are available for this purpose like AutoDock (AD), AutoDockVina (Vina) [10, 11, 12], DOCK [13], FlexX [14] and GOLD [15]. AutoDock programs are most widely used as they are excellent non-commercial docking program freely available for academic use. Both AD and Vina work more or less in same way such as in pairing an empirically-weighted scoring function with a global optimization algorithm, but differ in parameterization of the scoring function and local search function. Vina as compared to AD is considered to be more quick and accurate for re-docking protein-ligand complexes [12]. Also, it is considered to be more consistent and better than other docking programs for BD pose prediction [16]. Therefore, use of Vina for BD of ligands with target protein is a very good tool that can be used by researchers in CADD. The chances of success in different stages of drug discovery process increases by use of complementary experimental and informatics techniques, along with docking tools. The use of Vina for virtual screening is constrained theoretically only by ligand properties (rotatable bonds, hydrogen bond acceptor and donor) that can

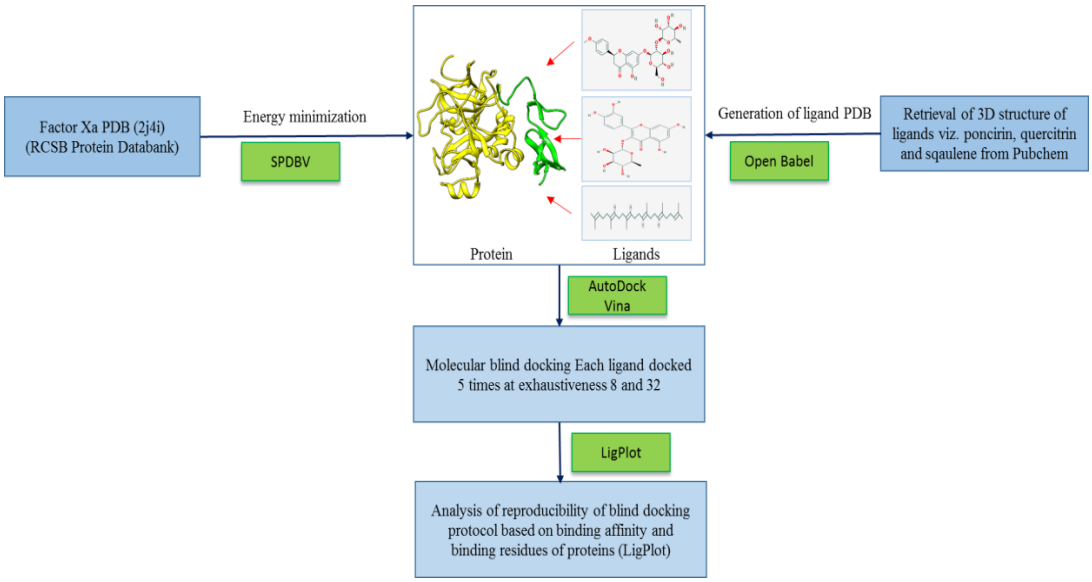
be calculated and relation between these properties and target protein [17]. But the major concern arises in practical implementation of Vina in molecular docking of ligands that requires several considerations. First and foremost is the reproducibility of BD protocol. As, in BD the ligand is free to acquire different poses or conformation on the entire grid area covering the protein structure since it allows more freedom of movement to ligand, thereby decreasing the probability of getting the same pose after every run. The present study is done to bridge the gap between the theoretical and practical implementation of Vina. The experimental approach used consists of five independent docking runs carried out on grid boxes centred on the macromolecule, at exhaustiveness 8 and 32. Poncirin, quercitrin and squalene are used as ligands to be docked with the protein factor Xa. Factor Xa is a serine protease involved in blood coagulation cascade. The ligands and protein are chosen at random only as examples to assess the BD protocol for its reliability and reproducibility. Selecting of known inhibitor of factor Xa could have resulted in a biased study. The same docking protocol was used for each independent run, but at different levels of exhaustiveness.

2. MATERIALS AND METHODS

Figure 1 exhibits the workflow of methodology adopted in present study as described below.

Fig 1. Workflow of blind docking analysis

Flow chart depicting the methodology and software adopted to conduct blind docking by integrating data acquiesced from different database.



Protein structure, data acquisition and preparation. The crystallographic structure file (PDB = 2J4I) for human factor Xa, was retrieved from the RCSB Protein Data Bank (PDB) [18]. The bound ligand and all crystallographic water molecules were removed from 3D structure of protein prior to BD. Energy minimization of protein structure was done using SPDBV- Swiss-Pdb Viewer [19]. Compounds structure acquisition and preparation. Three natural compounds viz. poncirin, quercitrin

and squalene were chosen randomly as ligands for BD study. The sdf files of 3D structure, of the three compounds were downloaded from PubChem, an open chemistry database. The sdf files were converted into pdb format prior to docking by using Open Babel software [20]. The BD of ligands with factor Xa was done by using AutoDock Vina that for global optimization for local minima search uses iterated local search global optimizer [21]. For docking the protein was kept as rigid moiety, while ligands were kept flexible. Each compound was docked 5 times with the target protein factor Xa, at exhaustiveness 8 and 32 to find out the relative binding affinity of a ligand at different hits (n=5) and assess the reproducibility of the protocol. Factor Xa (PDB ID = 2J4I), was processed in AutoDock tools (ADT) [22]. The polar hydrogens were added, Kollman charges were assigned to all atoms, and Gasteiger charges were calculated for the protein. ADT was also used to process the ligand PDB files of all three compounds. The polar hydrogens were added and Gasteiger charges were calculated for ligand PDB files. The rigid root and the number of rotatable bonds were defined by the Autotors tool of ADT. The BD was performed with affinity grid maps of 66 x 46 x 56 points, 66 x 46 x 55 points and 62 x 44 x 56 point for docking of poncirin, quercitrin and squalene respectively with 2J4I. The grid point spacing used is 1.00 □ centred on whole protein structure encompassing the active site of protein (catalytic domain) using the autogrid tool of ADT. The instructed command prompts were used to perform the docking procedure. Each ligand was docked against the protein at exhaustiveness 8 and 32. The ligands were docked five times at each exhaustiveness level using the same grid points. The minimum energy conformation state of ligands showing binding affinity in kcal/mol was taken into consideration. The protein ligand interaction for different hits was viewed by UCSF Chimera [23] by keeping protein as rigid molecule and overlaying each ligand at all five poses obtained after docking over it. The best docked pose with lowest energy conformation was selected from different hits, for all the ligands. The best pose image of ligand and protein bound complexes were prepared in Ligplot visualizing program [24] and the result included position of hydrogen bond formed between them.

3. RESULTS AND DISCUSSION

The details of the ligands viz. poncirin, quercitrin and squalene, along with their 2D structures are briefed in Table 1. The PubChem Id, chemical formula, molecular weight, hydrogen bond acceptors (HbA), hydrogen bond donors (HbD), number of rotational bonds (nRB), partition coefficient (LogP) and the 2-Dimensional (2D) structures of compounds employed in present work. The information is obtained from PubChem database. The different binding conformations and pose obtained after BD (number of hit =5) of ligands with factor Xa at exhaustiveness 8 are represented in Fig. 2a. The best docking conformation with binding affinity of -9.3 Kcal/mol, -9.1 Kcal/mol and -7.4 Kcal/mol (Fig. 2b) is selected for poncirin, quercitrin and squalene respectively.

Table 1. Structure information of ligands.

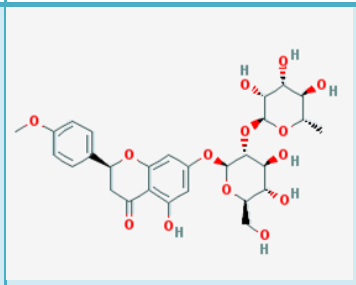
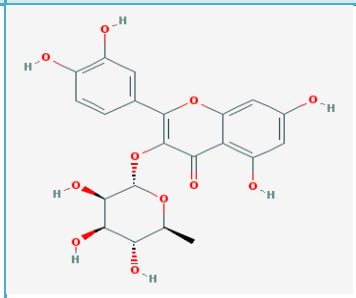
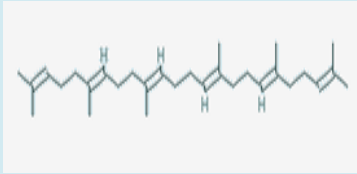
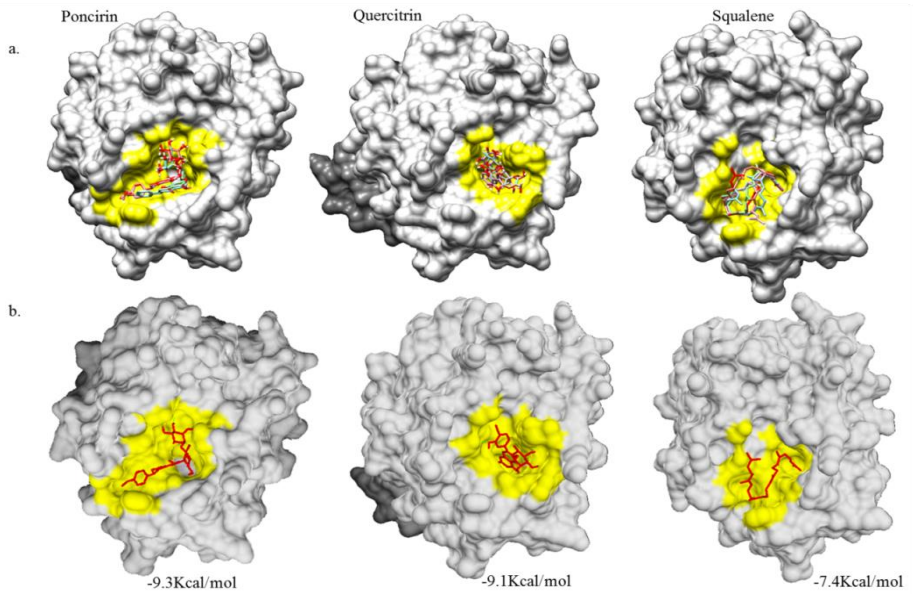
Compound Name	PubChem Id	Chemical Formula	Mol. Wt. (gm)	HbA	HbD	nRB	LogP	2D Structure
Poncirin	442456	$C_{28}H_{34}0_{14}$	594.56	14	7	7	-0.2	
Quercitrin	5280459	$C_{21}H_{20}O_{11}$	448.38	11	7	3	0.9	
Squalene	638072	$C_{30}H_{50}$	410.72	0	0	13	11.6	

Fig 2. Blind docking at exhaustiveness 8

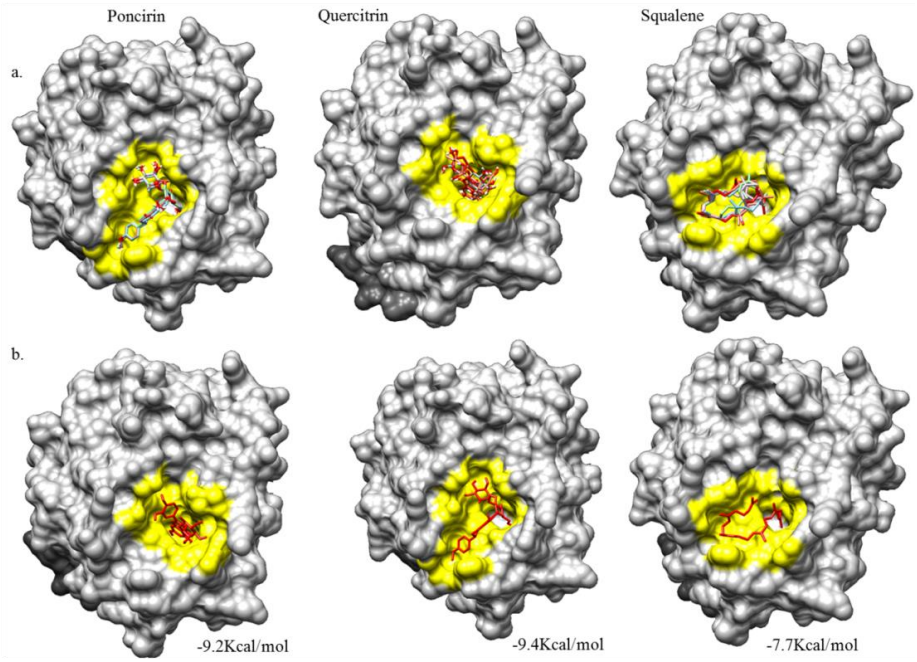
(a) Different binding conformations of poncirin, quercitrin and squalene with Factor Xa as viewed by UCSF Chimera (b) Best binding pose of poncirin, quercitrin and squalene with Factor Xa upon based binding affinity



The same experimental protocol of docking at exhaustiveness 32 was used to find out whether it increases the reliability over protocol. Fig. 3(a) and 2(b) elucidate the different binding poses and best binding pose of ligands at exhaustiveness 32.

Fig 3. Blind docking at exhaustiveness 32

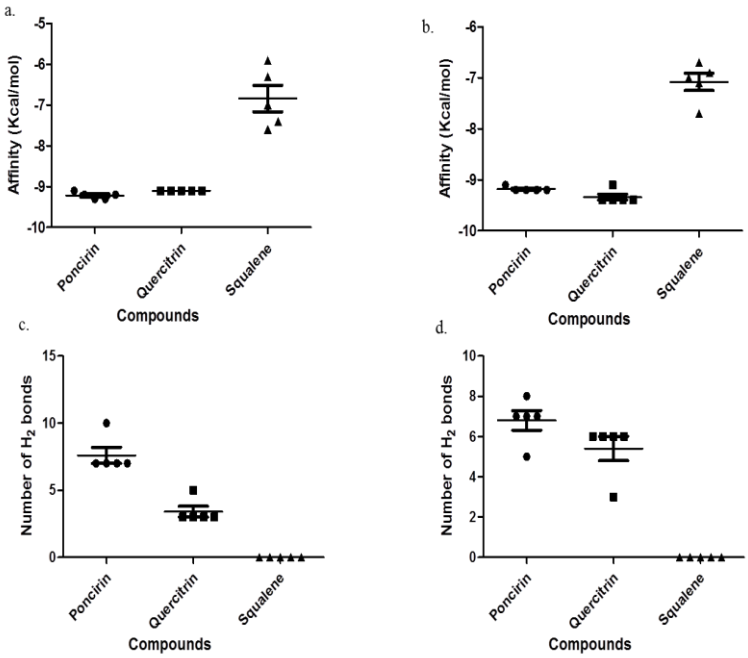
(a) Different binding conformations of poncirin, quercitrin and squalene with Factor Xa as viewed by UCSF Chimera (b) Best binding pose of poncirin, quercitrin and squalene with Factor Xa upon based binding affinity.



The best pose binding affinity of poncirin, quercitrin and squalene with factor Xa at exhaustiveness 32 is -9.2 Kcal/mol, -9.4 Kcal/mol and -7.7 Kcal/mol respectively. In order to assess which protocol is better in terms of reproducibility a comparison is drawn between the binding affinities (Fig. 4a and 4b) number of hydrogen bonds formed after docking (Fig. 4c and 4d) at exhaustiveness 8 and 32.

Fig 4. Binding affinity and number of hydrogen bond comparison

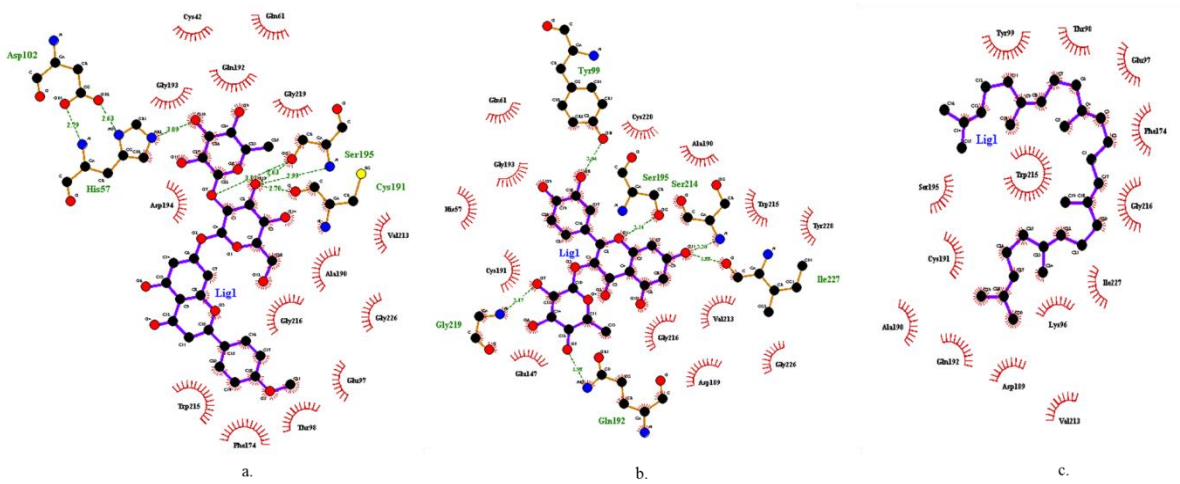
(a, b) Binding affinity of poncirin, quercitrin and squalene with Factor Xa measured in Kcal/mol at exhaustiveness 8 and 32 respectively (hit =5). (c, d) Number of hydrogen bonds formed by poncirin, quercitrin and squalene with Factor Xa at exhaustiveness 8 and 32 respectively (hit =5).



Based upon the above comparison best docking pose for poncirin (Fig. 5a), quercitrin (Fig. 5b) and squalene (Fig. 5c) was used to represent LigPlot analysis comprising of hydrogen bond formed between the ligand and protein.

Fig 5. LigPlot view

LigPlot view of best binding conformation of poncirin, quercitrin and squalene to factor Xa



The summary of LigPlot analysis of all three ligands at different hits (n=5), at exhaustiveness 8 and 32 is given in Table 2.

Table 2. Interacting residue details as obtained by LigPlot. Hydrogen bond formed by ligands with interacting residues of factor Xa, at different hits and level of exhaustiveness.

COMPOUND	RESIDUES FORMING HYDROGEN BOND	
	Exhaustiveness = 8	Exhaustiveness = 32
Poncirin	Ser195=3, His57=1, Asp102=2, Cys191=1	Ser195=2, His57=1, Asp102=2, Cys191=1 Asp189=1, Tyr99=1
	Ser195=2, Ser214=1, Gly219=1, Ala190=1 Cys191=1, Asp189=1	Ser195=3, His57=1, Asp102=2, Cys191=1
	Ser195=3, His57=1, Asp102=2, Cys191=1	Ser195=3, Cys191=1, Tyr99=1
	Ser195=2, His57=1, Asp102=2, Cys191=1, Ser214=1, Gly219=1, Ala190=1	Ser195=3, His57=1, Asp102=2, Cys191=1
	Ser195=3, His57=1, Asp102=2, Cys191=1	Ser195=3, His57=1, Asp102=2, Cys191=1
Quercitrin	Ser195=2, Gln192=1	Ser195=1, Gly219=1 Tyr99=1, Ser214=1 Gln192=1, Ile227=1
	Ser195=2, Gln192=1	Ser195=1, Gly219=1 Tyr99=1, Ser214=1 Gln192=1, Ile227=1
	Ser195=1, Gly219=2 Tyr99=1, Asp189=1	Ser195=1, Gly219=1 Tyr99=1, Ser214=1 Gln192=1, Ile227=1
	Ser195=2, Gln192=1	Ser195=1, Gly219=1 Tyr99=1, Ser214=1 Gln192=1, Ile227=1
	Ser195=2, Gln192=1	Ser195=2, Gln192=1

The aim of the in silico molecular docking is to predict the best ligand and target (enzyme or receptor) conformation and orientation. This is a two-step process, first the information about the different ligand conformations in the active site of protein is gathered and then these conformations are ranked based upon the scoring function for predicting binding affinity for each individual orientation [25]. While most of the applications of molecular docking tools concentrate on the predicted primary binding region, BD was introduced to scan the entire area of the target protein for binding sites and simultaneously optimizing the conformation of ligands [26]. The results obtained after BD were considered to be encouraging [27] and it has been used for answering various problems like that of designing inhibitor [28], comparing microtubule stabilizing agents [29] and finding out substrate binding modes [30]. But can BD be used blindly for drug discovery process and these results can be relied? To answer this we performed the present work by BD of three randomly chosen ligands poncirin, quercitrin and squalene with protein factor Xa as a target. The PDB structure used for factor Xa is 2J4I [18]. For BD two values of exhaustiveness were chosen 8 (default in Vina) and 32 to validate the effect of increase in exhaustiveness on result quality [31].
At exhaustiveness 8 and 32, different docking pose after BD (hit=5) was analysed (Fig. 1 and 2). It was observed that the different docking poses for each ligand at different hit partially overlap each other and sits in the same cavity that correspond to active site of protein. The difference in each overlap was analysed by comparing the number of hydrogen bond formed and difference in binding affinity of each ligand at different hits (Fig. 3). The binding affinity of poncirin varied and was -9.2Kcal/mol for two hits, -9.3Kcal/mol for two hits and -9.1 for one hit at exhaustiveness 8. The reproducibility in its binding increased at exhaustiveness 32 as the binding affinity was -9.2 for four hits and -9.1 for one hit. So, the best docking pose for poncirin will be the one with affinity -9.2Kcal/mol. The variation in number of hydrogen bond after each hit was more at exhaustiveness 32. This is because higher the exhaustiveness more vigorously the ligand will screen the protein for binding space. Quercitrin has only three rotatable bonds and hence can acquire less number of conformations. There by showed constant binding affinity of -9.1Kcal/mol at every hit at exhaustiveness 8. The binding affinity increased at exhaustiveness 32 due to more tight binding (-9.4Kcal/mol for four hits and -9.1Kcal for 1 hit). The number of hydrogen bond formed also increased from 3 to 6 after changing the level of exhaustiveness. The best pose selected was at exhaustiveness 32 with affinity of -9.4Kmol/cal and 6 hydrogen bonds. Squalene is an aliphatic molecule thus have more rotatable bonds (n=15). Thereby the binding affinity varied greatly at every hit at both exhaustiveness levels. However it was distributed more widely around the mean at exhaustiveness 8. Squalene lack hydrogen bond donor and acceptor therefore did not form any bond with protein at all. Based upon the above analysis best docking pose of poncirin, quercitrin and squalene among the various hits and level of exhaustiveness is selected and showed in Figure 4.

4. CONCLUSION

The study can be concluded on a note that though BD is an excellent approach, but its search parameters need to be checked and standardized before use. The reliability of the results obtained after BD can be increased by increasing exhaustiveness value. Though increase in exhaustiveness will slow down the speed of getting results as the protein surfaced would be screened more variously to provide the best docking pose of ligand. So, BD is a promising tool for obtaining the leads for drug discovery process but it can be trusted blindly only after setting the search parameters with great scrutiny. The results and data obtained after it need to be validated by more sophisticated in silico tools and wet lab assays.

5. ACKNOWLEDGEMENT

The authors are extremely thankful to Dr. Bhuvnesh Kumar (Director, DIPAS) and Dr. Lilly Ganju for their support during the study. We are grateful to Dr. Shashi Bala Singh, Directorate General Armed Forces Medical Sciences. SS is a Senior Research Fellow of the Council for Scientific and Industrial research (CSIR). This study was funded by DRDO, project DIP-263.

.6. CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Samy GB and Xavier L. Molecular docking studies on antiviral drugs for SARS. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 2015; 5, 75–79.
2. Anderson AC. The process of structure based drug design. Chem Biol 2003;10: 787–97
3. Schneider G. Virtual screening: an endless staircase? Nat Rev Drug Discov 2010; 9: 273–6.
4. Walters W, Stahl M and Murcko M. Virtual screening - An overview. Drug Discov Today 1998; 3:160-78.
5. Schneider G and Böhm H. Virtual screening and fast automated docking methods: combinatorial chemistry. Drug Discov Today 2002; 7: 64-70.
6. Waszkowycz B, Perkins T and Sykes R, Li J. Large-scale virtual screening for discovering leads in the postgenomic Era. IBM Systems J 2001; 40:360-76.
7. Shankar R, Frapaise X and Brown B. LEAN drug development in R&D. Drug Discov Development 2006:57–60.
8. Kapetanovic IM. Computer-aided drug discovery and development (caddd): in silico-chemico-biological approach. Chem Biol Interact. 2008; 171(2):165-76.
9. Kitchen DB, Decornez H, Furr JR and Bajorath J Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 2004; 3: 935–949.

10. Morris GM, Goodsell DS, Halliday RS, Huey R, and Hart WE, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry, 1999; 19: 1639–1662.
11. Huey R, Morris GM, Olson AJ and Goodsell DS. A semiempirical free energy force field with charge-based desolvation. J Comput Chem 2007; 28: 1145–1152.
12. Trott O, Olson AJ AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010; Jan 30; 31(2): 455–461.
13. Ewing TJ, Makino S, Skillman AG and Kuntz ID. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des 2001; 15: 411–428.
14. Rarey M, Kramer B, Lengauer T and Klebe G A fast flexible docking method using an incremental construction algorithm. J Mol Biol 1996; 261: 470– 489.
15. Jones G, Willett P, Glen RC, Leach AR and Taylor R Development and validation of a genetic algorithm for flexible docking. J Mol Biol. 1997; 267: 727– 748.
16. Azam SS and Abbasi, SW. Molecular docking studies for the identification of novel melatoninergic inhibitors for acetylserotonin O-methyltransferase using different docking routines. Theor. Biol. Med. Model. 2013; 10, 63.
17. Lazarova M. Virtual screening-models, methods and software systems. International Scientific Conference Computer Science 2008; 55-60.
18. Hetényi C, Van der, SD Young, Campbell RJ., Borthwick M. and Brown AD et al. Structure- and Property-Based Design of Factor Xa Inhibitors: Pyrrolidin-2-Ones with Acyclic Alanyl Amides as P4 Motifs. Bioorg. Med. Chem. Lett. 2006; 16: 5953.
19. Guex, N and Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 1997; 18, 2714-2723.
20. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, and Hutchison GR. "Open Babel: An open chemical toolbox." J. Cheminf. (2011), 3, 33.
21. Trott O, Olson AJ AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010; 31(2): 455–461.
22. Morris GM, Ruth H, Lindstrom W, “Software news and updates AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility,” Journal of Computational Chemistry, 2009; 16: 2785–2791,.
23. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004; 25(13):1605-12.

24. Wallace AC, Laskowski RA and Thornton JM, "LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions,"ProteinEngineering, 1995; 8:127–134.
25. Meng XY, Zhang HX, Mezei M and Cui M. Molecular docking: a powerful approach for structure-based drug discovery. Curr. Comput. Aided Drug Des. 2011; 7, 146
26. Hetényi C, Van der Spoel D Efficient docking of peptides to proteins without prior knowledge of the binding site.. Protein Sci. 2002; 11(7):1729-37.
27. Stephen J Campbell, Nicola D Gold, Richard M Jackson and David R Westhead. Ligand binding: functional site location, similarity and docking. Current Opinion in Structural Biology. 2003; 13, Pages 389–395.
28. Liu J, Irwan M, Helena YM, Vincent TKC, Teruna JS and Seetharama DSJ. Design, structure and biological activity of b-turn peptides of CD2 protein for inhibition of T-cell adhesion. Eur. J. Biochem. 2004; 271, 2873–2886.
29. Oriol P, Jaume F, Laura M, Fabrizio M, Maurizio B, and Jaume V. Computational comparison of microtubule-stabilising agents laulimalide and peloruside with taxol and colchicine. Bioorganic & Medicinal Chemistry Letters. 2004; 14(19):4825-9.
30. Bjelic S and Aqvist J. Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site. Biochemistry. 2004; 43(46):14521-8.
31. http://vina.scripps.edu/manual.html