Molecular Science

AIMS Molecular Science, 3(2): 238-245.
DOI: 10.3934/molsci.2016.2.238
Received 24 March 2016,
Accepted 15 May 2016,
Published 23 May 2016

http://www.aimspress.com/journal/Molecular

Communication

Assessing the utility and limitations of high throughput virtual screening

Paul Daniel Phillips[1], Timothy Andersen[2], and Owen M. McDougal[1,*]

[1] Boise State University, Department of Chemistry and Biochemistry, 1910 University Drive, Boise, Idaho 83725, USA
[2] Boise State University, Department of Computer Science and Engineering, 1910 University Drive, Boise, Idaho 83725, USA

* **Correspondence:** Email: owenmcdougal@boisestate.edu; Tel: 208-426-3964.

Abstract: Due to low cost, speed, and unmatched ability to explore large numbers of compounds, high throughput virtual screening and molecular docking engines have become widely utilized by computational scientists. It is generally accepted that docking engines, such as AutoDock, produce reliable qualitative results for ligand-macromolecular receptor binding, and molecular docking results are commonly reported in literature in the absence of complementary wet lab experimental data. In this investigation, three variants of the sixteen amino acid peptide, α-conotoxin MII, were docked to a homology model of the $\alpha_3\beta_2$-nicotinic acetylcholine receptor. DockoMatic version 2.0 was used to perform a virtual screen of each peptide ligand to the receptor for ten docking trials consisting of 100 AutoDock cycles per trial. The results were analyzed for both variation in the calculated binding energy obtained from AutoDock, and the orientation of bound peptide within the receptor. The results show that, while no clear correlation exists between consistent ligand binding pose and the calculated binding energy, AutoDock is able to determine a consistent positioning of bound peptide in the majority of trials when at least ten trials were evaluated.

Keywords: DockoMatic; AutoDock; high throughput virtual screening; conotoxin

1. Introduction

High throughput virtual screening (HTVS) is a widely used technique for determining the binding affinity of a large number of ligands to a target receptor [1-3]. The output of HTVS provides a

qualitative comparison of a ligand library, which is to say the ligands can be ranked from best to worst based on receptor binding affinity. The speed and accuracy of HTVS is dependent on how well the binding pocket of the receptor has been characterized. Commonly, the binding pocket for HTVS docking experiments is determined from the crystal structure of a biomacromolecular receptor with bound ligand.

Jacob et al. (2011) provided a step-by-step description of the process of initiating a docking experiment beginning with a pdb file for a ligand-receptor complex [4]. Briefly, the authors obtained the protein database pdb file for *Aplysia californica*-acetylcholine binding protein (*Ac*-AChBP) (accession number 2BR8) containing α-conotoxin (α-CTx) PnIA[A10L:D14K] bound to the ligand binding domain. The Cartesian coordinate information specifying each atom of the ligand was removed from the pdb file of the ligand-receptor complex leaving the remaining pdb file consisting of only apoenzyme information. AutoDock Tools was then used to generate a grid parameter file encompassing this binding pocket, allowing for HTVS docking experiments to determine how new α-CTx ligands bind to the *Ac*-AChBP. HTVS experiments have been used to discover DNA gyrase inhibitors, improve knowledge of dihydrodipicolinate reductase binding paradigms, discover STAT3 inhibitors, optimizing JAK2 inhibitors, kinase inhibitors, G-quadruplex ligands to demonstrate inert group 9 metals inhibiting protein-protein interfaces, and others [5-12].

In this study, α-CTx peptide ligands were docked to a homology model of the $\alpha_3\beta_2$-nicotinic acetylcholine receptor (nAChR) using DockoMatic version 2.0 software suite [13]. The HTVS was conducted using α-CTxMII as a ligand because experimental data supports a very high affinity of the peptide for the $\alpha_3\beta_2$-nAChR isoform (IC_{50} = 0.5 nM) [14]. The sequence for α-CTxMII is GCCSNPVCHLEHSNLC with disulfide bonds connecting Cys2-Cys8 and Cys3-Cys16. The peptide has a well-established α-helix originated at Pro6 and ending at His12. Coordinate files have been deposited in the Research Collaboratory for Structural Bioinformatics (RCSB) [15] representing α-CTxMII from two separate groups; in this investigation we used the original work submitted by Shon et al. [16] rather than Hill et al. [17]. The structure generated by Shon et al. was based on traditional aqueous solution conditions for NMR experiments in contrast to the further refinement in structure gained by Hill et al. as a result of solvent condition manipulation to reduce polarity by addition of either acetonitrile or trifluoroethanol. In their work, Hill et al. acknowledge the global fold of the peptide remains the same as that reported by Shon et al. For our purposes, either system would have sufficed as a starting point for computational studies. The structure of α-CTxMII contains electropositive amino acids at either end of the helix (Asn5 and His12) that have been proposed to be significant for ligand to receptor interaction [18, 19]. Two additional analogs of α-CTxMII were also evaluated, α-CTxMII[E11A] and α-CTxMII[N5R:E11A:H12K].

Conotoxins are small, cysteine-rich peptides found in the venom of snails belonging to the genus *Conus*. α-CTxs selectively bind and inhibit nAChRs leading to a great deal of research into their use as analgesics and treatments for Parkinson's disease [20-22]. α-CTxMII from the venom of *Conus magus*, is a well-studied molecular probe for $\alpha_3\beta_2$-nAChRs. The single point mutant α-CTxMII[E11A] was determined to shift binding selectivity from $\alpha_3\beta_2$-nAChR isoforms to preferential binding of $\alpha_6\beta_2\beta_3$-containing nAChRs [23]. The triple mutant, α-CTxMII[N5R:E11A:H12K], was designed to have positively charged amino acids at both ends of the α-helix, which was hypothesized to enhance receptor binding affinity while still retaining the E11A selectivity for $\alpha_6\beta_2\beta_3$-containing nAChRs [19]. Figure 1 shows the structures of the three peptides, where the top structures indicate the larger size of the triple mutant (right), which measures 16 Å from end-to-end of the α-helix, in comparison to either

α-CTxMII (left) or α-CTxMII[E11A] (middle), which are both on the order of 11 Å from end-to-end of the α-helix. The bottom set of structures shows the electrostatic topography of each peptide for visualization of the removal of negative charge (red) in position 11 for the α-CTxMII[E11A] mutant, and the enhancement of positive charge (blue) character for the triple mutant.

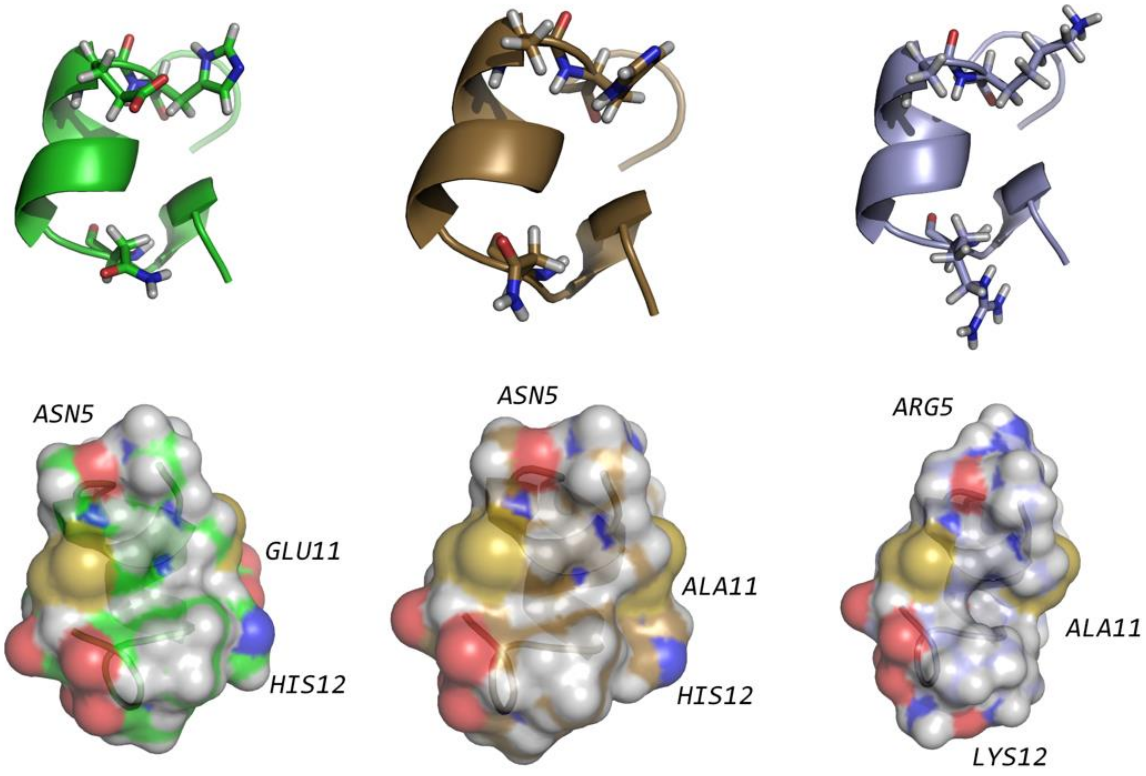

Figure 1. Top, from left to right: distance between amino acid positions 5 and 12 of α-CTxMII (~11 Å), α-CTxMII[E11A] (~11 Å), and α-CTxMII[N5R:E11A:H12K] (~16 Å). Bottom, from left to right: electrostatic topography of α-CTxMII, α-CTxMII[E11A], and α-CTxMII[N5R:E11A:H12K] showing the decrease in negative charge (red) and increase in positive charge (blue) with each successive amino acid mutation.

The aim of this study was to identify the reproducibility and limitations of HTVS for each ligand individually and for the three ligands compared to one another for binding to a homology model of the $\alpha_3\beta_2$-nAChR isoform.

2. Materials and Methods

DockoMatic is a comprehensive open source computational application with a graphical user interface (GUI) designed to facilitate HTVS; the docking engine used by DockoMatic is AutoDock v 4.2 [24]. DockoMatic was installed and made accessible on a high performance computing (HPC) cluster consisting of 32 AMD processors per node with 16 nodes available.

DockoMatic was then used to generate mutant analogs of α-CTxMII beginning with the pdb file (accession number 1M2C) originating from nuclear magnetic resonance spectroscopy structure elucidation [16]. The two mutations to α-CTxMII, α-CTxMII[E11A] and α-CTxMII[N5R:E11A:H12K],

were generated using the Treepack program, subordinate to the DockoMatic GUI [4]. These three peptides were docked to the $\alpha_3\beta_2$-nAChR for a total of ten trials each. Every trial consisted of 100 AutoDock cycles, resulting in 1000 docking trials per peptide [25].

3. Results and Discussion

The results from HTVS provide a qualitative ranking of ligands using a binding energy in kcal/mol as the unit of measurement. Analysis of the lowest free energy for binding (ΔG) resultant from the ten trials for each of the three peptides showed significant variation (Table 1).

Table 1. Lowest free energy of binding value for α-CTxMII, α-CTxMII[E11A], and α-CTxMII[N5R:E11A:H12K] docking to a homology model for the $\alpha_3\beta_2$-nAChR over 10 trials.

Conotoxin	trial1	trial2	trial3	trial4	trial5	trial6	trial7	trial8	trial9	trial10
MII	−14.96	−16.23	−15.32	−13.78	−16.27	−15.39	−14.53	−14.14	−15.32	−14.69
E11A	−15.98	−17.58	−16.22	−16.25	−16.26	−15.92	−17.58	−16.61	−7.14	−16.55
Triple	−17.3	−3.9	−8.21	−9.02	−16.95	−17.4	−17.34	−17.74	−7.43	−17.92

The average free energy of binding for the most favorable bound ligand, resultant from 100 docking cycles over all ten trials, for α-CTxMII was −15.063 ± 0.816 kcal/mol, α-CTxMII[E11A] was −15.069 ± 3.034 kcal/mol, and α-CTxMII[N5R:E11A:H12K] was −13.321 ± 5.483 kcal/mol (Table 2). From these data, one could conclude that more favorable interaction with the receptor is achieved for α-CTxMII and α-CTxMII[E11A] as compared to α-CTxMII[N5R:E11A:H12K], but the large standard deviation for the triple mutant indicated docking reproducibility was a problem and more stringent analysis of the results was required. We speculate that the triple mutant is subject to greater variability in docking result for two reasons: (1) the peptide is 5 Å larger than either α-CTxMII or α-CTxMII[E11A] resulting in greater difficulty to bind to the same receptor binding site, and (2) the triple mutant has greater charge character at either end of the peptide, lending itself to a larger number of binding opportunities that may constitute local energy wells rather than global energy minimum binding interactions.

Table 2. Correlation between average energy for docked ligand-receptor complex and binding pose similarity across ten docking trials consisting of 100 cycles per trial.

Peptide	Superimposed Docked Peptide	Average ΔG (kcal/mol)	Standard Deviation	Root Mean Square Deviation (Å)
MII	1,6,8,9,10	−15.063	±0.816	0.510
E11A	1,3–8,10	−15.609	±3.034	0.526
Triple	1,4,5,7,8,10	−13.321	±5.483	3.459

From the data provided in Table 2, it can be seen that the root mean square deviation (RMSD) between similar binding pose peptides for α-CTxMII and α-CTxMII[E11A] were roughly comparable and within specifications that are considered generally acceptable (e.g. ≤ 0.5 Å). These data are in marked contrast to the RMSD for the triple mutant, which is much higher (3.459 Å), indicating greater variability in ligand binding pose between docking trials. It should be noted that trial 2 for the triple

mutant was omitted from calculation of average free energy of binding because it was well outside of the standard deviation for calculated free energy, and the bound structure did not converge on a common binding site as was observed for the other ligand structures. As stated previously, the larger size and charge character of the triple mutant is suspected to affect convergence across AutoDock trials. Although AutoDock was the docking engine used in this investigation, we suspect other docking engines will be subject to the same limitations. The reason for this is because HTVS protocols limit the number of rotatable bonds (e.g. AutoDock default value = 32) allowed per docking trial to achieve efficient computational processing. To remove the rotatable bond limitation, requires more computationally intensive investigation which then restricts the number of screening experiments that can be run.

Protein data base files for the thirty most energetically favorable ligand binding poses (top results over ten trials for the three peptides) were visualized to identify whether inconsistency in binding energy correlated to significant differences in bound ligand positioning. Figure 2 shows the ligand-receptor complexes where the image to the left represents all thirty peptide binding poses for the three ligands and the image to the right provides a view of reproducible docking pose that is conserved for nineteen of the docked structures derived from HTVS.

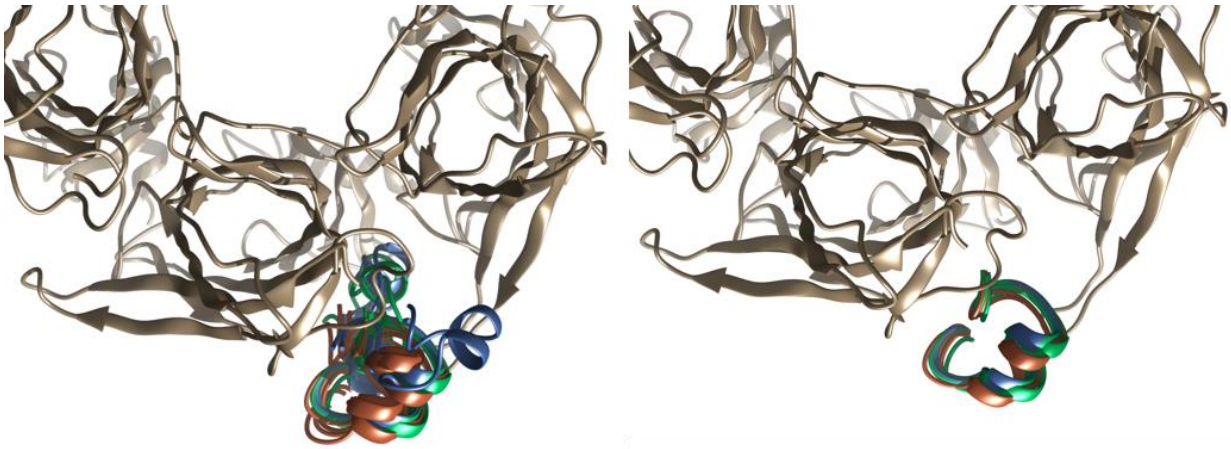

Figure 2. HTVS results for all thirty docking trials (left); nineteen of the thirty trials identified a single conserved binding pose (right).

Close visual inspection of the nineteen ligands with conserved binding pose to the receptor was performed to identify that five of the peptides are α-CTxMII, eight are α-CTxMII[E11A], and the remaining six ligands are α-CTxMII[N5R:E11A:H12K] (Figure 3). In order to limit computational time, AutoDock generates these binding poses in an efficient process by limiting the number of rotatable bonds to 32 for the ligand, and not allowing dynamic motion for the receptor [26]. This computational docking method is limited in its ability to optimize side chain interactions between the ligand and the receptor for large systems like those explored in this exercise. As a result, the calculated binding energy for peptide ligands is largely variable because of side chain immobility inherent to docking engines, such as AutoDock.

The conclusion from this investigation is that HTVS should be performed over multiple trials, followed by visualization of the docking results to identify commonalities in binding pose. The qualitative ranking of ligands in this exercise did not extend between ligands. Thus, it was only through

visual inspection of ligand binding pose and validation of binding pose across a minimum of ten trials, that the reproducibility of results could be assessed. To achieve ten trials for each ligand-receptor docking, the authors used DockoMatic, drawing on the docking engine AutoDock. Alternatively, AutoDock Vina, which also takes advantage of parallel processing across computing clusters, could have been used for efficient docking [27]. Regardless of the docking engine, it is our finding that manual inspection of results and visual comparison of binding pose across trials is required for confident binding pose identification. When data are available in the RCSB, visual inspection of docking results should be validated by comparison to ligand-receptor structure obtained experimentally by x-ray crystallography. This study suggests that results obtained through HTVS should segue into energy calculation using molecular dynamics simulations where ligand and receptor dynamics are permitted. HTVS can be a valuable tool to achieve a starting point for rigorous and time-intensive molecular dynamics simulation experiments or other more thorough methodologies [28].

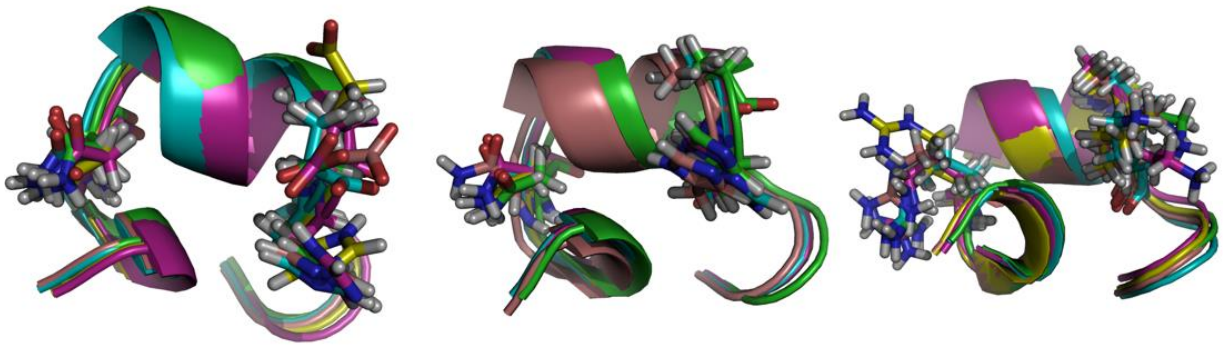

Figure 3. Five of the nineteen peptides with consistent binding pose represent α-CTxMII (left), eight of them are α-CTxMII[E11A] (middle), and the remaining six ligands are α-CTxMII[N5R:E11A:H12K] (right).

Acknowledgements

Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera was developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). Visualization and structure analysis was also performed with the PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC. This material is based in part upon work supported by the National Science Foundation under Grant No. 1229709

Conflict of interest

The authors declare no conflicts of interest in the reporting of this study.

References

1. Deng W, Verlinde C (2008) Evaluation of Different Virtual Screening Programs for Docking in a Charged Binding Pocket. *J Chem Inf Model* 48: 2010-2020.

2. Klebe G (2008) Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today* 11: 580-594.
3. Lavecchia A, Cerchia C (2016) In silico methods to address polypharmacology: current status, applications and future perspectives. *Drug Discov Today* 21: 288-298.
4. Jacob RB, Bullock CW, Andersen T, et al. (2011) DockoMatic: Automated Peptide Analog Creation for High Throughput Virtual Screening. *J Comput Chem* 32: 2936-2941.
5. Jorgensen WL (2004) The many roles of computation in drug discovery. *Science* 303: 1813-1818.
6. Clark DE (2008) What has virtual screening ever done for drug discovery? *Expert Opin Drug Discov* 3: 841-851.
7. Liu LJ, Leung KH, Chan DSH, et al. (2014) Identification of a natural product-like STAT3 dimerization inhibitor by structure-based virtual screening. *Cell Death Dis* 5: 96-104.
8. Ma DL, Chan DSH, Lee P, et al. (2011) Molecular modeling of drug-DNA interactions: Virtual screening to structure-based design. *Biochimie* 93: 1252-1266.
9. Brenk R, Schipani A, James D, et al. (2008) Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *Chemmedchem* 3: 435-444.
10. Ma DL, Chan DSH, Wei G, et al. (2014) Virtual screening and optimization of Type II inhibitors of JAK2 from a natural product library. *Chem Commun* 50: 13885-13888.
11. Ma DL, Lai TS, Chan FY, et al. (2008) Discovery of a drug-like G-quadruplex binding ligand by high-throughput docking. *Chemmedchem* 3: 881-884.
12. Ma DL, Chan DSH, Leung CH (2014) Group 9 Organometallic Compounds for Therapeutic and Bioanalytical Applications. *Acc Chem Res* 47: 3614-3631.
13. Sambasivarao SV, Roberts J, Bharadwaj VS, et al. (2014) Acetylcholine Promotes Binding of alpha-Conotoxin MII at alpha(3)beta(2) Nicotinic Acetylcholine Receptors. *Chembiochem* 15: 413-424.
14. Cartier GE, Yoshikami DJ, Gray WR, et al. (1996) A new alpha-conotoxin which targets alpha 3 beta 2 nicotinic acetylcholine receptors. *J Biol Chem* 271: 7522-7528.
15. Berman HM, Westbrook J, Feng Z, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235-242.
16. Shon KJ, Koerber SC, Rivier JE, et al. (1997) Three-dimensional solution structure of alpha-conotoxin MII, an alpha(3)beta(2) neuronal nicotinic acetylcholine receptor-targeted ligand. *Biochemistry* 36: 15693-15700.
17. Hill JM, Oomen CJ, Miranda LP, et al. (1998) Three-dimensional solution structure of alpha-conotoxin MII by NMR spectroscopy: Effects of solution environment on helicity. *Biochemistry* 37: 15621-15630.
18. Turner M, Eidemiller S, Martin B, et al. (2009) Structural basis for α-conotoxin potency and selectivity. *Bioorg Med Chem* 17: 5894-5899.
19. McDougal OM, Granum DM, Swartz M, et al. (2013) pKa Determination of Histidine Residues in α-Conotoxin MII Peptides by 1H NMR and Constant pH Molecular Dynamics Simulation. *J Physic Chem B* 117: 2653-2661.
20. Armishaw CJ (2010) Synthetic alpha-Conotoxin Mutants as Probes for Studying Nicotinic Acetylcholine Receptors and in the Development of Novel Drug Leads. *Toxins* 2: 1471-1499.
21. Azam L, McIntosh JM (2009) Alpha-conotoxins as pharmacological probes of nicotinic acetylcholine receptors. *Acta Pharmacol Sin* 30: 771-783.

22. Bordia T, Grady SR, McIntosh JM, et al. (2007) Nigrostriatal damage preferentially decreases a subpopulation of α6β2 nAChRs in mouse, monkey, and Parkinson's disease striatum. *Mol Pharmacol* 72: 52-61.
23. McIntosh JM, Azam L, Staheli S, et al. (2004) Analogs of α-Conotoxin MII Are Selective for α6-Containing Nicotinic Acetylcholine Receptors. *Mol Pharmacol* 65: 944-952.
24. Bullock C, Cornia N, Jacob R, et al. (2013) DockoMatic 2.0: High Throughput Inverse Virtual Screening and Homology Modeling. *J Chem Inf Model* 53: 2161-2170.
25. McDougal OM, Cornia N, Sambasivarao SV, et al. (2014) Homology Modeling and Molecular Docking for the Science Curriculum. *Biochem Mol Biol Educ* 42: 179-182.
26. Cavasotto CN, Singh N (2008) Docking and high throughput docking: Successes and the challenge of protein flexibility. *Curr Comput-Aided Drug Des* 4: 221-234.
27. Trott O, Olson AJ (2010) Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* 31: 455-461.
28. Kufareva I, Handel TM, Abagyan R (2015) Experiment-guided Molecular Modeling of Protein-Protein Complexes Involving GPCRs. *Methods Mol Biol* 1335: 295-311.

AIMS Press