Bioorganic & Medicinal Chemistry Letters 19 (2009) 5582–5585

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

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

Structure based virtual screening of GSK-3β: Importance of protein flexibility and induced fit

Nibha Mishra [a], Arijit Basu [a,*], Venkatesan Jayaprakash [a], Ashoke Sharon [b], Mahua Basu [c], Kiran K. Patnaik [d]

[a] Department of Pharmaceutical Sciences, Birla Institute of Technology, Mesra, Ranchi 835 215, India
[b] Department of Applied Chemistry, Birla Institute of Technology, Mesra, Ranchi 835 215, India
[c] Shri Shikshayatan College, 11, Lord Sinha Road, Kolkata 700 071, India
[d] Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi 835215, India

ARTICLE INFO

Article history:
Received 20 June 2009
Revised 8 August 2009
Accepted 11 August 2009
Available online 14 August 2009

Keywords:
GSK-3β
Virtual screening
Docking
Induced fit

ABSTRACT

GSK-3β, one of the vital enzymes responsible for various phosphorylation catalysis. Induced fit mechanism and the presence of conserved water molecule(s) in the active site poses complexity during the process of virtual screening. The present investigation reveals the practical strategy to handle the induced fit mechanism of GSK-3β though flexible docking protocol. This protocol provides an enrichment of 70% in top 1% of the dataset with a rank correlation of >0.9 and found better in comparison to earlier reported protocols.

Glycogen Synthase Kinase-3 (GSK-3) is a serine/threonine protein kinase that modulates the activity of various proteins by phosphorylation.[1,2] It has two isoforms α and β that are 97% homologous in their catalytic domain but diverse at the N and C terminus.[3,4] Both isoforms have a conserved N-terminal serine residue (S21 for GSK-3α and S9 for GSK-3β) and its phosphorylation plays an important role for further activity.[5] GSK-3β phosphorylates proteins such as glycogen synthase, acetyl CoA carboxylase, β-catenin, cyclin D1, Tau, etc. which plays a significant role in the signaling pathways. It controls several cellular process such as differentiation, proliferation, metabolic processes, apoptosis, inflammation, neuronal function, etc.[6,7] The small molecule inhibitors of GSK-3β have a therapeutic potential for the treatment of human diseases like neurodegenerative diseases, type II diabetes and cancer.[8,9] Therefore, the therapeutic significance of GSK-3β attracts researchers to understand its molecular basis of ligand binding properties through molecular docking studies.[10–12] The significant success in docking protocols provides a strategy to predict binding mode under rigid docking, however the flexible docking and consideration of conserved water is still a practical challenge. There are multiple co-crystal structures of GSK-3β (PDB codes 1Q4L, 1Q3W, 1Q41, 1UV5, 1R0E, 1Q5K, 1Q3D, 1PYX, 2OW3) with different ligands are available in the protein data bank. However, the binding mode investigation clearly reveals the significant change in the active site in terms of position of conserved water(s) and the orientation of the active-site residues. Therefore, the single co-crystal structure may not able to provide a universal solution for its structure based virtual screening, to develop new potential molecule(s). In lieu of this, we have initiated structural investigation to understand the molecular basis of induced fit mechanism associated with GSK-3β and discussed a plausible practical strategy for virtual screening within a reasonable computational time frame.

Significant induced fit mechanism[10,11] and presence of conserved water molecules in the active site of GSK-3β makes structure based virtual screening on this target very challenging. Due to induced fit mechanism the active site of the receptor could easily accommodate many structurally diverse ligands. Previously Polgr et al.[11] have done an elaborate study for choosing the correct protein structure. They classified the PDB structures into three groups as per the conformation of Gln185. It was also suggested to use 1UV5, 1Q5K and 1Q4L (one representative from each group) for the structure based virtual screening experiment. Amongst all the PDB structures 1UV5 provided the best enrichment in that particular experiment. A recent report highlights the crystal structural analysis of a ligand 'Bis-(indole)maleimide pyridinophane (BIM)' into the active site of GSK-3β (pdb: 2OW3).[13] The detail structural analysis (Fig. 1) provides a significant drift between the active site of 2OW3 in comparison to 1UV5, 1Q5K and 1Q4L. When BIM was cross-docked to 1UV5, 1Q5K and 1Q4L yielded very low score, which is not at par with its superior biological activity

* Corresponding author. Tel.: +91 9430353750.
E-mail address: arijit4uin@gmail.com (A. Basu).

0960-894X/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2009.08.042

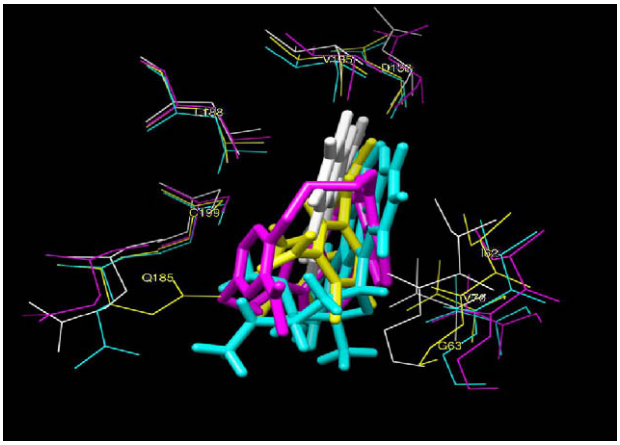

Figure 1. Active sites of the PDB used for the current study. The native position of the ligand with the receptors is shown. 1UV5-white, 1Q4L-yellow, 2OW3-magenta, and 1Q3D-cyan. Significant flexibility is observed in the active site. Q185 shows a marked flip. G63, V70 and I62 along with other residues shift its position in order to accommodate structurally diverse ligands.

(IC_{50} = 3 nM). The binding mode studies reveal a significant steric clash into the active site of 1UV5, 1Q5K and 1Q4L and BIM this explains for its low docking score. It can be seen from Figure 1. 2OW3 has a wider and more accommodative active site than any of the previous PDB structures.

The systematic structure based virtual screening was pursued on GSK-3β and protocols were summarized in Table 1**.** Thousand drug like decoys from Schrödinger Inc. were considered. Ten ligands were randomly eliminated and 10 known actives were seeded to have a random hit rate of 1% (10 * 100/1000). The known ligands (Supplementary data) were chosen on the basis of structural diversity. The four distinguished conformational structure (1UV5, 1Q5K, 1Q4L and 2OW3) of GSK-3β were selected for the virtual screening experiment. The freely available program AutoDock 4.0 has been selected to perform docking calculation. It can be used to explore the induced fit mechanism by providing the flexibility of the selected active site residues. To expedite the screening, a virtual screening script (ViSTA: Virtual Screening Tool for AutoDock, a unix shell script) were written and validated to perform automation during screening calculations. The initial docking was performed by rigid-receptor methodology, with or without the conserved water molecules in the active site. In the second phase, different combinations of active site flexible residues were considered to simulate induce fit mechanism, with or without conserved water molecule in the active site. The docking protocols[14] were validated by their enrichment factors[15–17] which describe the number of active compounds found by employing a certain virtual screening strategy. It is widely used a validation tool for assessing the quality of virtual screening protocol. Conceptually the enrichment factor metric is simply the measure of how many more actives we find within a defined 'early recognition' fraction of the ordered list relative to a random distribution. The enrichment factors **(1)** were calculated as follows:

$$E_f = \frac{N^{x\%}_{\text{experimental}}}{N^{x\%}_{\text{expected}}} = \frac{N^{x\%}_{\text{experimental}}}{N_{\text{active}} \cdot x\%} \quad (1)$$

($N_{\text{experimental}}$) number of experimentally found active structures in the top x% of the sorted database, (N_{expected}) number of expected active structures, (N_{active}) total number of active structures in database).

The dataset in the current study is enriched by 1% (10 known active and 990 inactive), the best probable outcome for any virtual screening protocol for this study will be 100% (10 out of 10) at the top 1%. Therefore, enrichment factor for this dataset has to be

Table 1
Different docking protocols adapted for virtual screening of GSK-3β

Docking code	PDB ID	Flexible protein residue	Water molecules
D1	2OW3	NONE	NONE
D2	1UV5	NONE	NONE
D3	1Q3D	NONE	NONE
D4	1Q4L	NONE	NONE
DW1	2OW3	NONE	HOH389
DW2	1UV5	NONE	HOH34, HOH104
DW3	1Q3D	NONE	HOH504, HOH554
DW4	1Q4L	NONE	HOH508
DFG1	2OW3	GLN185	NONE
DFG2	1UV5	GLN185	NONE
DFG3	1Q5K	GLN185	NONE
DFG4	1Q4L	GLN185	NONE
DFGW1	2OW3	GLN185	HOH389
DFGW2	1UV5	GLN185	HOH34, HOH104
DFGW3	1Q5K	GLN185	HOH504, HOH554
DFGW4	1Q4L	GLN185	HOH508
DFT1	2OW3	V70, I62, G63, Q185	NONE
DFT2	1UV5	V70, I62, G63, Q185	NONE
DFT3	1Q5K	V70, I62, G63, Q185	NONE
DFT4	1Q4L	V70, I62, G63, Q185	NONE
DFTW1	2OW3	V70, I62, G63, Q185	HOH389
DFTW2	1UV5	V70, I62, G63, Q185	HOH34, HOH104
DFTW3	1Q5K	V70, I62, G63, Q185	HOH504, HOH554
DFTW4	1Q4L	V70, I62, G63, Q185	HOH508
DFA1	2OW3	L188, V135, D133, C199, Q185, I62, G63, V70	NONE
DFA2	1UV5	L188, V135, D133, C199, Q185, I62, G63, V70	NONE
DFA3	1Q5K	L188, V135, D133, C199, Q185, I62, G63, V70	NONE
DFA4	1Q4L	L188, V135, D133, C199, Q185, I62, G63, V70	NONE
DFAW1	2OW3	L188, V135, D133, C199, Q185, I62, G63, V70	HOH389
DFAW2	1UV5	L188, V135, D133, C199, Q185, I62, G63, V70	HOH34, HOH104
DFAW3	1Q5K	L188, V135, D133, C199, Q185, I62, G63, V70	HOH504, HOH554
DFAW4	1Q4L	L188, V135, D133, C199, Q185, I62, G63, V70	HOH508

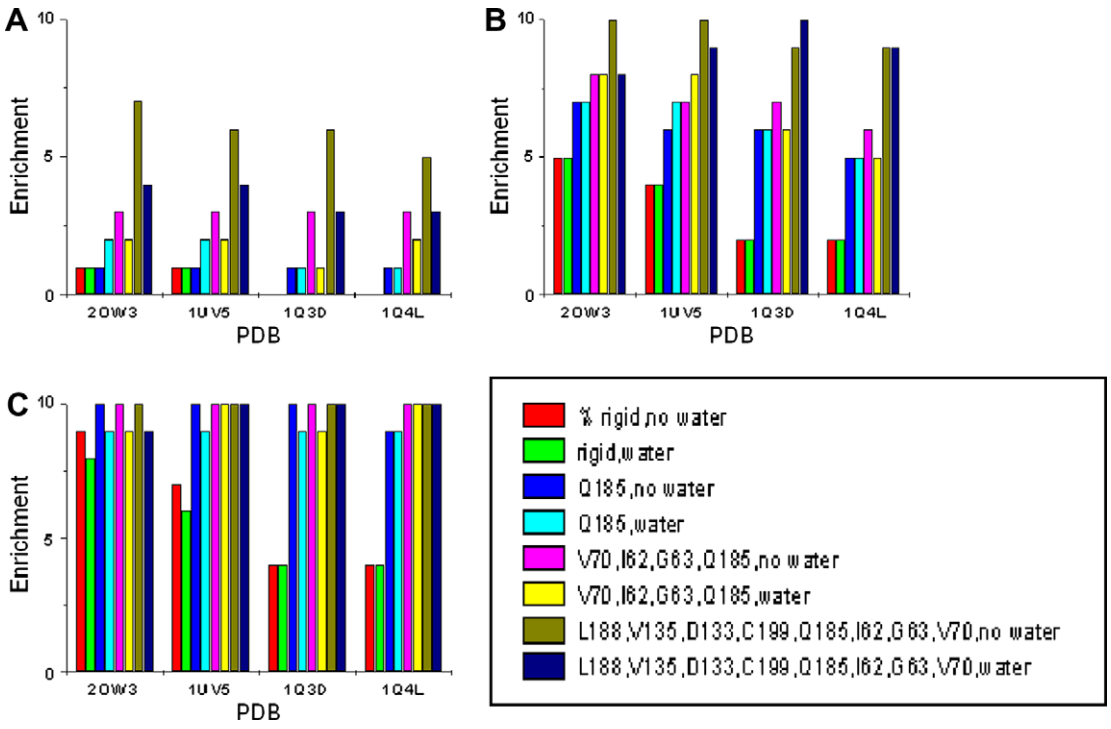

Figure 2. Enrichment for (A) top 1% dataset (B) top 5% dataset (C) top 10% dataset for 2OW3, 1UV5, 1Q3D and 1Q4L.

tested for at least top 1% of the dataset. The enrichment factors for the current study (Fig. 2A–C) were studied at top 1%, 5% and 10% of the dataset. Where, all the residues of the receptor were kept rigid, no significant difference in enrichment for any of the PDB structures were observed for top 1% of the dataset. However, 2OW3 and 1UV5 gave a better result as compared to other PDB structures when top 5% and 10% of the dataset were considered.

Inclusion of conserved water molecules did not improve the performance of the virtual screening experiment. It was presumed that docking to a wider and more accommodative active site of 2OW3 is likely to provide a better enrichment. But the obtained results were not at par to our expectation. It was presumed that docking to a wider and more accommodative active site of 2OW3 is likely to provide a better enrichment. The binding mode investigation after the rigid docking experiment clearly shows poor interactions between the structurally dissimilar ligands with active-site residues of 2OW3 in comparison to native ligand (BIM). Therefore, the induced fit methodology implemented to investigate in detail and to find out a general solution to overcome the enrichment problem. In addition, the role of conserved water was also studied.

Initially only Q185 was kept flexible, gradually V70, I62, G63, Q185 and finally L188, V135, D133, C199, Q185, I62, G63, V70 residues were made flexible. Enrichment factors gradually increased and an unprecedented enrichment[10,11] of 70% in top 1% of the dataset was observed when L188, V135, D133, C199, Q185, I62, G63, V70 residues of 2OW3 were kept flexible. Retention of conserved water molecules in the active site deteriorated the enrichment. Interestingly, when L188, V135, D133, C199, Q185, I62, G63, V70 residues were kept flexible other PDB structures like 1UV5, 1Q5K and 1Q4l also produced appreciable enrichment of 60%, 60% and 50% when top 1% of the dataset was analyzed. The largest ligand (BIM) was found in the top 1% of the dataset for all the PDB structures. When top 5% and 10% of the dataset was considered PDB structures 2OW3 and 1UV5 gave an acceptable enrichment ranging from 70% to 100% for almost any docking protocol that was adopted. But the enrichment for top 1% of the dataset was significantly dependent on the choice of docking protocol. Consideration of the flexible residues in the active site, successfully identified 50–70% of the known actives in top 1% and 100% of the known actives in top 5% of the dataset.

Ideally any docking program should have the power to sort the compounds according to their activity. Prediction of real binding affinities is an important aspect of any scoring function and their score (binding energy) should reflect their biological activity. As long as a scoring function can provide the correct rankings of candidate molecules, it will work perfectly. Therefore, to study the ability of the adopted protocols in this study to sort the known actives, 20 known inhibitors were selected from various literatures (Supplementary data). The compounds were selected so as to cover a wide range (1.1 nM–10 μM) of biological activities and structural diversities. Spearman's rank order correlation[18–21] coefficient ρ (**2**) was employed to evaluate the ability of the employed docking protocols to sort the compounds according to their activity. All the 20 compounds were docked on GSK-3β employing the same docking protocols used earlier; compounds were then ranked according to their docking scores (binding energy):

$$\rho = 1 - \frac{6\sum_i^n [r(x_i) - r(y_i)]^2}{n^3 - n} \qquad (2)$$

where n is the number of pairs $n = 20$ in our case, $r(x_i)$ and $r(y_i)$ are the rank of the activity and the interaction energy of the ith sample in the testing set. In theory, the Spearman correlation coefficient falls between −1 and +1.

Where +1 corresponds to a perfect correlation, −1 corresponds to a perfect inverse correlation, and zero corresponds to total disorder. The rank correlation was increased gradually (Fig. 3) as the residues were kept flexible. The best rank correlations >0.9 were obtained when L188, V135, D133, C199, Q185, I62, G63, V70 residues were kept flexible with PDB structures 2OW3 and not considering any of the conserved water molecule(s). Again, the consideration of water molecule did not provide any enhancement in the rank correlation. Therefore, it clearly reveals the potential of induced fit methodologies to overcome the screening problem. The induced fit simulation may be carried out through

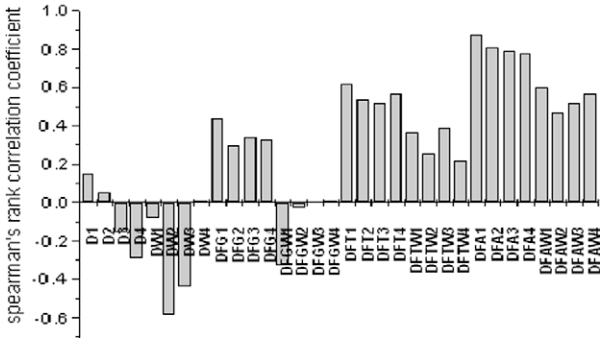

Figure 3. Spearman's rank order correlation coefficient for different docking protocols employed in the study. Refer to Table 1 for docking codes.

molecular dynamics but, it could be computationally expensive. Another significant observation, was that a satisfactory enrichment and rank correlation was obtained when L188, V135, D133, C199, Q185, I62, G63, V70 residues were kept flexible with PDB structures 1UV5, 1Q5K and 1Q4l. The flexible residues in the active site adopts a suitable conformations so as to make it wide enough for larger ligand e.g. BIM. Therefore it can be inferred that any of the PDB structures 2OW3, 1UV5, 1Q5K and 1Q4l can be considered for the structure based virtual screening protocol but, the residues suggested above has to be kept flexible.

Every ligand engages different water molecule(s) for binding with the receptor therefore it was not possible to identify the water molecule(s) which may provide a universal solution. Docking without considering the conserved water molecule(s) provided satisfactory enrichment as well as rank correlation hence; virtual screening experiment on this target can be carried out without retaining any of the active site water molecules.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2009.08.042.

References and notes

1. Li, G.; Iyengar, R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13254.
2. Ali, A.; Hoeflich, K. P.; Woodgett, J. R. *Chem. Rev.* **2001**, *101*, 2527.
3. Woodgett, J. R. *EMBO J.* **1990**, *9*, 2431.
4. Woodgett, J. R. *Methods Enzymol.* **1991**, *200*, 564.
5. Cross, D. A.; Alessi, D. R.; Cohen, P. *Nature* **1995**, *378*, 785.
6. Doble, B. W.; Woodgett, J. R. *J. Cell Sci.* **2003**, *116*, 1175.
7. Martinez, A.; Castro, A.; Dorronsoro, I.; Alonso, M. *Med. Res. Rev.* **2002**, *22*, 373.
8. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269.
9. Martinez, A. *Med. Res. Rev.* **2008**, *28*, 773.
10. Gadakar, P. K.; Phukan, S.; Dattatreya, P. M.; Balaji, V. N. *Curr. Sci.* **2007**, *93*, 1100.
11. Polgr, T.; Baki, A.; Szendrei, G. I.; Keseru, G. M. *J. Med. Chem.* **2005**, *48*, 7946.
12. Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. *Br. J. Pharmacol.* **2008**, *153*, S7.
13. Zhang, H. C.; Boñaga, L. V. R.; Ye, H.; Derian, C. K.; Damiano, B. P.; Maryanoff, B. E. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2863.
14. (a) *Protein preparation*: PDB codes 1Q4L, 1Q3W, 1Q41, 1UV5, 1R0E, 1Q5K, 1Q3D, 1PYX and 2OW3 (www.rcsb.org) were downloaded and analyzed for their active site residues and then only four 2OW3, 1UV5, 1Q3D, 1Q4L were considered. All the protein structures were initially corrected using MolProbity interactive server. The resulting structure is then further refined using Schrödinger protein preparation wizard. Protonation states were adjusted and finally a restrained energy minimization using OPLS2005 force field was carried out. (b) *Ligand preparation*: All ligands used were minimized using Macro Model (*MacroModel, version 9.6, Schrödinger, LLC, New York, NY, 2005*); OPLS2005 force field was used to minimize their structure. (c) *Docking simulations*: All the docking experiments were performed with AutoDock 4.0. Lamarckian Genetic Algorithm was employed as the docking algorithm. Virtual screening protocol was automated by a separate script was written and validated. (d) *Docking parameters*: Number of Genetic Algorithm (GA) runs: 10, Population size: 150, Maximum number of evaluation: 2,500,000, Maximum number of generation: 27,000.
15. Bender, A.; Glen, R. C. *J. Chem. Inf. Model.* **2005**, *45*, 1369.
16. Truchon, J.; Bayly, C. I. *J. Chem. Inf. Model.* **2007**, *47*, 488.
17. Chen, H.; Lyne, P. D.; Giordanetto, F.; Lovell, T.; Li, J. *J. Chem. Inf. Model.* **2006**, *46*, 401.
18. Wang, R.; Lu, Y.; Wang, S. *J. Med. Chem.* **2003**, *46*, 2287.
19. Moitessier, N.; Therrien, E.; Hanessian, S. *J. Med. Chem.* **2006**, *49*, 5885.
20. Seifert, M. H. J.; Kraus, J.; Kramer, B. *Curr. Opin. Drug Disc. Dev.* **2007**, *10*, 298.
21. Englebienne, P.; Moitessier, N. *J. Chem. Inf. Model.* **2009**, *49*, 1568.