Bioorganic & Medicinal Chemistry Letters 19 (2009) 5769–5772

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

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

Variation of protein binding cavity volume and ligand volume in protein–ligand complexes

N. Saranya, S. Selvaraj *

Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli 620 024, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 16 April 2009
Revised 17 July 2009
Accepted 30 July 2009
Available online 3 August 2009

Keywords:
Protein binding cavity volume
Protein flexibility
Ligand volume
Docking
Virtual screening

ABSTRACT

We have systematically analyzed the variation of protein binding cavity volume of 200 protein–ligand complexes belonging to eight protein families. Wide variation in protein binding cavity volume for the same protein is observed on binding different ligands. Analysis of individual protein families shows high correlation between atom–atom interactions in binding site and ligand volume. This study implies the significance of protein flexibility in docking small molecule inhibitors on the basis of protein binding cavity volume with respect to ligand volume.

Structure based drug design (SBDD) aims at the identification of potent lead compounds from the library of compounds and accurate prediction of binding affinity for the particular enzyme/receptor, by docking[1–3] and scoring[4] studies. When the 3D structure of a protein is available, potential binding sites can be identified using appropriate algorithms such as Pocket finder,[5] SURFNET,[6] PocketPicker,[7] CASTp[8] etc. and small molecules compatible with the binding sites are chosen for docking/virtual screening studies. In docking studies, other than ligand flexibility, protein flexibility is also a major issue to identify the correct binding mode (pose) of the ligand binding with receptor. In molecular recognition, binding partner influences the proper conformation from the ensemble of rapidly interconverting conformational species of unbound enzyme molecule.[9,10] In 'induced fit', binding of the ligand to a protein may cause conformational changes that align the residues involved in binding in their correct orientation.[11–13] Incorporation of protein flexibility and mobility in SBDD using methods such as molecular dynamics simulation methods, graph theoretical and geometry based approach and harmonic methods have been discussed elaborately by Teague[14] and Fradera and Mestres.[15]

Ligand binding causes variation in protein binding cavity volume (PCV) due to protein flexibility. A change in PCV for each conformation in a single binding site has been observed upon complex formation. In the case of docking studies, mostly flexibility of the ligand alone is considered and protein/receptor conformation remains rigid. A number of studies have highlighted the importance of protein flexibility in docking.[16–22] Prior to docking, knowledge of the extent of flexibility of receptor obtained from experimental studies will enrich the procedure of finding the correct pose of ligand and receptor.

In the present work, we have analyzed PCV in 200 protein ligand complexes from eight different protein families viz. Thrombin, Trypsin, Cyclin Dependent Kinase-2 (CDK-2), Carbonic anhydrase, HIV-1 Protease, FactorXa, Acetylcholine esterase and HSP 90. We have compared the variation of predicted PCV change on complex formation with a set of diverse ligands in individual protein families. An extensive analysis on the atom–atom interactions in the binding site and its relationship with ligand volume has been carried out where the number of interactions in the binding site at a distance of 2.5–4.5 Å was calculated.

Information regarding the 200 protein–ligand complexes for all the eight protein families, as well as their corresponding PDB IDs were collected from Binding MOAD database.[23] The atomic coordinates of both apo and holo structures were obtained from RCSB Protein Data Bank (PDB).[24] PDB IDs and their corresponding ligand hetero atom (HETATM) IDs (www.rcsb.org) of protein–ligand complexes of eight different protein families are given in Supplementary data in Table S1. PCV of a protein–ligand complex was predicted using Pocket finder.[5] Ligand volume (LV) was calculated using molinspiration server (www.molinspiration.com).[25] For each ligand, SMILES obtained from Protein Data Bank (PDB) was used as input to calculate the LV. Protein–ligand complexes included in the dataset are of resolution less than 2.5 Å.

Atom–atom interactions analysis in the protein–ligand complexes were carried out by calculating the number of interactions

* Corresponding author. Tel.: +91 431 2407072x620; fax: +91 431 2407045.
E-mail address: sel_emi@yahoo.co.uk (S. Selvaraj).

0960-894X/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2009.07.140

involving C (carbon), H (hydrogen), O (oxygen) and N (nitrogen) atoms at a distance of 2.5–4.5 Å from the ligand atoms in the binding site using an in-house perl program. Multiple linear regression analysis of number of interactions involved with C (carbon), H (hydrogen), O (oxygen) and N (nitrogen) at the protein binding site versus calculated LV was evaluated to predict the LV in each individual protein family.

Protein–ligand complexes and their corresponding apo protein structures of the eight protein families were analyzed for PCV. Variations in PCV for various ligands were analyzed systematically in comparison with the respective apo protein structure. Figure 1 shows the variation of the minimum and maximum PCV for the eight families in relationship to apo PCV.

From the graph, wide variation in PCV of the bound state in comparison with the apo protein for each protein family was observed. Further, it showed how PCV changes when different ligands bind to the protein. More or less rigid binding site was observed in Trypsin[26] and Factor Xa. Maximum, minimum PCV for Trypsin and Factor Xa was found to be 100 $Å^3$ (PDB ID: 1GHZ) (Fig. 2a), 181 $Å^3$ (PDB ID: 1AQ7) and 151 $Å^3$ (PDB ID: 1EZQ), 285 $Å^3$ (PDB ID: 1NFW), respectively. High PCV variation was observed in CDK2 with the minimum and maximum PCV of 515 $Å^3$ (PDB ID: 1OIQ) and 1694 $Å^3$ (PDB ID: 1PXN) (Fig. 2b), respectively, which is due to loop flexibility.[27,28] In HIV-1 protease, apo PCV tends to be much smaller than minimum PCV because of the conformational flexibility in the flap domains.[29]

Variation in PCV in increasing order for the eight protein families are given as Trypsin < FactorXa < Carbonic anhydrase < Thrombin < Acetylcholine esterase < HIV-1 Protease < HSP 90 < CDK-2. Erickson et al. (2004) have studied the effects of ligands and protein flexibility on molecular docking accuracy and have reported the effect of protein movement in decreasing the docking accuracy.[30] The decrease in docking accuracy shows the trend Trypsin < Thrombin < HIV-1 Protease, which seems to accord with the results observed in the present study. There is low correlation between PCV and LV (Fig. 3), PCV and ligand molecular weight in all protein families. Earlier studies by Liang et al. (1998)[31] showed a linear correlation between the ligand volume (LV) and binding site volume, when binding site volume is $\leqslant 700$ $Å^3$, inspite of the fact that binding site volume varies largely for different protein families.

Evidence of any relationship between LV and PCV can be obtained from Figures 1 and 4 by comparing Trypsin and CDK-2. Inspite of its rigid binding site Trypsin accommodates ligands with high LV, as ligands are only partially buried in the binding cavity. On the other hand, inhibitors are completely buried in the binding cavity of CDK-2; hence, it implicitly binds smaller ligands with higher PCV.

The ratio was evaluated between maximum and minimum LV, and maximum and minimum PCV to observe the proportion of change in PCV and LV in eight protein families (Fig. 5). A ratio of 1:5 was obtained for LV in Thrombin, followed by a uniform ratio of 1:4 (approximately) for Trypsin, Cyclin Dependent Kinase-2 (CDK-2), Carbonic anhydrase and Acetylcholine esterase families and 1:2 for HSP 90, HIV-1 Protease and FactorXa families. A ratio of 1:2 (approximately) was observed for PCV in most of the protein families. Virtual screening of ligands not far from the minimal and maximal LV values may reduce the ligand decoys from the large group of small molecular compounds. Further, an average LV of 200 PDB structures included in this analysis was found to be 426 $Å^3$ approximately.

Flexible nature of protein shows diverse binding modes for even structurally similar ligands.[32] In contrast single binding site has similar binding volume for structurally diverse ligands.[33] In our study, ligands with identical volume which shows different PCV in the same protein (1DX6, 1QTI, 1W6R, 1GPK, 1VOT, 2C4H, 2C58, 1I9L, 1G1D, 1IF8, 1IF7, 1CE5, 1F0T, 1GI4, 1GI1, 2J34, 2J38, 1NFY, 2CJI, 1UY7, 1UY8, 2BT0 and 2CDD) and vice versa (1AID, 1HPV, 1HPO, 1G2K, 1G4J, 1I9L, 1BN1, 2FNN, 1F0T, 1GI4, 1G36, 1QBO, 1O3I, 1TNG, 1C1R, 1TNH, 1C5S, 1F0U, 2FX6, 2BZA, 1C5Q, 1DI8, 1R78, 2CJI, 1F0S, 2CDD, 2BT0, 1UY8 and 2FWY) were observed in 52 protein–ligand complexes.

Ligand has major impact on receptor plasticity. Along these lines, receptor flexibility should be incorporated in view of the ligand to be docked. Range of variability of PCV remains different for different protein families.

Multiple linear regression analysis of atom–atom interactions and LV showed direct correlation between calculated and predicted LV in all protein families which are plotted in a single graph as given below (Fig. 6), which is clearly due to protein plasticity. Numbers of interactions involving C, H, O and N at protein binding site interacting with ligand at a distance of 2.5–4.5 Å in protein–ligand complexes are given in Figure S1 and Table S2 of Supplementary data. In the above mentioned distance criteria almost all the atoms in the ligand are found to interact with the protein/receptor.

In most protein families, H interactions were more numerous compared to C, O and N interactions. Hydrogen atoms involved in non-bonded interactions have a crucial part of energetic contribution in protein–ligand complexes. Variation in H interaction within individual protein family has high impact in recognition of protein and ligand in each protein–ligand complex. Number of interactions (C, H, O and N) do not increase or decrease in accordance with a protein of high or low flexibility.

When a protein binds a ligand, it alters its conformation suitably to accommodate the same. As the extent of conformational changes varies widely depending on the ligands, the interactions also vary accordingly at the binding site for the same protein. Thus, the correlation between the atoms involved in interaction (C, H, O and N) and ligand volume demonstrates the ligand defined protein conformational variability and hence emphasizes the need to account protein plasticity in docking studies. This suggests that in docking/virtual screening studies to explore the binding of differ-

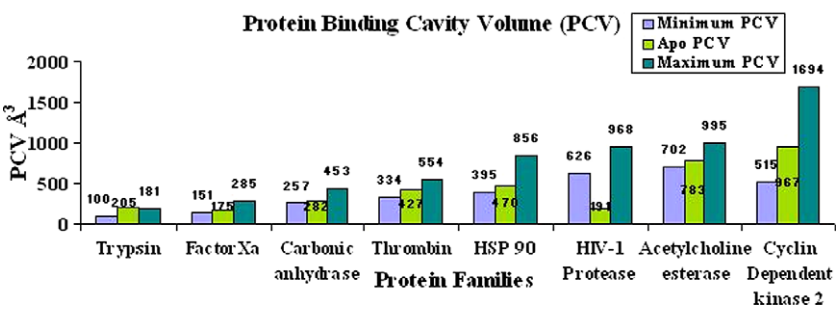

Figure 1. Bar graph showing the minimum PCV, Apo PCV and maximum PCV of eight protein families.

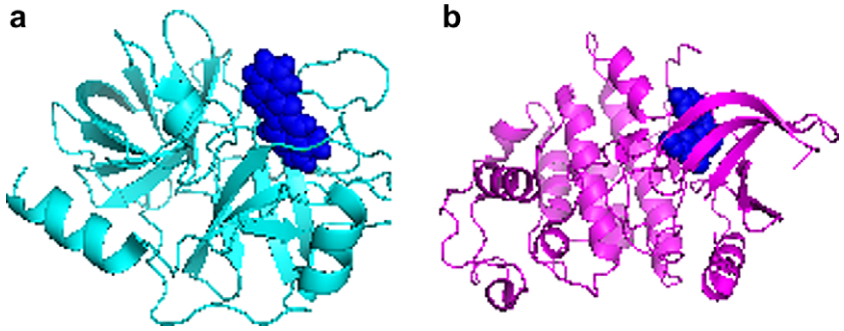

Figure 2. (a) Trypsin (PDB ID: 1GHZ) in complex with inhibitor 120 partially buried, having lower PCV of 100 Å. (b) Inhibitor (CK6) accommodates whole binding pocket of CDK2 (PDB ID: 1PXN) with larger PCV of 1694 Å.

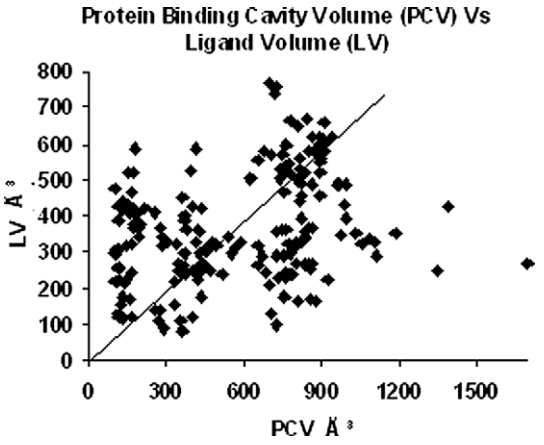

Figure 3. Two hundred complexes have less correlation ($r^2 = 0.10$) between PCV and LV.

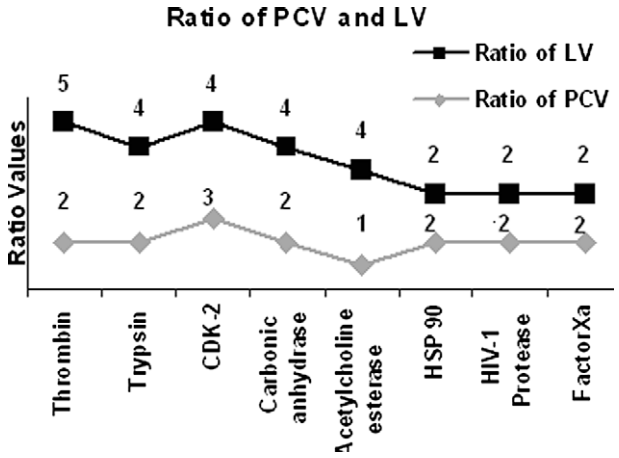

Figure 5. Ratio of PCV and LV for eight protein families.

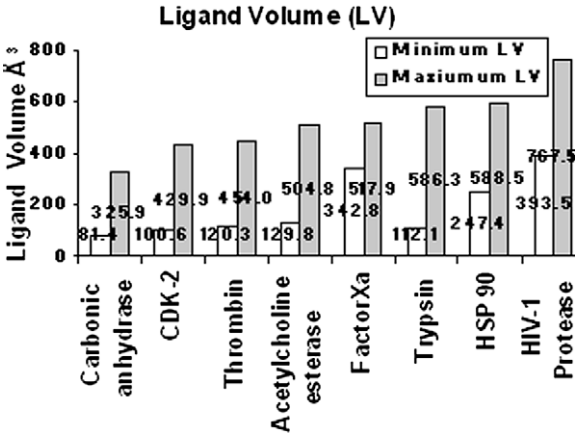

Figure 4. Minimum and maximum ligand volume of eight protein families shown as bar graph.

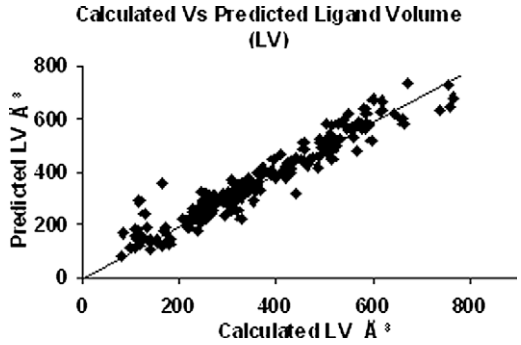

Figure 6. Observed linear correlation ($r^2 = 0.92$) between calculated and predicted ligand volume in all eight protein families.

ent ligands to a specific receptor, one needs to pay greater attention to ligand induced protein conformational variability.[8,14] Also the volume of the ligands to be docked need not be restricted to the PCV observed in apo form of the protein. Chemical information present in the ligand, that is, key functional groups essential for the activity and their possible atom–atom interactions with the protein and knowledge of the native protein binding pocket as to whether it is flexible are very crucial for the success of virtual screening in SBDD.

In summary, an attempt has been made to analyze the variation of protein plasticity for different protein families of mostly studied drug targets consisting of 200 protein–ligand complexes. The present study has provided insight into the extent of variation of PCV and LV in the eight protein families considered. This information could be effectively incorporated in docking/virtual screening studies of these drug target families in the future. Improvement in binding affinity prediction from docking studies results from identifying the appropriate binding mode of protein/receptor, which remains difficult in SBDD. Single binding site accommodates ligands of different volume and binding cavity volume also changes accordingly. In a virtual screening or SBDD procedure, a structure with a different or no inhibitor is generally available. Using these protein structures in apo or ligand bound form often results in drop of docking accuracy.[29] Selecting a protein structure with appropri-

ate conformation and cavity volume based on the LV and chemical properties of ligand molecule to be docked can produce a better binding pose. In addition to selection of protein conformation, incorporation of flexibility of both protein and ligand may yield accurate results. Results obtained on variation in PCV for different ligands in individual protein families can also be used in post processing of docking/virtual screening studies. Molecular interactions exhibit the recognition of a particular ligand to the protein in its specific conformation. Protein flexibility is mainly revealed by the ligand to which it binds.[8] Future studies involving characterization of binding site for conserved interactions, as well as, the knowledge of the extent of protein flexibility available from the crystal structure information will throw light on over all improvement in docking/virtual screening studies.

Acknowledgement

We thank the University Grants Commission, New Delhi for the award of Fellowship to N.S.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2009.07.140.

References and notes

1. Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins* **2002**, *47*, 409.
2. Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J. Comput. Aided Mol. Des.* **2002**, *16*, 151.
3. Glen, R. C.; Allen, S. C. *Curr. Med. Chem.* **2003**, *10*, 763.
4. Gohlke, H.; Klebe, G. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644.
5. Laurie, A. T.; Jackson, R. M. *Bioinformatics* **2005**, *21*, 1908.
6. Laskowski, R. A. *J. Mol. Graph.* **1995**, *13*, 323.
7. Weisel, M.; Proschak, E.; Schneider, G. *Chem. Central J.* **2007**, *1*, 7.
8. Binkowski, T.; Naghibzadeh, S.; Liang, J. *Nucleic Acids Res.* **2003**, *31*, 3352.
9. Tsai, C. J.; Ma, B.; Nussinov, R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9970.
10. Tsai, C.-J.; Kumar, S.; Ma, B.; Nussinov, R. *Protein Sci.* **1999**, *8*, 1181.
11. Koshland, D. E. *Science* **1967**, *156*, 540.
12. Fersht, A. R.; Knill-Jones, J. W.; Bedouelle, H.; Winter, G. *Biochemistry* **1988**, *27*, 1581.
13. Koshland, D. E. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 2375.
14. Teague, S. J. *Nat. Rev. Drug Disc.* **2003**, *2*, 527.
15. Fradera, X.; Mestres, J. *Curr. Top. Med. Chem.* **2004**, *4*, 687.
16. Carlson, H. A.; McCammon, J. A. *Mol. Pharmacol.* **2000**, *57*, 213.
17. Cavasotto, C. N.; Abagyan, R. A. *J. Mol. Biol.* **2004**, *337*, 209.
18. Rauh, D.; Klebe, G.; Stubbs, M. T. *J. Mol. Biol.* **2004**, *335*, 1325.
19. Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Protein Struct. Funct. Bioinform.* **2006**, *65*, 15.
20. Huang, S. Y.; Zou, X. *Protein Struct. Funct. Bioinform.* **2007**, *66*, 399.
21. Krovat, E. M.; Steindl, T.; Langer, T. *Curr. Comput. Aided Mol. Des.* **2005**, *1*, 93.
22. Yang, C. Y.; Wang, R.; Wang, S. *J. Med. Chem.* **2005**, *48*, 5648.
23. Hu, L.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A. *Protein Struct. Funct. Bioinform.* **2005**, *60*, 333.
24. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
25. www.molinspiration.com.
26. Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. *Acta Crystallogr., Sect. B* **1983**, *39*, 480.
27. Park, H.; Yeom, M. S.; Lee, S. *ChemBioChem* **2004**, *5*, 1662.
28. Bartova, I.; Koca, J.; Otyepka, M. *Protein Sci.* **2008**, *17*, 22.
29. Heaslet, H.; Rosenfeld, R.; Giffin, M.; Lin, Y. C.; Tam, K.; Torbett, B. E.; Elder, J. H.; McRee, D. E.; Stout, C. D. *Acta Crystallogr., Sect. D* **2007**, *63*, 866.
30. Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. *J. Med. Chem.* **2004**, *47*, 45.
31. Liang, J.; Edelsbrunner, H.; Woodward, C. *Protein Sci.* **1998**, *7*, 1884.
32. Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Dardy, G.; Jones, Y.; Stuart, D.; Stammers, D. *Nat. Struct. Biol.* **1995**, *2*, 293.
33. Klabunde, T.; Petrassi, H. M.; Oza, V. B.; Raman, P.; Kelly, J. W.; Sacchettini, J. C. *Nat. Struct. Biol.* **2000**, *7*, 312.