proceedings

Structure-Based Site of Metabolism (SOM) prediction of ligand for CYP3A4 Enzyme: Comparison of Glide XP and Induced Fit Docking (IFD)†

Deepak K. Lokwani[1]*, Aniket P. Sarkate[2], Anna Pratima G. Nikalje[3], Julio A. Seijas[4]

[1] Department of Pharmaceutical Chemistry, R. C. Patel Institute of Pharmaceutical Education and Research, Shirpur, Dist-Dhule, Maharashtra, India
[2] Department of Chemical Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
[3] Wilson College, Chowpatty Seaface Road, Mumbai, India
[4] Departamento de Química Orgánica, Facultad de Ciencias, Universidad of Santiago De Compostela, Alfonso X el Sabio, Lugo 27002, Spain
* Correspondence: dklokwani@gmail.com
† Presented at The 23rd International Electronic Conference on Synthetic Organic Chemistry (15 Nov 2019-15 Dec 2019).

Abstract: Metabolism is one of the prime reasons where most of the drugs fail to accomplish their clinical trials. The enzyme CYP3A4, which belongs to the superfamily of cytochrome P450 enzymes (CYP) helps in the metabolism of a large number of drugs in the body. The enzyme CYP3A4 catalyzes mainly the oxidative chemical processes and also shows a very broad range of ligand specificity. Understanding of compound's structure where oxidation would take place is crucial for the successful modification of molecules in order to avoid unwanted metabolism and to increase its bioavailability. For this reason, it is required to know the site of metabolism (SOM) of the compounds, where compounds undergo enzymatic oxidation. It can be identified by predicting the accessibility of the substrate's atom toward oxygenated Fe atom of heme in a CYP protein. The CYP3A4 enzyme is a highly flexible enzyme and can take significantly different conformations depending on the ligand with which it is being bound. Here in, we studied the ability of the Glide XP and Induced Fit docking (IFD) tool of Schrodinger software suite to reproduce the binding mode of co-crystalized ligands into six X-ray crystallographic structures. We extend our studies for the prediction of SOM for compounds whose experimental SOM is reported but ligand-enzyme complex crystal structure is not available in Protein Data Bank (PDB). It was observed that IFD reproduces the exact binding mode of available co-crystallized structures and correctly predicted the SOM of experimentally reported compounds.

Keywords: CYP3A4; Glide XP; Induced Fit Docking (IFD); Site of Metabolism (SOM)

1. Introduction

CYP3A4 is the most important enzyme in the superfamily Cytochrome P450 that, besides endogenous compounds, metabolizes about half of the currently marketed drugs [1]. For a given substrate, CYP enzymes can catalyze a wide variety of reactions, such as hydroxylation, epoxidation or heteroatom oxidation, dealkylations, and desaturation [2]. The CYPs metabolize foreign compounds into polar hydrophilic metabolites by integrating one oxygen atom into the substrate, making it water-soluble and consequently more easily excreted [3, 4]. The CYP3A4 exhibits very wide ligand specificity and catalyzes a large kind of chemical processes. Therefore inhibition of metabolic activity of CYP3A4 by one substrate, can extensively influence the metabolism of other substrate. For this reason, prediction of CYP inhibition by compounds and/or its metabolic stability

along with other ADMET profile are important for assessing the quality of lead molecule. To reduce the CYP inhibition and enhance the metabolic stability of compounds, predictions of sites of metabolism (SOM) or binding modes with CYPs are quite helpful in addition to assays of the inhibition and stability in experiments. Using the predicted SOM, a substitute around the sites can be converted into a functional group which is metabolically more stable. Thus, in order to design new compounds which are less prone to metabolism it is necessary to know the site of metabolism (SOM) in their structure. The SOM can be identified by predicting the accessibility of the substrate's atom toward oxygenated Fe atom of heme in a CYP protein.

To metabolize such a variety of compounds, CYP3A4 has to be an extremely flexible structure and its X-ray crystals structure have shown that the CYP3A4 can adopt substantially different conformations depending on the ligand with which it is being co-crystallized [5]. The available native (ligand unbound) and ligand bound crystal structures of CYP3A4 in Protein Data Bank (PDB) have showed the high flexibility of amino acid residues in side chain and loop region, thus making binding pocket in a closed and open conformation respectively. As reported by Yuki H. et. al. [6], we also superimposed six different crystal structures of CYP3A4 and root-mean-square deviation (RMSD) between the docked ligand poses and their native poses were calculated. The RMSD values of superimposition of C-alpha carbon atoms of crystal structures were obtained in range of 1.58 Å to 10.54 Å **(Table 1)**. Based on RMSD values it is cleared that CYP enzyme is highly flexible and confirmed that every ligand induces different conformational changes into binding pocket of CYP3A4 enzyme **(Figure 1)** and thus make ligand-enzyme interaction studies difficult. The main difference found between two unbound crystal structures of CYP3A4 is in the orientation of Arg212. In one structure (PDB code 1W0E), Arg 212 is orientated away from heme group and in another unbound structure (Pdb code 1TQN), it occupies the orientation towards heme group [7]. Whereas in a structure bounded with ketoconazole (PDB code 2V0M), CYP3A4 have lot conformational changes and Arg212 is found away from active site. Some hydrophobic cluster in above ketoconazole bound structure is shown broken which increases the volume of active site. Dan Fishelovitch et. al., have reported that Arg212 may strongly interact with Phe304 and thus affect the binding of substrates/inhibitors on the enzyme.

Table 1: Superimposition of C-alpha carbon atoms of crystal structure of CYP3A4

	1TQN[1]	**1W0E**[1]	**1W0F**[2,3]	**1W0G**[2]	**2V0M**[2]	**3NXU**[2]
1TQN	-	7.23	7.25	7.40	6.16	4.58
1W0E	7.23	-	1.66	2.10	10.54	8.65
1W0F	7.25	1.66	-	1.58	10.59	8.72
1W0G	7.40	2.10	1.58	-	10.67	8.83
2V0M	6.16	10.54	10.59	10.67	-	4.19
3NXU	4.58	8.65	8.72	8.83	4.19	-

[1] Native (ligand unbound) Crystal Structures; PDB ids 1TQN and 1W0E

[2] Ligand bound Crystal Structures: PDB ids 1W0F (Progesterone), 1W0G (Metyrapone), 2V0M (Ketoconazole) and 3NXU (Ritonavir)

[3] PDB Id 1W0F: Progesterone bound at peripheral site, 17 Å away from heme

In this study, we have taken theses six PDBs of CYP3A4 and first performed re-docking of all bound co-crystalized ligands and then performed cross-docking on all six PDBs using Glide XP and Induced fit docking methodology. We extended our studies for prediction of SOM for ten different ligands obtained from literature whose experimental SOM is reported.

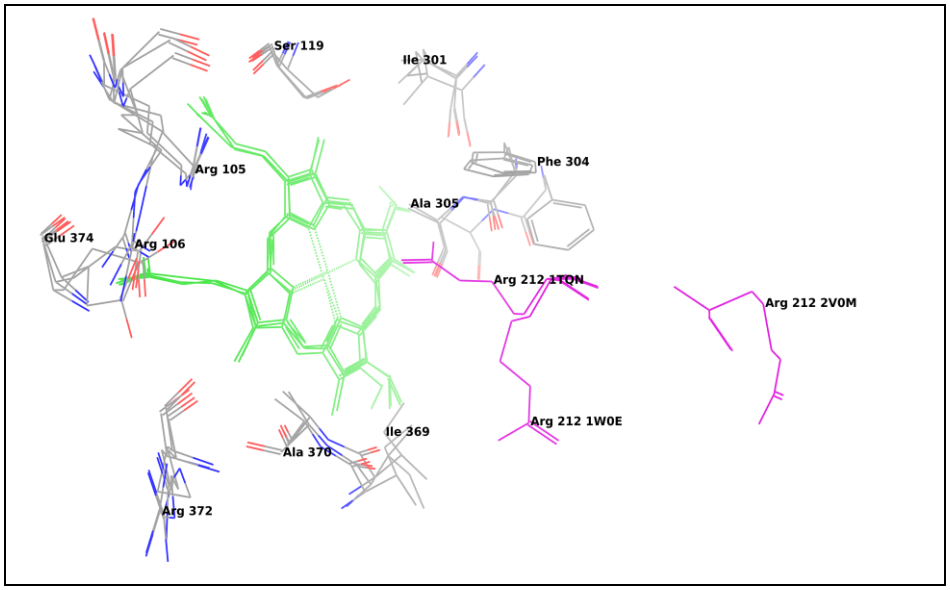


Figure 1: Superimposition of three Crystal Structures of CYP3A4

2. Materials and Methods

2.1. Glide ligand docking

Molecular Docking Studies were performed in Maestro 9.1 using Glide v6.8 (Schrodinger, LLC, New York, NY, USA). All compounds were built using Maestro build panel and optimized to lower energy conformers using Ligprep v3.5.9 (Schrodinger, LLC). The PDB's 1W0F (Progesterone), 1W0G (Metyrapone), 2V0M (Ketoconazole) and 3NXU (Ritonavir) were taken from RCSB Protein Data Bank and prepared for docking using 'protein preparation wizard' in Maestro v10.3. (Schrodinger, LLC) The bond orders and formal charges were added for heterogroups and hydrogens were added to all atoms in the structure. Side chains that are not close to the binding cavity and do not participate in salt bridges were neutralized and termini were capped by adding ACE and NMA residue. After preparation, the structure was refined to optimize the hydrogen bond network using OPLS_2005 force field. The minimization was terminated when the energy converged or the RMSD reached a maximum cutoff of 0.30 Å. The extra precision (XP) docking mode for all compounds was performed on generated grid of protein structure. The final evaluation of ligand-protein binding was done with Glide score.

2.2 Induced fit docking

IFD was performed using the module Induced Fit Docking of Maestro v9.1. (Schrodinger, LLC). The entire receptor molecule constrained minimized with an RMSD cutoff of 0.18 Å was selected for generation of centroid of the residues and the box size was generated automatically. The initial Glide docking for each ligand was carried out. Side chains were trimmed automatically based on B-factor, with receptor and ligand van der Waals scaling of 0.70 and 0.50, respectively; and the number of poses generated were set to be 20. Prime side chain prediction and minimization was carried out in which residues were refined within 5.0 Å of ligand poses and side chains were optimized. This leads to a ligand structure and conformation that is induced fit to each pose of the receptor structure. Finally, Glide XP redocking was carried out into structures within 30.0 kcal/mol of the best structure, and within the top 20 structures overall. The ligand was rigorously docked into the induced-fit receptor structure and the results yielded an IFD score for each output pose.

3. Result and Discussion

3.1. Re-docking study

The re-docking study was conducted to review the docking methodology to predict the correct ligand pose within the active enzyme site. Of the four ligand-bound crystal structures, the crystal structure belonging to the PDB I d 1W0F was not chosen for re-docking as the inbound ligand, Progesterone bound at the peripheral site 17 Å away from the heme. The inbound ligand was extracted and redocked using Glide XP and induced fit docking (IFD) technique for the remaining three crystal structures. For each docked pose, the RMSD was calculated by superimposing on their respective inbound crystallized ligand. Table 2 overview RMSD values for poses of docked ligands when superimposed on inbound ligand. To compare the docking methodology, the mean RMSD of all poses of each ligand was also calculated. From mean RMSD values of all poses and RMSD of best poses with low RMSD value, it was observed that IFD is more thoroughly predicted and recollected the accurate poses of the ligand. Although the mean RMSD of poses of Ritonavir in PDB 3NXU was higher for IFD as compare to Glide XP docking, but both methodology predict the accurate pose for Ritonavir.

Table 2: Superimposition of ligands after docking over corresponding crystal structure of CYP3A4

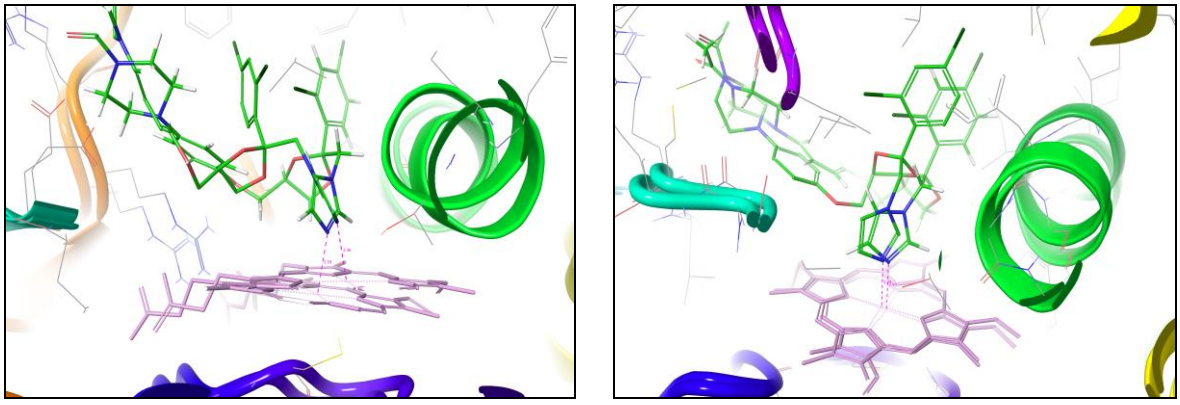
PDB ID	RMSD After Superimposition on Crystal Structure (Å)					
	IFD			Glide XP		
	Max.	Min.	Mean	Max.	Min.	Mean
			Ketoconazole			
2V0M	3.34	1.18	2.11	2.17	2.12	2.14
			Metyrapone			
1W0G	2.23	2.23	2.23	2.55	2.55	2.55
			Ritonavir			
3NXU	6.21	2.39	4.36	5.36	2.31	3.79

3.2. *Cross docking Studies*

All three ligands ketoconazole, metyrapone and ritonavir have been docked to all six crystal structures to validate both the IFD and Glide XP docking procedures further. Similar to redocking studies, the RMSD value was determined by superimposing all docked pose of each compound over respective inbound ligand pose. In most of PDBs, IFD regenerate the same poses for both compounds having RMSD value for best pose below 2.0 Å whereas best pose predicted by Glide XP have RMSD value above 2.0 Å **(Table 3)**. It was noted from **Figure 2** that IFD more accurately predict pose for ketoconazole and metyrpone which is similar to their corresponding reported crystal structures.

Table 3: Superimposition of ligands after docking over all crystal structure of CYP3A4

PDB ID	RMSD After Superimposition on Crystal Structure (Å)					
	IFD			Glide XP		
	Max.	Min.	Mean	Max.	Min.	Mean
			Ketoconazole			
1TQN	3.41	1.98	2.60	3.52	2.65	3.08
1W0E	3.12	1.76	2.38	2.79	2.79	2.79
1W0F	3.64	2.30	3.02	3.49	3.10	3.25
1W0G	4.06	1.82	2.87	2.83	2.83	2.83
2V0M	3.34	1.18	2.11	2.17	2.12	2.14
3NXU	3.78	2.26	3.11	3.59	3.24	3.42
			Metyrapone			
1TQN	-	-	-	-	-	-
1W0E	1.89	1.89	1.89	-	-	-
1W0F	-	-	-	2.30	2.30	2.30
1W0G	2.23	2.23	2.23	2.55	2.55	2.55
2V0M	1.39	1.39	1.39	2.51	2.51	2.51
3NXU	2.18	2.18	2.18	-	-	-
			Ritonavir			
1TQN	4.92	4.33	4.67	4.27	4.27	4.27
1W0E	4.61	3.49	4.05	4.69	3.91	4.15
1W0F	-	-	-	4.21	4.00	4.10
1W0G	4.99	3.88	4.43	4.59	3.74	4.09
2V0M	5.62	3.31	4.32	5.01	2.98	3.89
3NXU	6.21	2.39	4.36	5.36	2.31	3.79



(a) **(b)**

Figure 2: Superimposition of pose of ketoconazole docked in (a) PDB id 1TQN and (b) PDB id 2V0M by IFD over crystal structure of ketoconazole bound CYP3A4 (PDB id 2V0M)

3.3 Prediction of SOM

After the comparison and validation of both docking methodology, we extend our aim to predict the SOM in those compounds whose ligand-enzyme complex crystal structure is not available in PDB. For this, the ten compounds were selected from literature whose SOM was reported by experimental methodology but their bound to co-crystalized CYP3A4 structure not available **(Figure 3).** All these compounds are docked in all 6 PDBs by both Induced fit and Glide XP methodology. The best ligand-enzyme complexes from cluster were selected for each ligand as per desired metabolic pose of drug candidate, and further analyzed by measuring distance of ligand reported SOM atom from heme Fe atom of a CYP protein **(Table 4).**

It was cleared that IFD generated number of desired metabolic poses for each compounds except Verapamil as compared to Glide XP docking. After analyzing data, it was also seen that distance between reported SOM atom of Ketoconazole, Metyrapone, Ritonavir and Tamoxifen and heme Fe of CYP was found to within 2.5 Å for docking poses genereated by IFD. Whereas Glide XP generated the poses for each compounds where SOM atom is far away from heme Fe atom. Thus it can reveled that IFD generates desired metabolic poses for most of compounds **(Figure 4)** and Glide XP fails to predict desirable poses in active site of CYP3A4.

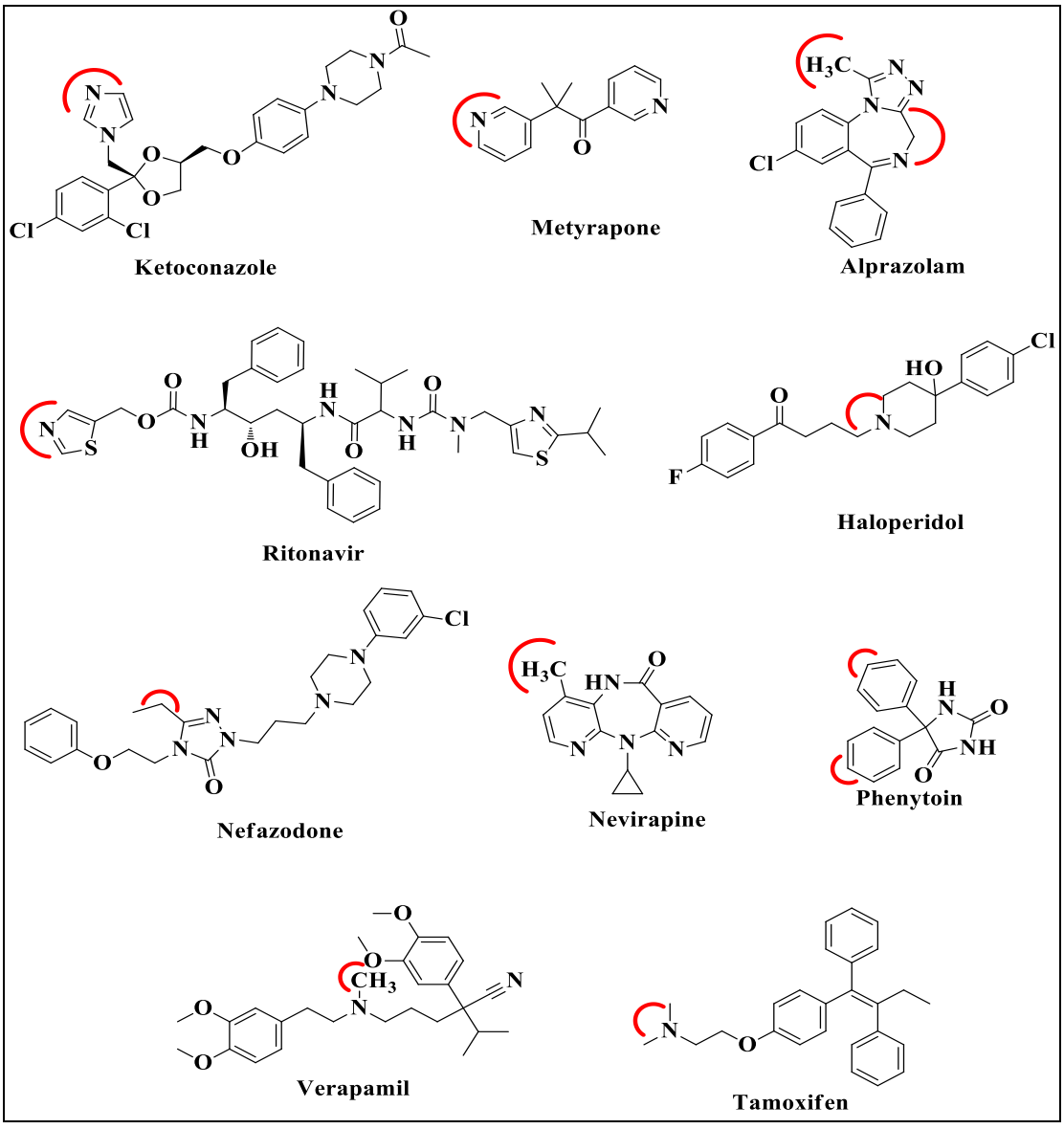
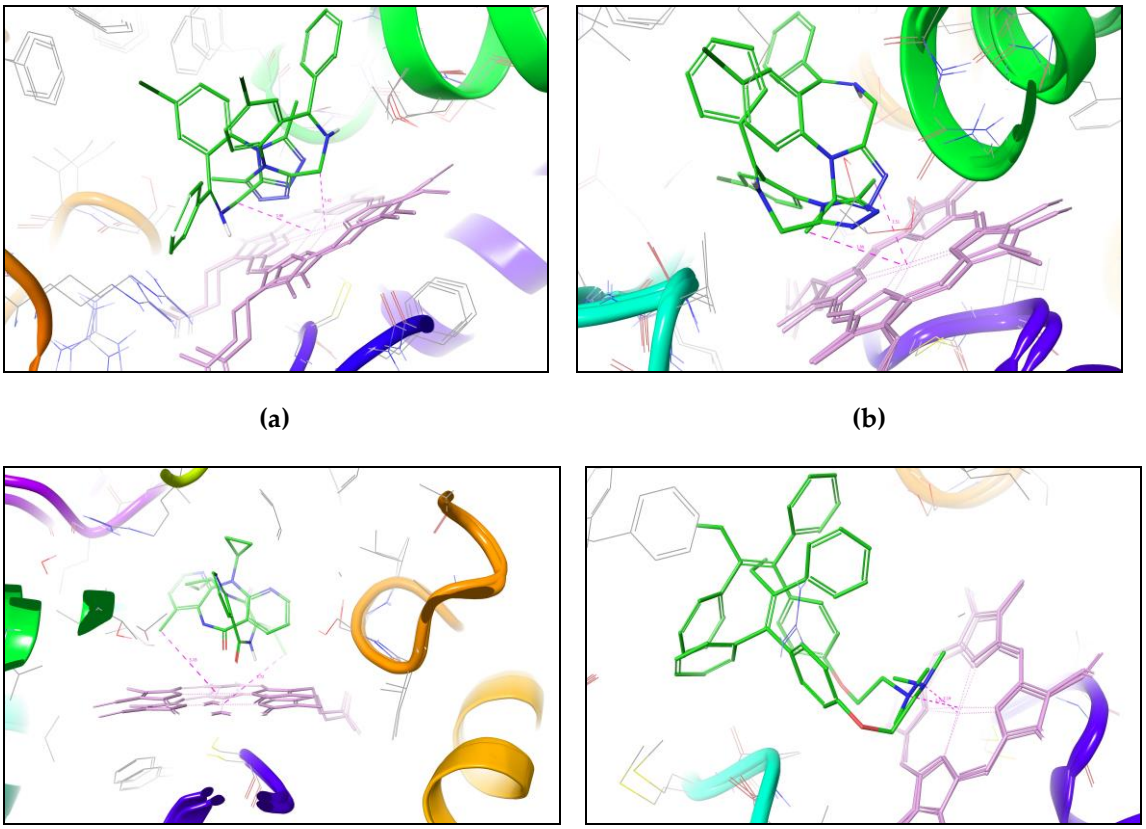


Figure 3: Chemical Structures of compounds. Red color indicate experimentally reported SOM

Table 4: Comparison of IFD and Glide XP results for docking of ligands in active site of CYP3A4

PDB ID	IFD					Glide XP				
	Generated Pose	Desired Metabolic Pose	Distance of Atom			Generated Pose	Desired Metabolic Pose	Distance of Atom		
			Max	Min	Mean			Max	Min	Mean
Ketoconazole (PDB Code 2V0M, Dist from Heme Fe: 2.59Å)										
1TQN	21	5	2.40	2.17	2.32	8	2	3.78	3.21	3.49
1W0E	27	11	2.39	2.21	2.33	6	1	3.26	3.26	3.26
1W0F	31	6	2.40	2.33	2.36	8	3	3.92	3.61	3.76
1W0G	42	15	2.40	2.19	2.32	4	1	3.82	3.82	3.82
2V0M	88	18	2.39	2.15	2.31	3	2	2.81	2.52	2.66
3NXU	88	16	2.40	2.28	2.36	4	4	3.92	3.40	3.56
Metyrapone (PDB Code 1W0G, Dist from Heme Fe: 2.36Å)										
1TQN	15	0	-	-	-	2	0	-	-	-
1W0E	9	1	2.34	2.34	2.34	0	0	-	-	-
1W0F	17	0	-	-	-	2	1	3.54	3.54	3.54
1W0G	3	1	2.39	2.39	2.39	1	1	2.69	2.69	2.69
2V0M	17	1	2.39	2.39	2.39	1	1	3.35	3.35	3.35
3NXU	22	1	2.39	2.39	2.39	1	0	-	-	-
Ritonavir (PDB Code 3NXU, Dist from Heme Fe: 2.42Å)										
1TQN	4	3	2.34	2.09	2.24	32	1	3.20	3.20	3.20
1W0E	4	2	2.27	2.17	2.22	10	4	3.68	2.83	3.34
1W0F	0	0	-	-	-	32	3	3.57	3.23	3.35
1W0G	23	2	2.37	2.35	2.36	14	7	3.64	2.53	3.00
2V0M	58	26	2.40	1.92	2.31	29	18	3.73	2.47	3.00
3NXU	61	53	2.40	1.96	2.30	24	12	3.99	2.65	3.21
Alprazolam[a] (includes both SOM)										
1TQN	6	6	4.31	3.90	4.14	2	0	-	-	-
1W0E	52	52	4.77	3.42	4.18	2	2	3.90	3.84	3.87
1W0F	28	28	4.36	3.81	3.99	2	1	4.66	4.66	4.66
1W0G	72	72	4.52	3.55	4.07	1	1	4.58	4.58	4.58
2V0M	78	78	4.66	3.30	4.03	1	0	-	-	-
3NXU	74	74	4.65	3.28	4.15	1	1	3.56	3.56	3.56
Haloperidol[b]										
1TQN	40	2	5.33	5.33	5.33	2	0	-	-	-
1W0E	82	38	5.98	5.45	5.75	0	0	-	-	-
1W0F	78	12	5.99	5.66	5.91	3	0	-	-	-
1W0G	72	10	5.85	5.37	5.68	0	0	-	-	-
2V0M	96	26	5.99	4.91	5.63	1	0	-	-	-
3NXU	52	10	5.75	5.49	5.64	3	2	5.73	5.64	5.68

Nefazodone[b]										
1TQN	2	0	-	-	-	2	0	-	-	-
1W0E	8	0	-	-	-	0	0	-	-	-
1W0F	24	2	5.25	5.25	5.25	2	2	6.00	6.00	6.00
1W0G	60	6	5.74	5.12	5.44	1	0	-	-	-
2V0M	12	4	6.00	5.11	5.55	1	1	3.98	3.98	3.98
3NXU	32	32	4.15	3.32	3.72	0	0	-	-	-
Nevirapine[b]										
1TQN	9	0	-	-	-	2	0	-	-	-
1W0E	26	6	5.59	5.42	5.51	0	0	-	-	-
1W0F	40	14	5.96	5.28	5.62	2	0	-	-	-
1W0G	62	30	5.95	4.76	5.40	0	0	-	-	-
2V0M	74	30	5.94	4.98	5.56	0	0	-	-	-
3NXU	16	4	5.95	5.72	5.83	0	0	-	-	-
Phenytoin[b]										
1TQN	1	0	-	-	-	3	0	-	-	-
1W0E	62	0	-	-	-	1	0	-	-	-
1W0F	52	0	-	-	-	3	2	4.50	3.62	4.06
1W0G	60	0	-	-	-	2	0	-	-	-
2V0M	84	16	5.91	5.70	5.84	2	0	-	-	-
3NXU	80	6	5.78	5.16	5.57	3	0	-	-	-
Tamoxifen[a]										
1TQN	10	10	2..37	2.26	2.32	2	0	-	-	-
1W0E	2	2	2.33	2.33	2.33	1	1	2.96	2.96	2.96
1W0F	38	34	5.00	2.19	2.79	2	0	-	-	-
1W0G	60	46	4.96	2.19	4.06	0	0	-	-	-
2V0M	18	18	4.61	2.38	2.88	1	1	4.26	4.26	4.26
3NXU	50	32	2.40	2.35	2.38	2	1	3.85	3.85	3.85
Verapamil[b]										
1TQN	46	0	-	-	-	6	0	-	-	-
1W0E	80	0	-	-	-	6	0	-	-	-
1W0F	48	0	-	-	-	5	0	-	-	-
1W0G	72	0	-	-	-	6	0	-	-	-
2V0M	146	0	-	-	-	6	0	-	-	-
3NXU	128	0	-	-	-	6	0	-	-	-



(a) **(b)**

(c) **(d)**

Figure 4: Superimposition of IFD docking pose of lignads in CYP3A4 enzyme (a) Alprazolam (1W0E-1W0G)-Site 1 (b) Alprazolam (1W0E-1W0G)-Site 2 (c) Nevirapin (1W0F-3NXU) (d) Tamoxifen (1W0G-1TQN)

3. Conclusion

As the protein flexibility issue concerned with CYP3A4, Glide XP docking is unable to predict the desired metabolic pose of some of ligands and IFD produces at least one desired ligand-enzyme complex for number of ligands for all crystal structures. IFD therefore was found to be one of the reliable methods for predicting and analyzing the site of metabolism (SOM) of ligands in the flexible binding pocket of CYP3A4.

Conflicts of Interest: "The authors declare no conflict of interest."

References

1. Brown, C.M.; Reisfeld, B.; Mayeno, A.N. Cytochromes P450: A structure-based summary of biotransformations using representative substrates. *Drug Metab. Rev.* 2008, *40*, 1–100.
2. Guengerich, F.P. Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity. *Chem. Res. Toxicol.* **2001**, *14*, 611–650.
3. Anzenbacher, P.; Anzenbacherová, E. Cytochromes P450 and metabolism of xenobiotics. *Cell. Mol. Life Sci.* 2001, *58*, 737–747.
4. Prusis, P.; Afzelius, L. Improvement of Site of Metabolism Predictions for CYP3A4 by Using Discriminant Analysis of Compound Preference of CYP3A4 X-Ray Structural Conformers

and Subsequent Docking. **2009**, 865–868.

5. Williams, P.A.; Cosme, J.; Matak Vinković, D.; Ward, A.; Angove, H.C.; Day, P.J.; Vonrhein, C.; Tickle, I.J.; Jhoti, H. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science (80-.).* **2004**, *305*, 683–686.
6. Yuki, H.; Honma, T.; Hata, M.; Hoshino, T. Prediction of sites of metabolism in a substrate molecule , instanced by carbamazepine oxidation by CYP3A4. *Bioorg. Med. Chem.* **2012**, *20*, 775–783.
7. Fishelovitch, D.; Hazan, C.; Shaik, S.; Wolfson, H.J.; Nussinov, R. Structural Dynamics of the Cooperative Binding of Organic Molecules in the Human Cytochrome P450 3A4. **2007**, 1602–1611.