Ligand Docking Methods to Recognize Allosteric Inhibitors for G-protein Coupled Receptors

Harini K
National Centre for Biological Sciences
Jayashree S
Royal Melbourne Institute of Technology: RMIT University
Vikas Tiwari
NCBS: National Centre for Biological Sciences
Sneha Vishwanath
University of Cambridge
Ramanathan Sowdhamini (mini@ncbs.res.in)
National Centre for Biological Sciences

Research

Keywords: G-protein coupled receptors, allosteric ligands, AutoDock, cognate ligands, Tanimoto co-efficient

Posted Date: March 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-218684/v1

Abstract

Background

G-protein coupled receptors (GPCRs) are large protein families known to be important in many cellular processes. They are well known for their allosteric activation mechanisms. They are drug targets for several FDA-approved drugs. We have investigated the diversity of the ligand binding site for these class of proteins against their cognate ligands using computational docking, even if their structures are known in the ligand-complexed form.

Results

The cognate ligand of some of these receptors dock at allosteric binding site, with better score than the binding at the conservative site. Further, ligands obtained from GLASS database, which consists of experimentally verified GPCR ligands, also show allosteric binding to GPCRs. The allosteric binders show strong affinity to the binding site, though the residues at the binding site are not conserved across GPCR subfamilies.

Conclusions

Based on our computational approach it was found that the residues at the allosteric site are not as conserved as in the cognate binding site, which might explain the specificity of a particular GPCR. Further, for certain GPCRs, some of their known cognate ligands were predicted to have better binding preference towards the allosteric site than orthosteric site and therefore this computational approach can assist in the prediction of allosteric binders for GPCRs.

Background

Membrane proteins constitute nearly 40% of the human genome [1,2]. Amongst the membrane proteins, G-protein coupled receptors (GPCRs) are one of the most studied [3–5] since they are important drug targets. There are close to 800 GPCRs known in the human genome and they are grouped into 5 families depending on their substrates, such as peptides, amines [6]. GPCRs act as one of the important environmental sensors and are key to diverse signaling processes. As a result, apart from the variation in substrates, there is enormous diversity in their position in different biological pathways. Mutations in either GPCRs or their interacting proteins have been widely implicated in neurodegenerative and other diseases [7,8].

The hallmark of GPCRs is the presence of seven transmembrane helices (TMHs), where extracellular loop regions and parts of TMHs recognize the cognate substrates and indentations in the intracellular loop regions provide the capacity to participate in diverse biological pathways [9]. Since the past two decades, there has been substantial insights into the structural features of few important GPCRs, despite the inherent challenges in the structure determination of membrane proteins. This has meant that we now

have a great deal of information about the conformational changes that occur subsequent to ligand binding, distinct sites for agonist and antagonist binding etc. Characteristic functional motifs in the intracellular regions, the presence of conserved charged residues in the intracellular face and the presence of prolyl residues are known to contribute to signaling and conformational changes [10]. We now have close to 389 PDB entries (GPCR-EXP database) which pertain to GPCRs, either in the apo-form or ligand-bound form, as well as agonist or antagonist-bound states. There has been successful design of drugs, such as haloperidol [11], over the years. Till November 2017, FDA has approved drugs against 134 GPCRs [12].

Despite the structural insights available, drug design for GPCRs remains highly challenging due to the inherent characteristics of substrate promiscuity, structural similarities of drug molecules and sequence similarity within subfamilies of GPCRs. However, there has been continuing demands to provide drug solutions to address this prevailing feature of GPCRs. Some of the important GPCR drug targets are in the area of neurodegenerative diseases, where there is appalling amounts of side effects observed in patients who are treated with GPCR drugs. Recent efforts have, therefore, focused on the design of allosteric inhibitors [13–15].

In this paper, we report systematic computational ligand docking experiments using select GPCRs with known information on ligand binding, to present which factors enable best capture of near-native ligand binding and how this can be employed to identify allosteric inhibitors. Our choice of ligand docking algorithm is AutoDock which is one of the well-known software that has gone through rigorous analyses by other groups as well [16]. We first describe the analysis of binding poses, where different levels of known information on structural or evolutionary conservation can guide the docking process. Objective measures, such as Tanimoto co-efficient, have been employed to assess the comparison between docking poses and the structural data for ligand binding. This approach enables the identification of structurally reasonable docking poses which are not close to the native pose, namely allosteric binding. Finally, we have performed blind docking on one of the GPCRs, chemokine receptors, using specialized set of ligands known to bind GPCRs as organized in GLASS database. Few allosteric binding sites can be recognized using this novel computational approach, which was found to be stable as evidenced using molecular dynamics simulations.

Methods

Selection of GPCR structures and ligands for analysis

All the available GPCR structures were downloaded from PDB (Protein Data Bank, April 2017) [17]. There were 71 GPCR structures. There are several structures with different resolution and agonist/antagonist bound forms for single receptor. Thus, for every GPCR, one best structure with highest resolution was selected that was bound to ligand. A set of 27 GPCRs were selected for final analysis (Table 1). GPCR alignments were obtained from GPCRdb for the subfamily of GPCR receptors to study evolutionary conservation[18]. Ligands were retrieved from PDB, PubChem [19] or GLASS database [20]. GLASS

database is a resource to retrieve ligands that are known to bind to GPCRs and were used to explore allosteric binding. Out of 871 ligands, 45 were chosen based on Lipinski's drug likeness and XlogP (less than 2). The 45 ligands were clustered using ChemMine tool and finally 29 compounds were used for docking. These compounds were converted to PDB format using Open babel tool.

Table 1: List of GPCRs used in the study

Serial No.	GPCR name
GPCR-1	5-HT1B receptor
GPCR-2	5-HT2B receptor
GCPR-3	A2A receptor
GPCR-4	M2 receptor
GPCR-5	M3 receptor
GPCR-6	β1-adrenoceptor
GPCR-7	β2-adrenoceptor
GPCR-8	AT1 receptor
GPCR-9	CCR5
GPCR-10	CRF1 receptor
GPCR-11	CXCR4
GPCR-12	D3 receptor
GPCR-13	FFA1 receptor
GPCR-14	mGlu5 receptor
GPCR-15	H1 receptor
GPCR-16	LPA1 receptor
GPCR-17	δ receptor
GPCR-18	κ receptor
GPCR-19	μ receptor
GPCR-20	NOP receptor
GPCR-21	Rhodopsin
GPCR-22	OX2 receptor
GPCR-23	P2Y1 receptor
GPCR-24	P2Y12 receptor
GPCR-25	PAR1
GPCR-16	S1P1 receptor
GPCR-27	SMO

GPCR proteins were selected based on the highest resolution structure available in PDB**.** All class-A GPCRs structures were selected from PDB and a non-redundant dataset was obtained.

Grid setting for GPCR-ligand docking

The protein and ligand coordinates were separated from PDB and used for docking. In two out of three experiments, ligand coordinates were derived from PubChem (please see below). Both the agonist and the antagonist bind in the extracellular domains for most of the GPCRs in distinct sites. Hence, the grid region of choice for docking, were chosen in the extracellular region. However, different grid settings were considered in order to identify the parameter in grid setting that will help in correct prediction of ligand binding to GPCRs. These are as follows:

Level 1: Ligand coordinates were taken as such from PDB, but ligand coordinates were kept flexible for docking. Only the extracellular half of the receptor region was selected as grid box (semi-blind docking).

Level 2: Ligand coordinates were taken from PubChem and ligand was kept flexible. A smaller grid that covers just the known ligand binding site was selected as grid box (guided docking).

Level 3: Ligand coordinates were taken from PubChem and ligand was kept flexible. Only the extracellular half of the receptor region was selected as grid box (semi-blind docking). (**Fig 1**)

Molecular docking

AutoDock 4.2.6 was used for docking studies [16]. Protein and ligand structures were prepared using ADT (AutoDock Tool). The hydrogen atoms were added and water molecules were removed followed by addition of gasteiger charges. The grid parameter file was generated with default distance (0.375 Å) between grid points. Grid space was defined around the extracellular site of GPCRs. The search parameter was set to 100 Genetic Algorithm (GA) runs and Lamarckian genetic algorithm was used. Autogrid was run followed by AutoDock. Similar parameters were used for all ligands docking to a single protein. The docking results were analyzed and protein-ligand complex was made using ADT.

Calculation of Tanimoto co-efficient and identification of allosteric sites

To identify how similar the protein-ligand docking results are to the X-ray crystallographic structure of complexes, we used the Tanimoto co-efficient score. Any atom within 4Å distance was considered to be a contact with the ligand.

Let ‘a’ be the contacts in PDB structure and ‘b’ be the contacts in the AutoDock complex then the Tanimoto co-efficient (TC) is calculated as,

$$\text{TC value} = \frac{\text{No.of common contacts } (a \sqcap b)}{\text{---}}$$

(No. of contacts in a + No. of contacts in b)- (a ⊓ b)

Further, for each ligand, all 100 poses were checked for their binding site compared to native ligand. The minimum distance between all the atoms of native ligand and all the atoms of docked ligand was calculated and if the minimum distance is more than 5 Å for any pose, then that pose was considered as allosteric site binding pose.

Molecular Dynamics (MD) simulation

The protein-ligand complex structure was subjected to MD simulation using Desmond module of Schrodinger to assess the stability of complex [21]. Initially the complex structure was processed using protein preparation wizard of maestro (Schrodinger Release 2019-4: Maestro, Schrodinger, LLC, New York, NY, 2019) which assigns bond orders, sets protonation state, optimizes H-bonds and then minimizes the structure [22]. The structure (CXCR4-ligand complex) was protonated at pH 5.5 before minimization in protein preparation wizard. The membrane information was retrieved from OPM database [23]. After processing, the structure was solvated in the presence of membrane (POPC) using TIP4P water model. Orthorhombic box, with buffer distance of 10 Å, was used followed by minimization of box size. The system was neutralized and 150 mM salt (NaCl) was added. System builder was run and output of system builder (solvated system) was used for MD simulation. The system was subjected to default relaxation protocol of Desmond followed by production MD run for 100 nanoseconds. Simulation results were analyzed using Simulation Interaction Diagram and Simulation Event analysis modules of Schrodinger. The analysis was performed for entire range of simulation time. RMSD is calculated for each frame by aligning the complex to protein backbone of the reference frame. Significantly higher values of "Lig fit Prot" than protein RMSD signifies the diffusion of ligand away from its initial binding site. "Lig fit lig RMSD" is calculated by aligning the ligand on the reference ligand conformation and it indicates the internal fluctuation of ligand.

Results

Prior knowledge of the binding mode helps to predict correct protein-ligand interaction

Twenty-seven GPCRs from PDB were selected for analysis. Each of the protein ligand complex was separated and docking was performed as mentioned in Methods. The three different grid conditions were used as mentioned in Methods (Level 1, 2 and 3) (**Fig 1**). We observed that 50% of the instances, when the ligand coordinates were taken from PDB and semi-guided docking was performed, we could get the protein-ligand complex with high TC score with the original PDB structure and AutoDock energy. In Level 2 of docking, runs with use of ligand coordinates from PubChem and guided docking, complexes with high TC score and low energy could be obtained in 30% of instances. Level 3 (ligand from PubChem and semi-guided docking), however, did not result in complexes similar to the original structure in most of the cases (only 3 out of 27 GPCRs had high TC scores) (**Fig 2**: TC score Vs Energy correlation plot for Level 1).

Identifying novel binding modes using the TC score and energy values for GPCR-ligand complexes

We selected complexes which had low TC score (that indicates a different binding mode compared to original ligand binding site (hereafter referred as 'OLBS')) and good (low) docking energy (that indicates stable binding) on the basis of Level 1 results. First three highest scoring complexes were selected for detailed analysis of the docking poses. These include beta-2-adrenergic receptor, chemokine receptor and 5-HT2B receptor. It was observed that for beta-2-adrenergic receptor and 5-HT2B receptors the ligand binds in the same pocket as in original complex, but is displaced along Y-axis of the protein with a large overlap with OLBS. (**Fig 3a and 3b**). This results in low TC score for these complexes. For chemokine receptor, however, we could find novel binding sites for the cognate ligand with minimum overlap with OLBS (**Fig 3c**). Thus, we used chemokine receptor-ligand complex for further analysis on allosteric binding in GPCRs.

Residues at the alternate novel binding sites are not conserved

Twenty three homologous sequences of chemokine receptors were used to identify conservation of residues at the endogenous binding site and allosteric binding site. It was observed that 7 out of 12 residues at OLBS were conserved, while only 3 out of 12 residues were evolutionarily conserved at the allosteric site, even within the same subfamily of receptors. This clearly suggests that allosteric binding site is novel and evolving (**Fig 4**).

Identifying novel allosteric binders from GLASS database for chemokine receptor

In order to explore allosteric binders for CXCR4, ligands were retrieved from GLASS database. The docking results for each ligand were analyzed in terms of best docking pose with minimum docking energy among all poses and best docking pose among the largest cluster. The docking energy of best docking pose for ligands varied from -6.25 kcal/mol (for ligand "25178561")to -11.27 kcal/mol (for ligand "483559"). Out of 29 ligands, the overall best pose (in terms of docking energy) for 6 ligands also belongs to the largest cluster. Similar analysis was performed for all the ligands. Interestingly, it was observed that all 100 poses are allosteric poses for ligand "76381" (**Fig 5**). The best pose of ligand "76381" in terms of best docking energy is also the best pose among the largest cluster. Therefore, this best pose for "76381" ligand to CXCR4 was subjected to MD simulation to assess the stability of the complex.

The CXCR4-76381 complex was subjected to 100 ns simulation and the RMSD was calculated. RMSD plot indicates that the ligand moves from its original binding position during simulations (**Fig 6a**). Major interacting residues of CXCR4 are Tyr256 and Gln200 (**Fig 6b**). Most of the interactions involve H-bonding and atom wise interactions of 76381 are shown in **Fig 6c**. All contacts with the CXCR4 are lost at around 40 ns, but regains subsequently, as indicated by total contact plot (**Fig 6d**). H-bond analysis between 76381 and CXCR4 (including H-bond with water) was also performed and it was found that at multiple instants during simulation the H-bond contact was lost and re-formed (**Fig 7)**. However, at no point during

simulation, there was any overlap between the native ligand binding mode (ITD) and 76381 binding mode (**Fig 8; Supplementary Video file**) suggesting that these are true allosteric binding sites.

Discussion

G-protein coupled receptors are important drug targets for several approved human drugs. Since all the GPCRs have similar seven-transmembrane structure, it becomes challenging to specifically design drugs for target GPCRs. Thus, designing specific allosteric modulators becomes more challenging and useful [13]. In this study, we first understand how well different protocols (Levels 1-3) play a role in modeling GPCR-ligand interaction. We find that accurate guidance of docking (smaller grid) is a crucial parameter and any biochemical knowledge on ligand binding sites (as in Level 1 and 2) greatly improves the chances of reproducing native ligand pose using computational docking methods. We selected structural entries of GPCRs with alternate binding sites and GLASS, a specialized database of ligands known to bind GPCRs for an in-depth study. We performed docking to check the binding of ligands from GLASS database with GPCRs. We were able to obtain ligands that are reported as targets for particular GPCRs that bind better to the allosteric site as compared to the endogenous binding site. We conclude that few of the known GPCR-specific ligands may bind to the allosteric site and alter the function of the GPCRs. The residues at the allosteric site are not as conserved as in the cognate binding site, further suggesting their likely specificity to a particular GPCR. This computational approach for ligand binding can be used for prediction of allosteric binders for GPCRs in general.

Abbreviations

5-HT2B: 5-Hydroxytryptamine receptor 2B

CXCR4: CXC-chemokine receptor 4

FDA: Food and Drug Administration

GLASS: GPCR-Ligand Association

GPCR: G-protein coupled receptor

OLBS: Original ligand binding site

RMSD: Root-Mean-Square Deviation

TC: Tanimoto co-efficient

Declarations

Conflict of Interest

The authors disclose no conflict of interests.

Funding

RS would like to acknowledge her JC Bose Fellowship (JC Bose fellowship (SB/S2/JC-071/2015) from Science and Engineering Research Board and NCBS (TIFR) for infrastructural facilities.

Author Contributions

KH, SJ and VT carried out all the work and analyses. RS conceptualized the work and both RS and SV participated in discussions. KH and VT wrote first draft of the manuscript and SJ, SV and RS improved it.

Acknowledgments

The authors would like to acknowledge Prof. N. Srinivasan of Indian Institute of Science for useful discussions.

References

1. Ahram M, Litou ZI, Fang R, Al-Tawallbeh G. Estimation of membrane proteins in the human proteome. In Silico Biol. 2006;6:379–86.
2. Almén MS, Nordström KJV, Fredriksson R, Schiöth HB. Mapping the human membrane proteome: A majority of the human membrane proteins can be classified according to function and evolutionary origin. BMC Biol. 2009;7:50.
3. Rosenbaum DM, Rasmussen SGF, Kobilka BK. The structure and function of G-protein-coupled receptors. Nature. 2009.
4. Nagarathnam B, Kannan S, Dharnidharka V, Balakrishnan V, Archunan G, Sowdhamini R. Insights from the analysis of conserved motifs and permitted amino acid exchanges in the human, the fly and the worm GPCR clusters. Bioinformation. 2011;7:15–20.
5. Venkatakrishnan AJ, Flock T, Prado DE, Oates ME, Gough J, Madan Babu M. Structured and disordered facets of the GPCR fold. Curr Opin Struct Biol. 2014;27:129–37.
6. Fredriksson R, Lagerström MC, Lundin LG, Schiöth HB. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol. 2003;
7. Heng BC, Aubel D, Fussenegger M. An overview of the diverse roles of G-protein coupled receptors (GPCRs) in the pathophysiology of various human diseases. Biotechnol. Adv. 2013.
8. Stoy H, Gurevich V V. How genetic errors in GPCRs affect their function: Possible therapeutic strategies. Genes Dis. 2015.
9. Zhang D, Zhao Q, Wu B. Structural studies of G protein-coupled receptors. Mol Cells. 2015;
10. Schwartz TW, Frimurer TM, Holst B, Rosenkilde MM, Elling CE. Molecular mechanism of 7TM receptor activation - A global toggle switch model. Annu. Rev. Pharmacol. Toxicol. 2006.

11. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: A major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46:D1074–82.
12. Sriram K, Insel PA. G protein-coupled receptors as targets for approved drugs: How many targets and how many drugs? Mol Pharmacol. 2018.
13. Jeffrey Conn P, Christopoulos A, Lindsley CW. Allosteric modulators of GPCRs: A novel approach for the treatment of CNS disorders. Nat Rev Drug Discov. 2009;
14. Lindsley CW, Emmitte KA, Hopkins CR, Bridges TM, Gregory KJ, Niswender CM, et al. Practical Strategies and Concepts in GPCR Allosteric Modulator Discovery: Recent Advances with Metabotropic Glutamate Receptors. Chem. Rev. 2016.
15. Azam S, Haque ME, Jakaria M, Jo S-H, Kim I-S, Choi D-K. G-Protein-Coupled Receptors in CNS: A Potential Therapeutic Target for Intervention in Neurodegenerative Disorders and Associated Cognitive Deficits. Cells. 2020;
16. Garrett M. Morris, Ruth Huey, William Lindstrom, Michel F. Sanner, RIchard K. Belew, David S Goodsell, et al. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. J Comput Chem. 2009;
17. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nat. Struct. Biol. 2003. p. 10, 980.
18. Pándy-Szekeres G, Munk C, Tsonkov TM, Mordalski S, Harpsøe K, Hauser AS, et al. GPCRdb in 2018: Adding GPCR structure models and ligands. Nucleic Acids Res. 2018;
19. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: Improved access to chemical data. Nucleic Acids Res. 2019;
20. Chan WKB, Zhang H, Yang J, Brender JR, Hur J, Ozgur A, et al. GLASS: A comprehensive database for experimentally validated GPCR-ligand associations. Bioinformatics. 2015;
21. Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. Proc 2006 ACM/IEEE Conf Supercomput SC'06. 2006.
22. Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. 2013;27:221–34.
23. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: Resources for positioning of proteins in membranes. Nucleic Acids Res. 2012;

Figures

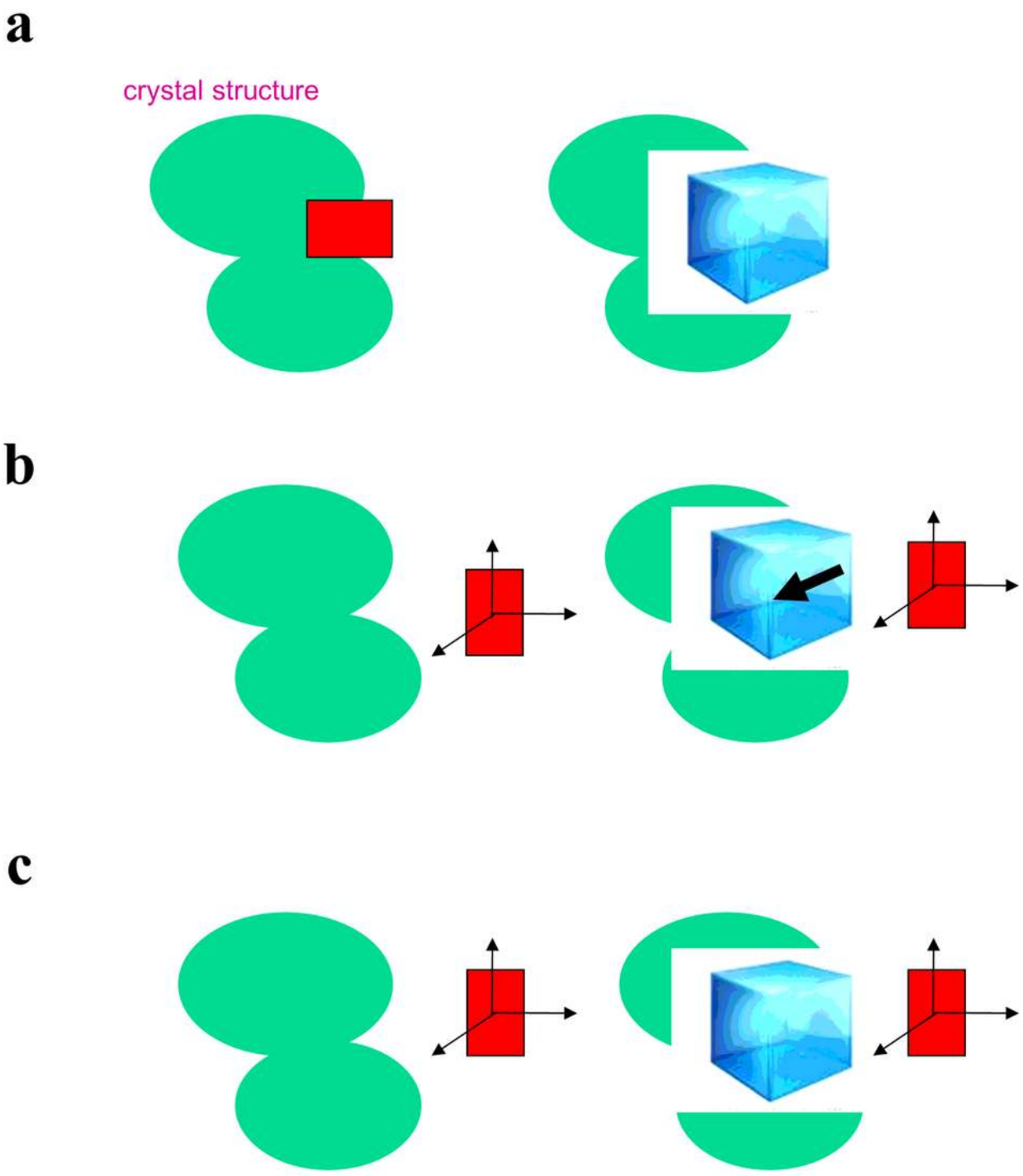


Figure 1

Different grids used for docking of ligands to GPCRs. For Level 1, extracellular half of GPCRs was considered for grid box generation and ligand coordinates were taken from PDB. For Level 2, smaller grid box covering only the known ligand binding site was defined and ligand coordinates were retrieved from PubChem. For Level 3, the ligands coordinates were retrieved from PubChem while the grid box was defined by considering the extracellular half region of GPCRs.

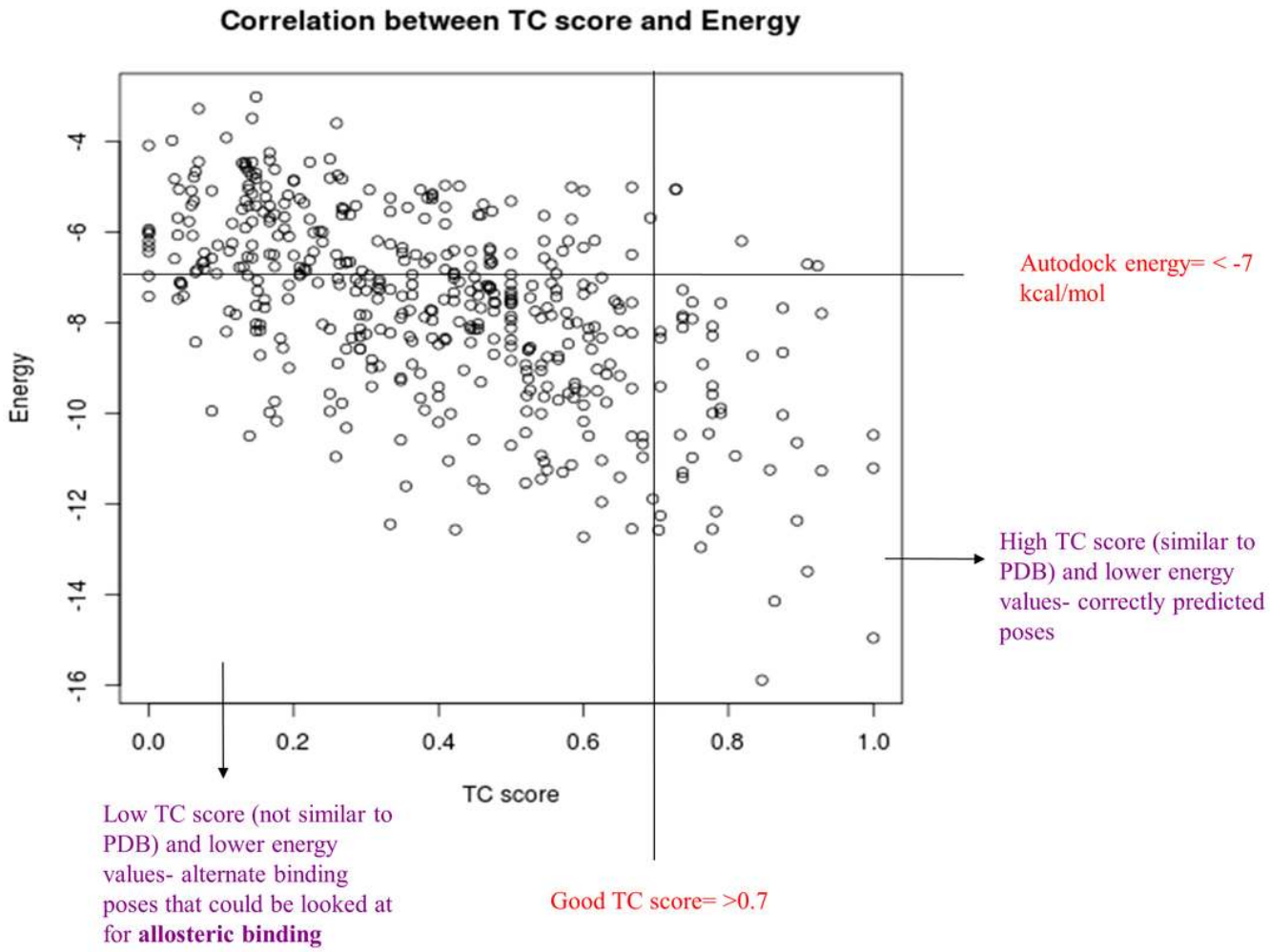


Figure 2

TC score vs Docking energy score plot. High TC score indicates a high similarity between docked ligand pose and crystal pose. Lower Docking score indicates the stable binding of ligand to GPCR

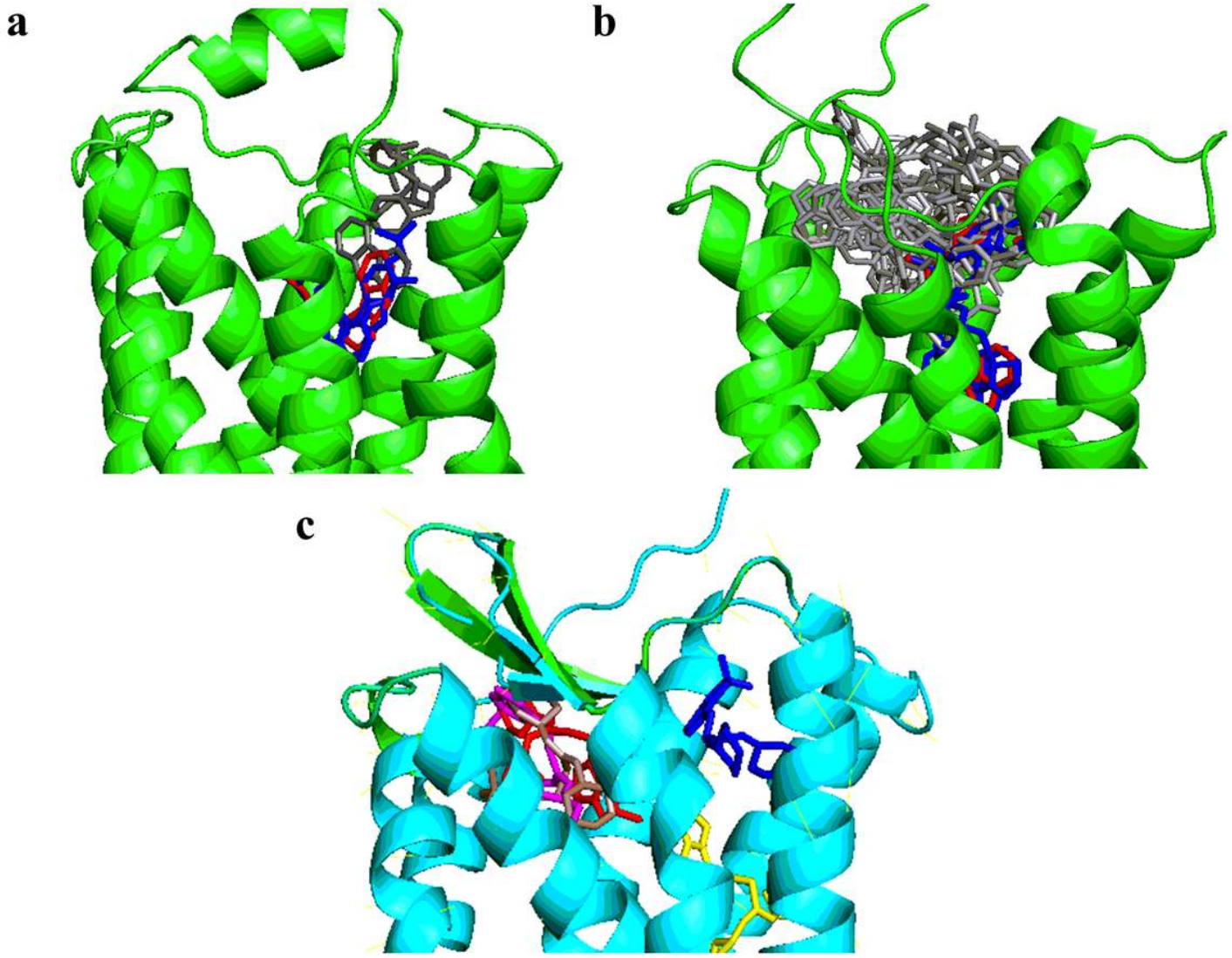
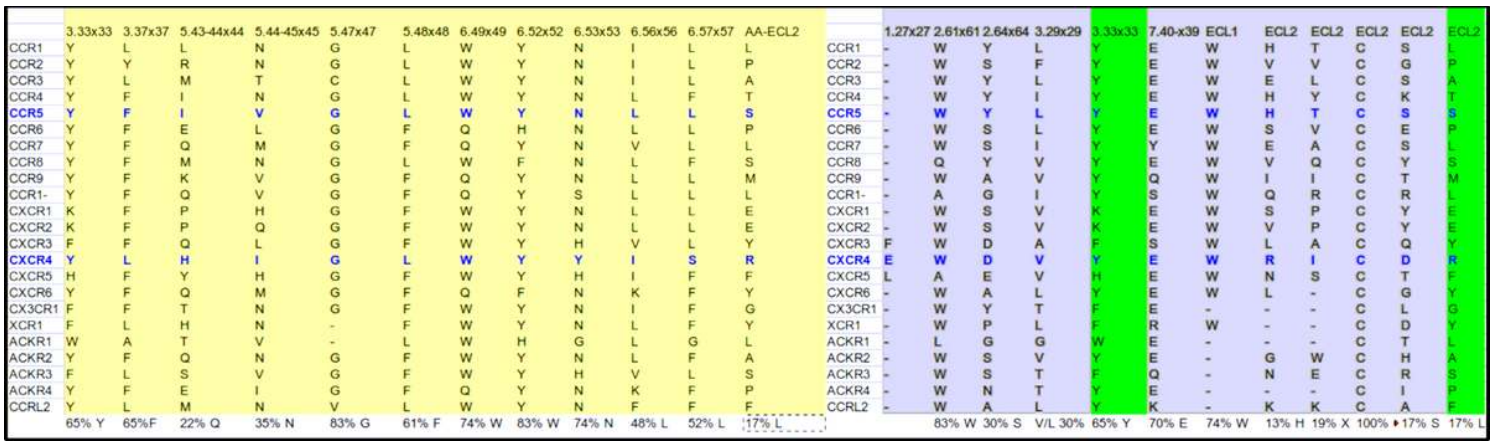


Figure 3

Novel ligand binding poses for high scoring complexes. (A) Ligand poses for beta-2 adrenergic receptor indicating the overlap between novel docked ligand pose (grey) with native ligand binding pose (red) and best scoring ligand pose is shown in blue(B) Ligand poses for 5HT2b receptor. The novel binding sites (grey) are close to the PDB-complex ligand binding site (red) and the best scoring pose (blue). (C) Ligand poses for Chemokine receptor CXCR4. Different structures of CXCR4 are bound to different ligands in PDB. Native ligand binding pose (red) and best scoring ligand pose is shown in blue. Other binding modes observed in PDB are represented by yellow, grey and magenta color ligand poses.



	3.33x33	3.37x37	5.43-44x44	5.44-45x45	5.47x47	5.48x48	6.49x49	6.52x52	6.53x53	6.56x56	6.57x57	AA-ECL2
CCR1	Y	L	L	N	G	L	W	Y	N	I	L	L
CCR2	Y	Y	R	N	G	L	W	Y	N	I	L	P
CCR3	Y	L	M	T	C	L	W	Y	N	I	L	A
CCR4	Y	F	I	N	G	L	W	Y	N	L	F	T
CCR5	**Y**	**F**	**I**	**V**	**G**	**L**	**W**	**Y**	**N**	**L**	**L**	**S**
CCR6	Y	F	E	L	G	F	Q	H	N	L	L	P
CCR7	Y	F	Q	M	G	F	Q	Y	N	V	L	L
CCR8	Y	F	M	N	G	L	W	F	N	L	F	S
CCR9	Y	F	K	V	G	F	Q	Y	N	L	L	M
CCR1-	Y	F	Q	V	G	F	Q	Y	S	L	L	L
CXCR1	K	F	P	H	G	F	W	Y	N	L	L	E
CXCR2	K	F	P	Q	G	F	W	Y	N	L	L	E
CXCR3	F	F	Q	L	G	F	W	Y	H	V	L	Y
CXCR4	**Y**	**L**	**H**	**I**	**G**	**L**	**W**	**Y**	**Y**	**I**	**S**	**R**
CXCR5	H	F	Y	H	G	F	W	Y	H	I	F	F
CXCR6	Y	F	Q	M	G	F	Q	F	N	K	F	Y
CX3CR1	F	F	T	N	G	F	W	Y	N	I	F	G
XCR1	F	L	H	N	-	F	W	Y	N	L	F	Y
ACKR1	W	A	T	V	-	L	W	H	G	L	G	L
ACKR2	Y	F	Q	N	G	F	W	Y	N	L	F	A
ACKR3	F	L	S	V	G	F	W	Y	H	V	L	S
ACKR4	Y	F	E	I	G	F	Q	Y	N	K	F	P
CCRL2	Y	L	M	N	V	L	W	Y	N	F	F	F
	65% Y	65%F	22% Q	35% N	83% G	61% F	74% W	83% W	74% N	48% L	52% L	17% L

	1.27x27	2.61x61	2.64x64	3.29x29	3.33x33	7.40-x39	ECL1	ECL2	ECL2	ECL2	ECL2	ECL2
CCR1	-	W	Y	L	Y	E	W	H	T	C	S	L
CCR2	-	W	S	F	Y	E	W	V	V	C	G	P
CCR3	-	W	Y	L	Y	E	W	E	L	C	S	A
CCR4	-	W	Y	I	Y	E	W	H	Y	C	K	I
CCR5	**-**	**W**	**Y**	**L**	**Y**	**E**	**W**	**H**	**T**	**C**	**S**	**S**
CCR6	-	W	S	L	Y	E	W	S	V	C	E	P
CCR7	-	W	S	I	Y	Y	W	E	A	C	S	L
CCR8	-	Q	Y	V	Y	E	W	V	Q	C	Y	S
CCR9	-	W	A	V	Y	Q	W	I	I	C	T	M
CCR1-	-	A	G	I	Y	S	W	Q	R	C	R	L
CXCR1	-	W	S	V	K	E	W	S	P	C	Y	E
CXCR2	-	W	S	V	K	E	W	V	P	C	Y	E
CXCR3	F	W	D	A	F	S	W	L	A	C	Q	Y
CXCR4	**E**	**W**	**D**	**V**	**Y**	**E**	**W**	**R**	**I**	**C**	**D**	**R**
CXCR5	L	A	E	V	H	E	W	N	S	C	T	F
CXCR6	-	W	A	L	Y	E	W	L	-	C	G	Y
CX3CR1	-	W	Y	T	F	E	-	-	-	C	L	G
XCR1	-	W	P	L	F	R	W	-	-	C	D	Y
ACKR1	-	L	G	G	W	E	-	-	-	C	T	L
ACKR2	-	W	S	V	Y	E	-	G	W	C	H	A
ACKR3	-	W	S	T	F	Q	-	N	E	C	R	S
ACKR4	-	W	N	T	Y	E	-	-	-	C	I	P
CCRL2	-	W	A	L	Y	K	-	K	K	C	A	F
		83% W	30% S	V/L 30%	65% Y	70% E	74% W	13% H	19% X	100%	17% S	17% L

Figure 4

Sequence alignment of chemokine receptor subfamily. Yellow indicates the residues involve in binding with novel ligand pose. Grey residues interact with native ligand pose. Green indicates the common residues that are involved in interaction with novel pose as well as native pose.

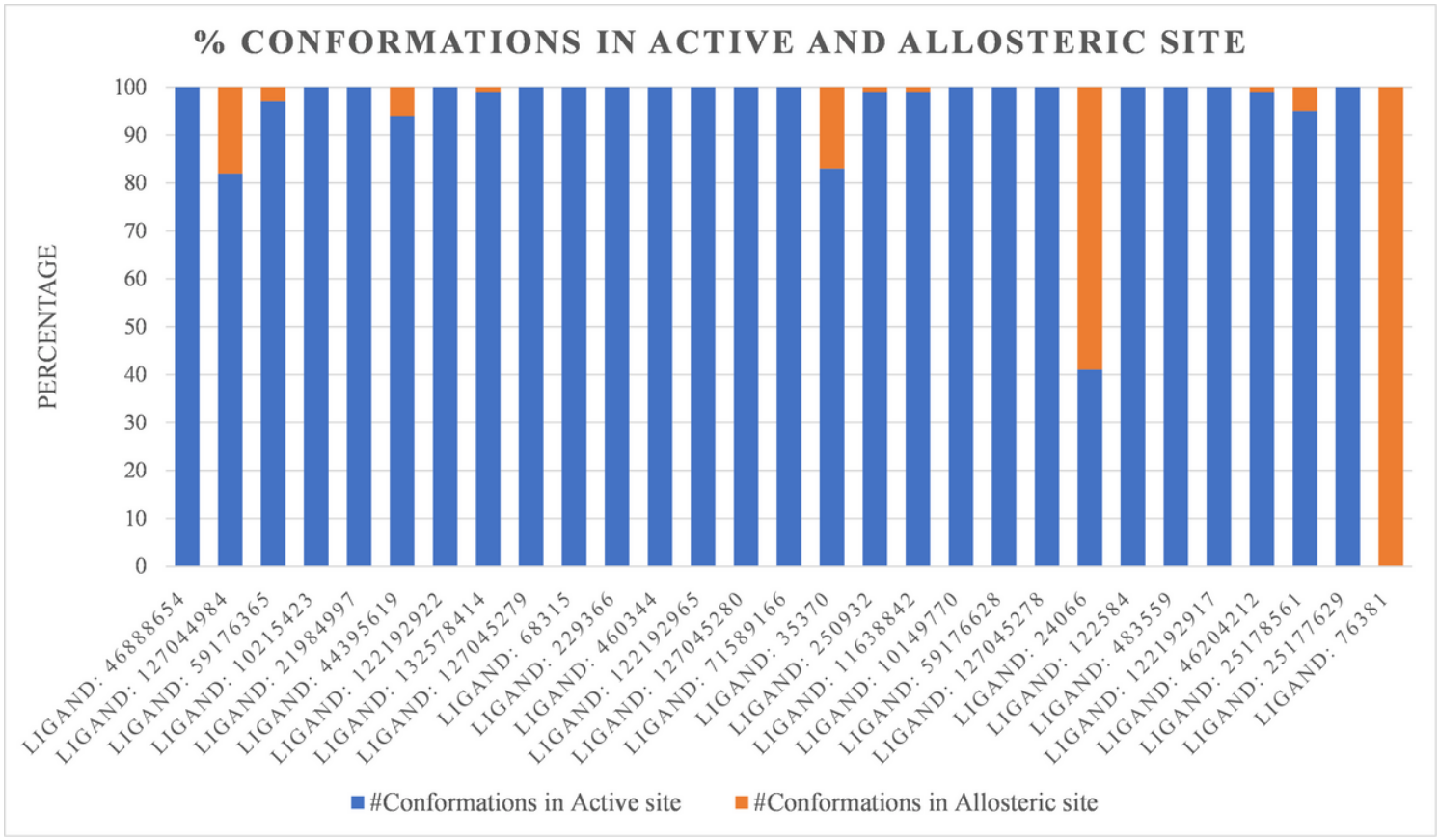


Figure 5

Bar graph indicating the percentage of each ligand conformation in either native site or allosteric site.

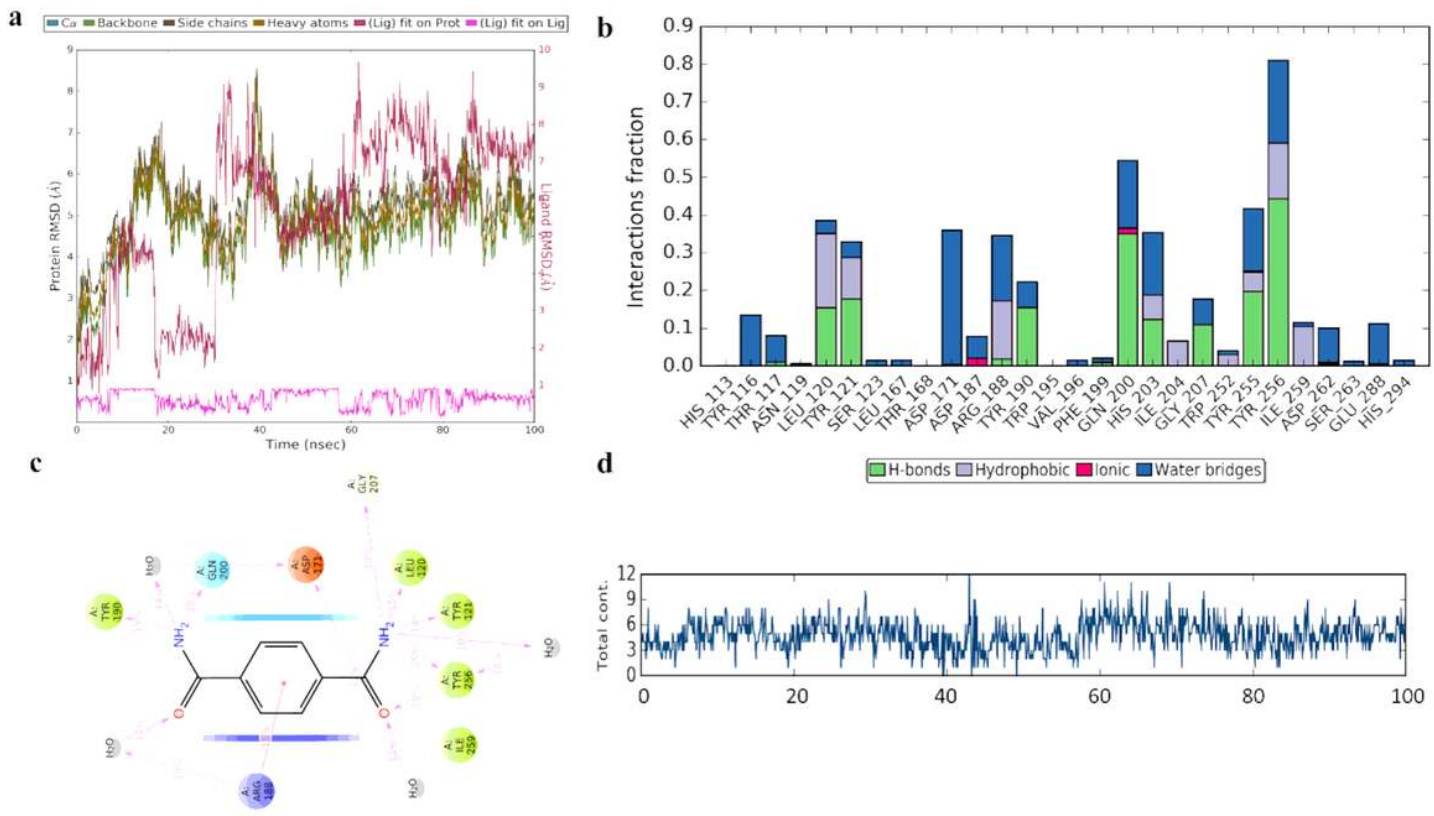


Figure 6

MD simulation results of CXCR4-76381 complex. (A) RMSD plot of CXCR4 and 76381. (B) Interacting residues and Interaction types of CXCR4 with 76381 over the course of simulation time. Normalized stacked bars indicate the fraction of simulation time for which a particular type of interaction was maintained. Values more than 1.0 suggests that the residue forms multiple interactions of same subtype with ligand (C) Interactions of 76381 atoms with residues of CXCR4 along with type and duration of interactions. Interactions that persist for more than 10% of simulation time has been shown. If some residues form multiple interaction of same type with the same atom of ligand then interaction value can be more than 100% (D) Total number of contacts (H-bonds, Water bridges, Hydrophobic, Ionic) between CXCR4 and ligand 76381 throughout the simulation.

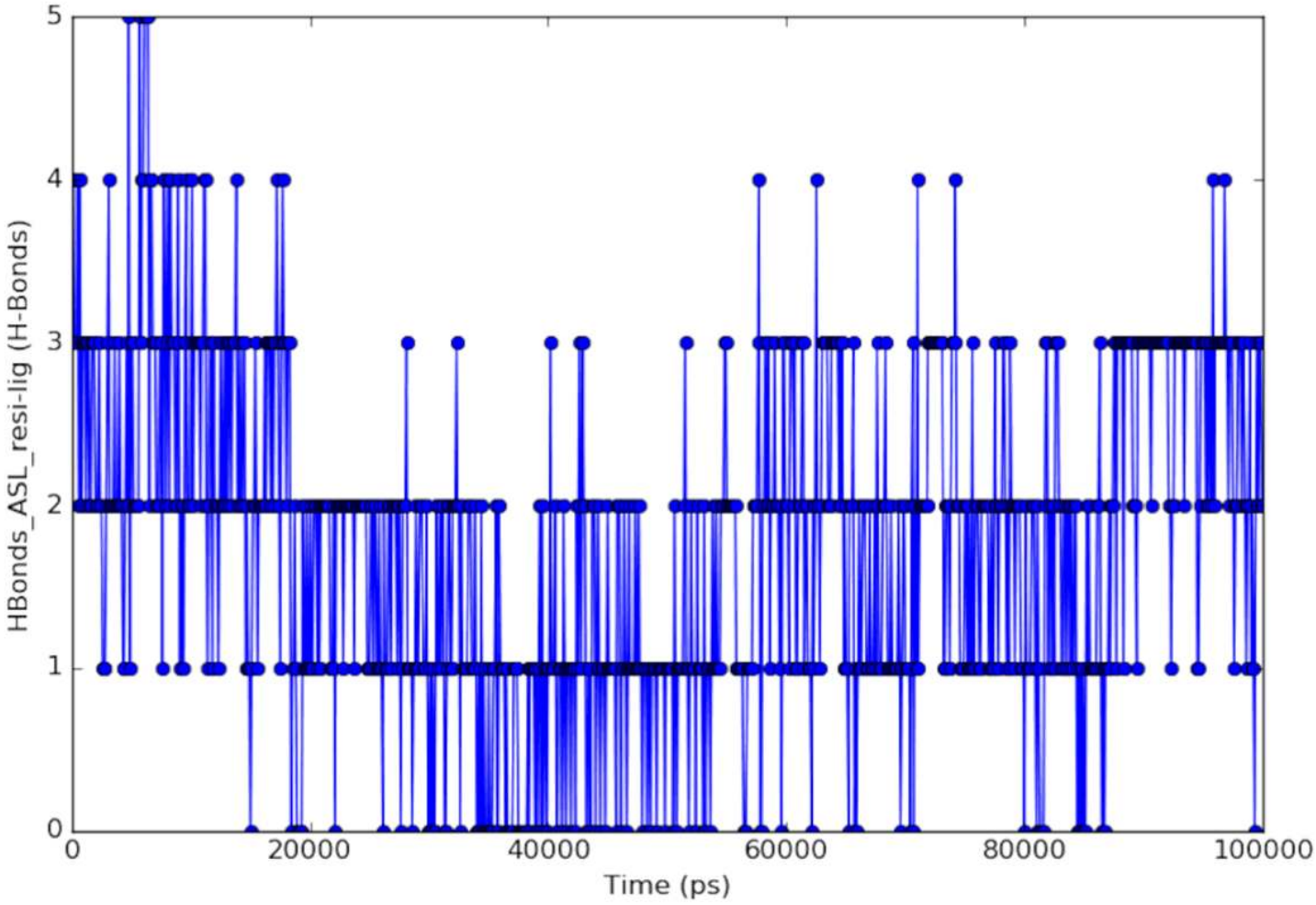


Figure 7

H-bond interactions between 76381 ligand and CXCR4 (and water) throughout the simulation period. Simulation time is in picoseconds.

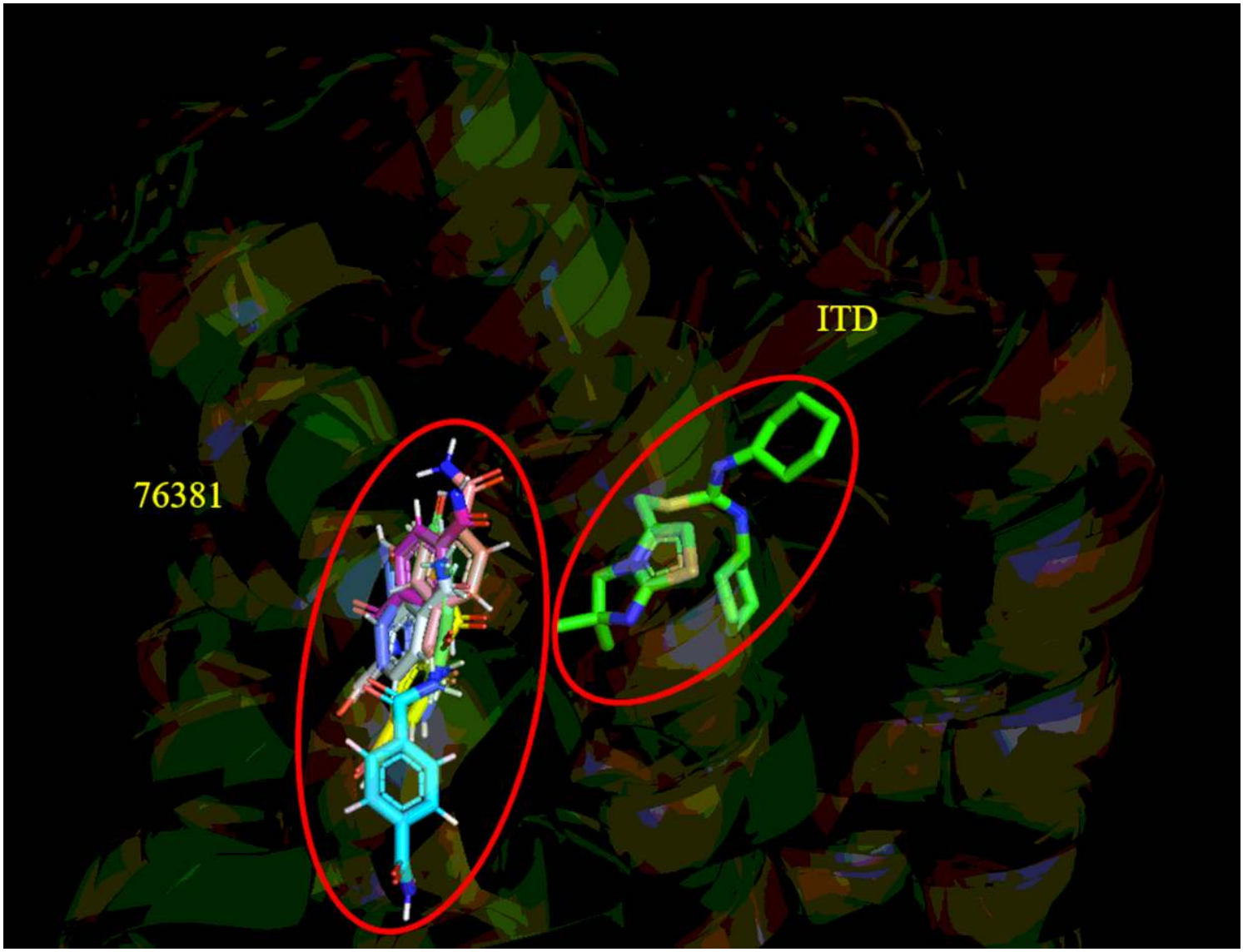


Figure 8

76381 poses with CXCR4 compared to native ligand (ITD) of CXCR4 at different time points (in nanoseconds) during simulation. Native ligand pose (ITD) is shown as green-blue sticks. Cyan: 0ns; Magenta: 33ns; Yellow: 55ns; Pink: 61ns; White:77ns; Violet: 85ns; Orange: 90ns; Green: 100ns