Supporting Information

Discovery of GlyT2 Inhibitors using structure-based pharmacophore screening and FEP+ Calculations

Filip Fratev[1, 2], Manuel Miranda-Arango[3], Elvia Padilla[3] and Suman Sirimulla[1]

[1]Department of Pharmaceutical Sciences, School of Pharmacy; University of Texas at El Paso, El Paso, Tx, USA
[2]Micar21 Ltd., Persenk 34B, 1407, Sofia, Bulgaria
[3]Department of Biological Sciences, The University of Texas at El Paso, El Paso, Tx, USA

May 20, 2019

Contents

1 Methods **2**
- 1.1 Homology modeling . . . 2
- 1.2 Molecular dynamics . . . 2
- 1.3 Virtual screening approach . . . 3
- 1.4 Pharmacophore, regular and induced fit (IFD) docking . . . 4
- 1.5 Metadynamics refinements of IFD solutions . . . 4
- 1.6 Free energy perturbation FEP+ calculations . . . 5
- 1.7 ADMET . . . 5
- 1.8 Development of an in-vitro cell system for screening of small molecules . . . 6

2 Figures **8**

References **24**

1 Methods

1.1 Homology modeling

A detail analysis of all possible homologous structures to GlyT2 was initially performed. For our homology modeling we used both the Prime package of Schrodinger software and I-TASSER server. Recently obtained high resolution X-ray structures of *Drosophila* dopamine (pdb id: 4xpt) and human serotonin (pdb id: 5i6x) transporters were employed as templates [1, 2]. They shared higher similarity compared to the older bacterial (*Aquifex aeolicus*) structure of leucine receptor (pdb id: 3f3a). The later showed only 22% identity to GlyT2 whereas the pdb id 4xpt and 5i6x structures have 42% and 38% similarity, respectively. For the aim of our study, the identity of amino acids in the glycine binding site was also very important. For instance, the serotonin transporter has a 68% identity with GlyT2, which was much better than that of the bacterial leucine transporter (41%). In summary, several of the previous models of glycine transporters have been based on the leucine receptor structure but they should be used with caution. For instance, pdb id 3f3a is populated with more than one ligand, which can produce significant changes in LBD, and also the similarity of helix 12 (H12) is less pronounced to GlyT2 than those of serotonin and dopamine transporters. A chimeric model was used in Prime. After comparison of the Ramachandran plots and visual inspection the I-tasser model was used as an input for our further MD simulations.

1.2 Molecular dynamics

All molecular dynamics (MD) simulations were carried out using the Amber 16 program [3]. CHARMM36 force field (FF) was used for the calculations of the apo transporter form and those bound with glycine (holo form). The GlyT2 was enabled embedded in a POPC membrane by the CHARMM-GUI membrane builder webserver (charmm-gui.org) [4]. Membrane orientation was obtained by OPM server [5] and was essentially same as for serotonin, which was used also for a control run, and dopamine transporters. Default parameters were employed for both heterogeneous lipids creation. Water thickness of 17.5Å and 0.15M NaCl ion concentration were used. Simulations with NaK were also provided in order to examine potential difference in Na^+ and K^+ at the 3^{rd} sodium binding site. These was mixed model with Na and K ions keeping the system neutral. The number of these ions was automatically chosen by CHARMM-GUI membrane builder server. The system size was X=99 Å, Y=99 Å and Z=152 Å and it contained about 98 000 atoms for both the apo and holo forms. The glycine parameters as calculated by CHARMM General Force Field (CGenFF) were employed. We used the default relaxation times, as generated for Amber software, by CHARMM-GUI but used friction coefficient of 2 ps-1 in order to allow higher temperature (T=303.03K) fluctuations. The non-bonded cutoff was set to 12 Å whereas the force-based switching was 10 Å. For these calculations the MC Barostat was used for the pressure control. After 8000 steps minimization (2500 with the conjugate gradient method), five equilibration steps were executed and five independent 500-ns-long production runs were performed for each system. Finally, based on all of these 2.5μs MDs we retrieved both average and cluster (center of the most populated cluster) structures for GlyT2 transporter, which were used for our further analysis, VS and FEP+ calculations. However, for identified hit

compounds we executed simulations with Amber14SB, along with Lipid 14, and GAFF2 force fields [3]. The reasons were: 1) comparison of the structures produced by different FFs, which would provide more confidence about the quality of obtained structures and 2) GAFF2 seems to perform better during ligand-protein MD simulations according to our experience. Another difference is that for these calculations we used Berendsen instead of MC Barostat because of our recent results [6] showing that the MC Barostat should be used with caution for ligand sampling in similar systems. For these MD runs, the systems were energy-minimized in two steps. First, only the water molecules and ions were minimized in 6000 steps while keeping the protein and ligand structures restricted by weak harmonic constrains of 2 kcal mol–1 Å–2. Second, a 6000 steps minimization with the conjugate gradient method on the whole system was performed. Furthermore, the simulated systems were gradually heated from 0 to 310 K for 50 ps (NVT ensemble) and equilibrated for 3 ns (NPT ensemble). The production runs were performed at 310 K in a NPT ensemble. Temperature regulation was done by using a Langevin thermostat with a collision frequency of 2 ps–1, and the pressure regulation via Berendsen barostat. The time step of the simulations was 2 fs with a non-bonded cutoff of 8 Å using the SHAKE algorithm [7] and the particle-mesh Ewald method [8]. In both Charmm and Amber simulations we used TIP3P water model. Two independent 250-ns-long production simulations were executed for all of the hit compounds complexed with GlyT2.

1.3 Virtual screening approach

The structural and ligand-based *in silico* screens are of great help during drug discovery nowadays, and their combination has been shown to provide better results than using a single approach. However, many targets have not had any ligands discovered, thus making this combination impossible.

In the current study we tested whether the recently introduced structural based pharmacophore approach in Phase software (Schrödinger package) in combination with docking followed by rescoring with the new Glide-SP scoring function can achieve reasonable virtual screening (VS) results. These methods were also tested in our lab on the full (102 targets) DUD-E benchmark. Surprisingly, we found that the structure based pharmacophore hypothesis created by the docking of more than 600 small fragments (the default number in Phase) performs well for most of the targets with an average active ligand recovery rate of 10-15% during the first 10% database screen [9] (see Figures S15 and S16 for examples). The MD prepared structures further increased these values. These results are equal to the ligand based pharmacophore virtual screening. The docking technique performed much better on DUD-E set. Early enrichment performance shows on average of about 30% of known actives are recovered in screening the top-ranked 1% of recovered decoys, similarly to the previous DUD-E test results. Further, an improvement of the Glide-SP scoring function, as it is also claimed on Schrödinger website, was confirmed as it is shown on the ACE test target, which usually produced unsatisfied results previously (Figure S17). However, the main aim of a real drug discovery project is the very early enrichment performance, i.e. how many active compounds will be discovered between the first 10 to 50 ligands, which we call here also success rate (SR) or efficiency (EF). According to our data the combination of these methods achieves an impressive result of nearly 75% success rate for all 102 targets, even reaching

often 100%, for the top 10 ligands [9]. Based on our data we developed a new VS protocol using the top 1% of the pharmacophore search (3.5 million "lead-like" ZINC compounds in our case, which were available to download during the execution of the study) as an input to the docking and scoring by Glide-SP (35,000 ligands). All calculations were performed by the Schrodinger's Phase and Glide software packages as implemented in Schrodinger 2017-3 suite [10]. We hope that our approach will be in help during the drug discovery projects and in particular in a case that no know ligands are available for the studied protein. The aforementioned short description provides only some brief introduction to our approach, and a detailed paper about the topic is underway. Figures S15, S16 and S17 are in support to the idea that the combination of structural pharmacophore and docking methods is a good strategy which should be further tested, which has been done also in our study. However, as this methodology and results were based on X-ray data [9], for us it was also interesting to find out to which extent it would be applicable for generated by homology modelling structures as in the case of GlyT2 transporter. As it is evident from the achieved in this study hit rate of 75 percent it seems that this approach can be successfully applied in a real hit discovery VS projects. For more description see also the results and discussion part of the paper.

1.4 Pharmacophore, regular and induced fit (IFD) docking

As it was described above we used the combination of structure-based pharmacophore and docking approaches as they are implemented in Schrodinger Suite 2017-3. For both of these methods the default settings were employed. Exceptions were that during pharmacophore search all points match (4 from 4 pharmacophores (Figure 4)) were requested whereas during the docking we kept 10 000, instead of 5000, initial docking poses and the rewards of intramolecular H-Bonds were employed as an option (Figure S18). The ligand-protein H-bond reward was included by default in the Glide scoring function. Thus, we focused on ligands which are more capable to make H-bonds with the key residues such as Tyr232 and Ser473. The induced Fit Docking (IFD) [11] were also performed using the default settings but either an increased accuracy to XP docking mode or enhanced sampling option were also used to test results stability. This was done in order to investigate if other docking poses can be obtained when using a different settings.

1.5 Metadynamics refinements of IFD solutions

Metadynamics refinements of the IFD output poses is an important step in order to find the accurate binding pose. Very often the regular IFD suggests variety of docking solutions which are similar to two major conformation of the ligand; in our case the aromatic ring conformation rotated by 180 degree. On the other hands the advanced MD sampling techniques, such as accelerated MD (aMD) and metadynamics, provide a good opportunity the most likely binding mode to be recovered [12, 6]. In current study we used recently suggested approach based on execution of several independent metadynamics simulations and then scoring the ligand docking poses by observed ligand-protein interactions during the simulations course [12]. This protocol is now implemented in Schrodinger suite and can be automatically performed. We used the default settings for these calculations. Essentially,

8 independent 10-ns-long metadynamics simulations were executed for selected two possible conformations and the pose scores were retrieved as previously suggested (Figure S6) [12]. For a final screening, we combine the RMSD stability and hydrogen bond persistence measures, obtained from the metadynamics trajectories into a scoring function whose parameters have been optimized using the data from the 42 test cases discussed in [12]. For instance, the stability of Hit 1 for the two likely poses were almost same (after 4ns the RMSD was below 0.2A) but the H-bonding, persistence was better for the pose one. We selected the better ranked pose from the metadynamics (represented in blue color in Figure S6) based on its lower score for further MD simulations In the same case the order of the rank was changed and the ranked by IFD top compound became second position after the metadynamics refinement.

1.6 Free energy perturbation FEP+ calculations

All calculations have been conducted using the Schrödinger molecular modelling suite 2017-3 [10]. Free energy perturbation calculations were performed using the FEP+ methodology, which combines the accurate modern OPLS3 force field [13], GPU-enabled high-speed molecular dynamics simulations with Desmond MD package [14] , the REST algorithm for locally enhanced sampling [15], a cycle-closure correction [16] to incorporate redundant information into free energy estimates. The FEP+ calculations were based on the aforementioned MD obtained structures and were conducted using the default protocols: The systems were solvated in an orthogonal box of SPC water molecules with buffer width (minimum distance between box edge and any solute atom) of 5 Å for the complex and 10 Å for the solvent simulations. The full systems were relaxed and equilibrated using the default Desmond relaxation protocol, consisting of an energy-minimization with restraints on the solute, then 12 ps length simulations at 10 K using an NVT ensemble followed by an NPT ensemble. A total of 12 λ windows were used for all calculations. Replica exchanges between neighbouring λ windows were attempted every 1.2 ps. Finally, for the actual FEP+ pre-REST (prior REST) and REST calculations we employed our new FEP+ sampling protocol which demonstrated superior results especially with MD derived structures [6, 17] . Thus, for these FEP+ simulations we employed 2x10 ns/λ pre-REST and 8 ns/λ REST calculations. We analyzed the convergence of the free energies (deltaG) during the course of our FEP+ calculations. All calculations were run on Nvidia Pascal architecture GPUs (cluster of 8xGTX 1080Ti GPUs).

1.7 ADMET

All ADMET calculations were performed by QikProp as implemented in Schrödinger suite 2017-3 [10]. QikProp predicts the widest variety of pharmaceutically relevant properties - octanol/water and water/gas log Ps, log S, log BB, overall CNS activity, Caco-2 and MDCK cell permeabilities, log Khsa for human serum albumin binding, and log IC_{50} for HERG K^+-channel blockage - so that decisions about a molecule's suitability can be made based on a thorough analysis. QikProp bases its predictions on the full 3D molecular structure; unlike fragment-based approaches, it can provide equally accurate results in predicting properties for molecules with novel scaffolds as for analogs of well-known drugs.

The ADMET profile of Hit1 is not optimal and should be optimized. It does not show either some activity pertinent to CYP, HERG K+, androgen and estrogen receptors, or PgP as predicted by our calculations, but features not good membrane permeability, orally absorption, and water solubility as well as a logBBB ratio of 3. However, after FEP+ guided lead optimization currently underway in our lab, this hit may be relatively easy transformed to a promising preclinical drug candidate. The computational ADMET profiles of the compounds were attached as separate excel sheet as part of the Supporting Information. We also speculate that due to the presence of the carbonyl group in Hit1 as a key functional group in the interactions with the first sodium ion and the binding as a whole, as in the case of ALX1393, this inhibitor will be reversible (not like Org-25543) and thus will not produce undesired toxic affects due to decreasing GlyT2 expression levels [18]. This is also supported by our MD calculations (Figure S4 and Movie S1), which clearly indicated that the first NA+ unbinding leads also to the Glycine unbinding due to the loss of Carbonyl-Sodium interactions. The calculated Hit2 ADMET profile was superior, indicating superior BBB penetration and no significant interactions with major proteins that can produce side effects. Additional selectivity screening against the remaining members of these transporters and an additional optimization will be necessary. We are continuing these efforts toward improved drug development for chronic pain management.

1.8 Development of an in-vitro cell system for screening of small molecules

GlyT2

To set up a cell assay for screening and testing of GlyT2 inhibitors, we cloned the human GlyT2 gene into the expression vector pcDNA3.1 and used it for transfection and selection of porcine endothelial aortic (PAE) stable cell lines. We selected PAE cells as our preferred model cells because the cells firmly attach to the plastic and do not overexpress ectopic proteins, Moreover, they are flattened which allow easy visualization of the plasma membrane by conventional fluorescence microscopy. After selection of stable cell lines, we analyzed the localizatiporters in PAE cells, we measured [3H] Glycine uptake over different concentrations of substrate (Figure 4B). The cell line exhibited specific transport defined as the difference between [3H] Glycine uptake from GlyT2 cells minus parental PAE. This kinetic analysis shows a Michaelis-Menten behavior with an affinity for glycine of 12.3 μM and a Vmax of 8.2 nmol/min/mg (Figure S7). These results together validate that this cell line has the correct localization and good quantities of functional GlyT2, making this cell line ideal for screening of a novel class of specific inhibitors.

GlyT1

Porcine Aortic Endothelial (PAE) cells stable cell lines expressing mouse GlyT1a were grown in 24-well plates with a high confluency of 90-100 percent. First, confluent cells were washed twice with 0.25mL of reaction buffer (10mM HEPES pH 7.4, 135mM NaCl, 2mM KCl, 1mM CaCl2, 1mM MgSO4 and 10mM glucose). After washes, we added 0.25mL of reaction buffer with 4 Ci of [3H] glycine per mL, and a final concentration of 500 M of cold glycine along with the inhibitor in different concentrations for 10 minutes at 37C. After 10 minutes, we terminated glycine uptake by washing the cells twice with 0.25mL of

cold reaction buffer without [3H] glycine and glucose. Lastly, we extracted [3H] glycine by incubating the cells with 0.25mL of 0.2N NaOH for one hour at room temperature. Then, we proceed to measure protein concentration using the Bradford method. Glycine uptake was measured using scintillation spectroscopy and glycine specific transport was calculated by subtracting the transport of wild type PAE cells.

2 Figures

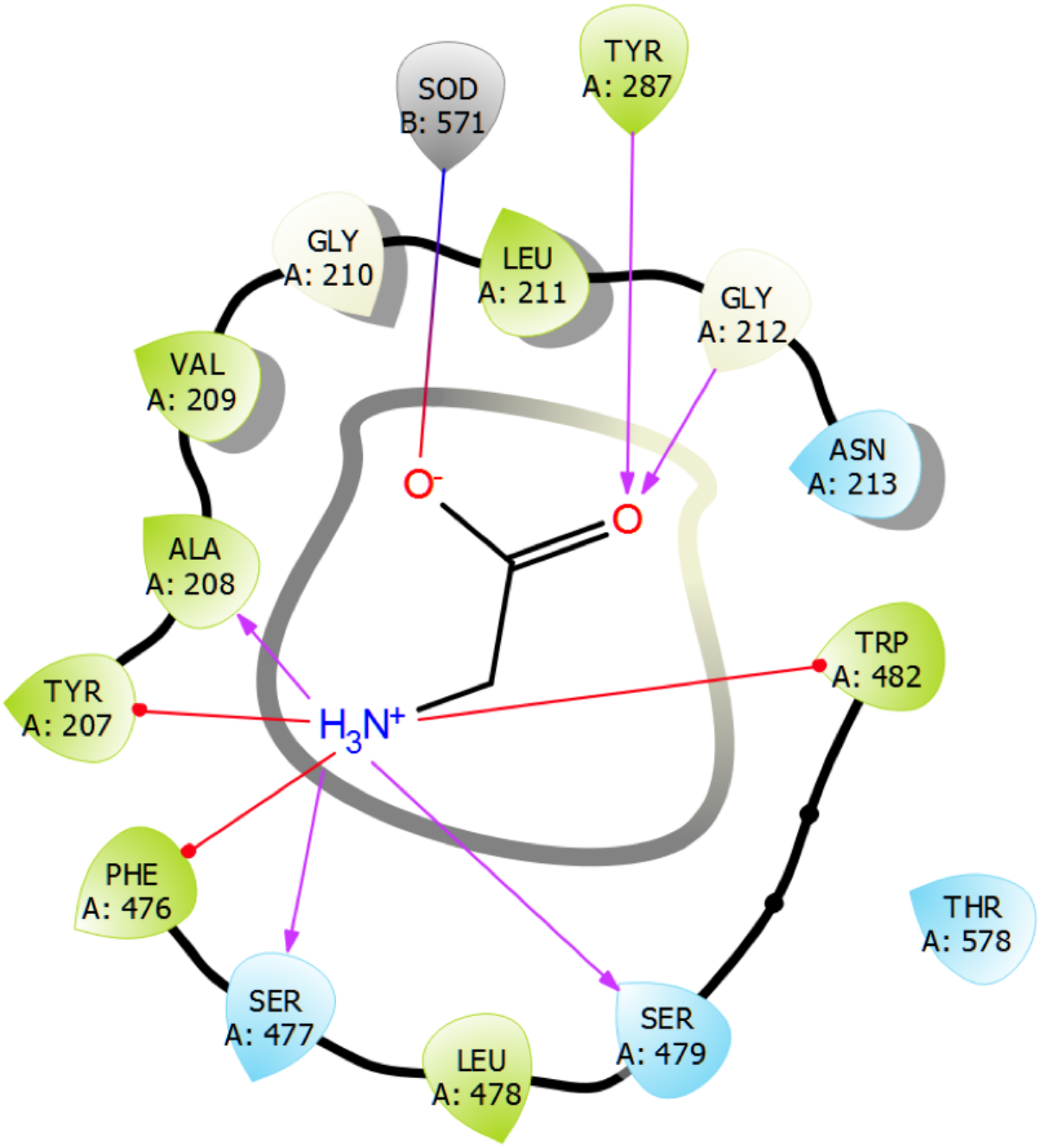


Figure S1: 2D representation of Glycine-GlyT2 interactions as identified by our MD studies.

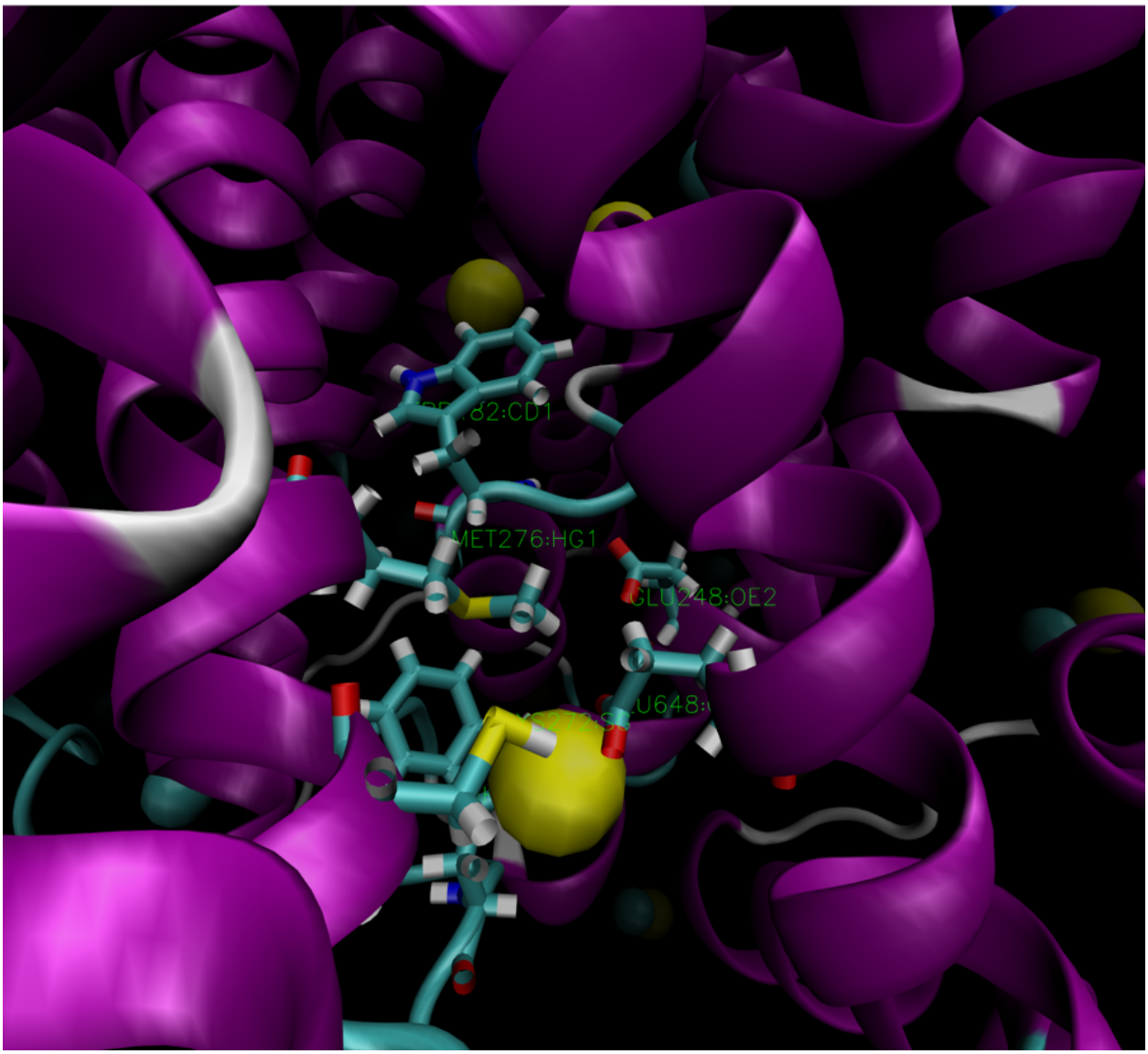
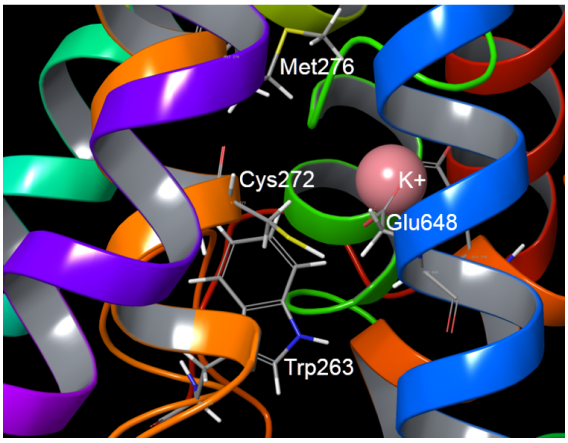
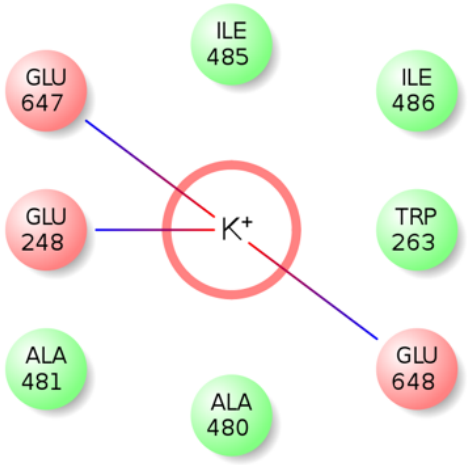


Figure S2: 3D representation of the 3^{rd} sodium binding site.



(a) 3D represntation



(b) 2D representation

Figure S3: (a) 3D and (b) 2D representation of the K^+ ion position in the 3^{rd} sodium binding site.

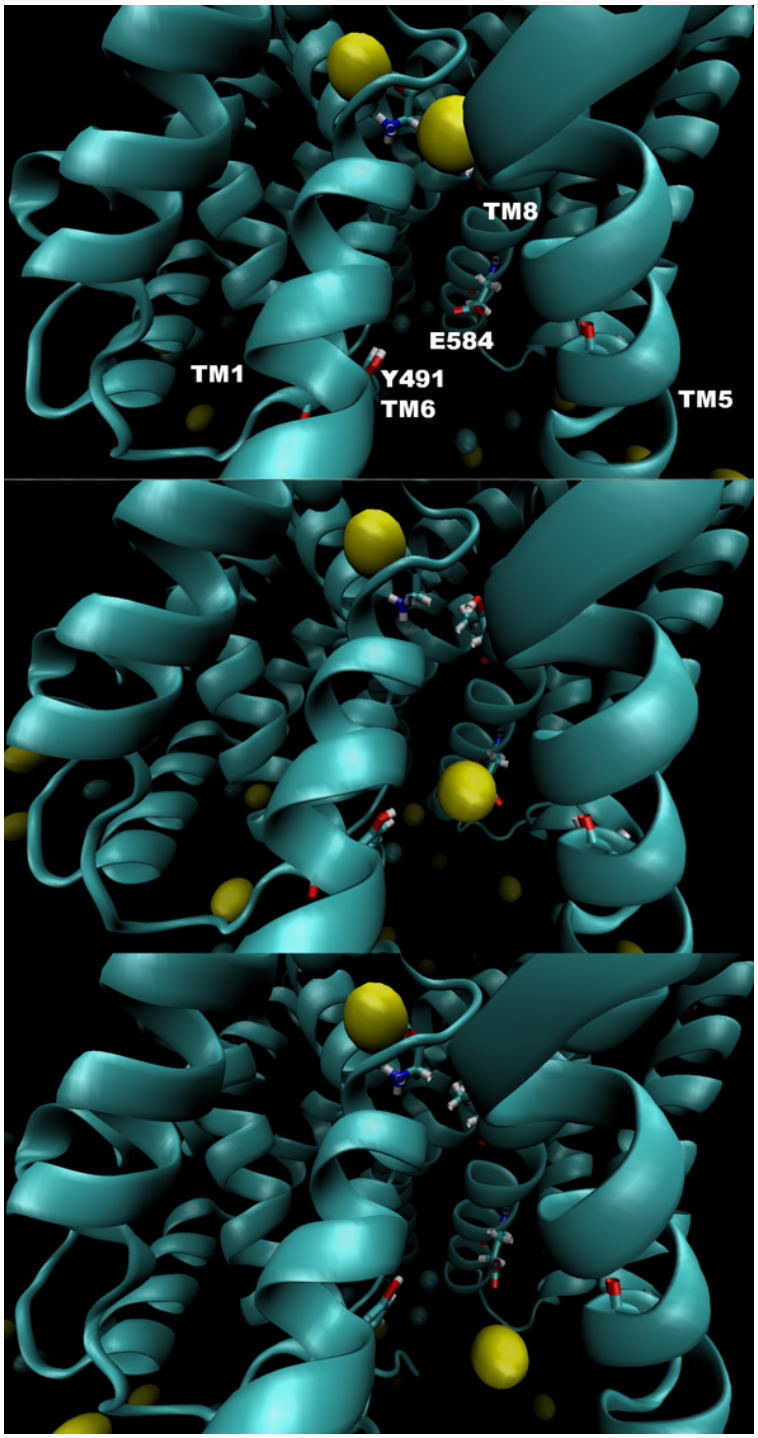


Figure S4: Identified unbinding path of the second Na^+ ion in GlyT1.

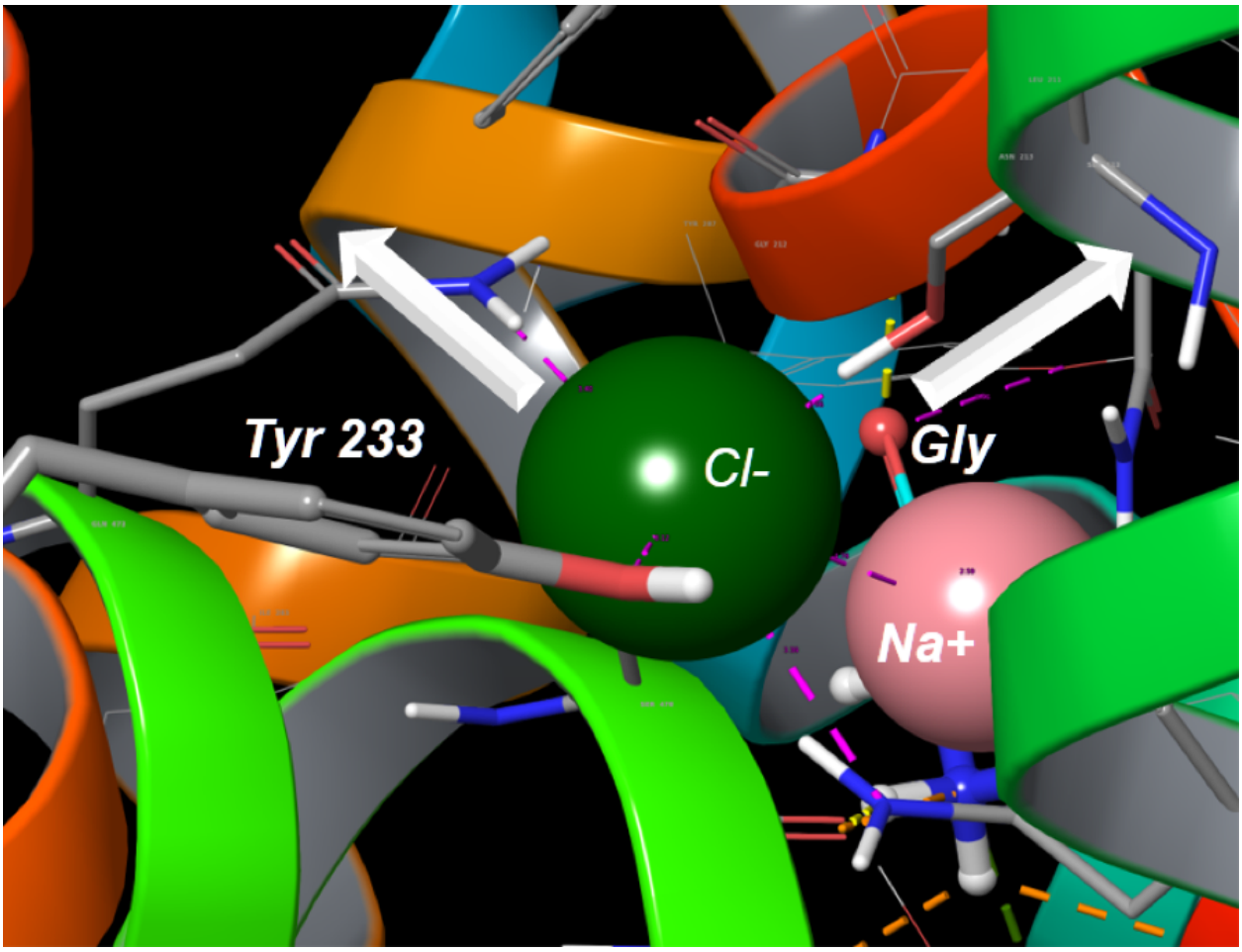


Figure S5: Identified unbinding paths of the second Cl^- ion in GlyT2. With white arrows are shown the unbinding paths of Cl^- (left) and neighboring first sodium ion.

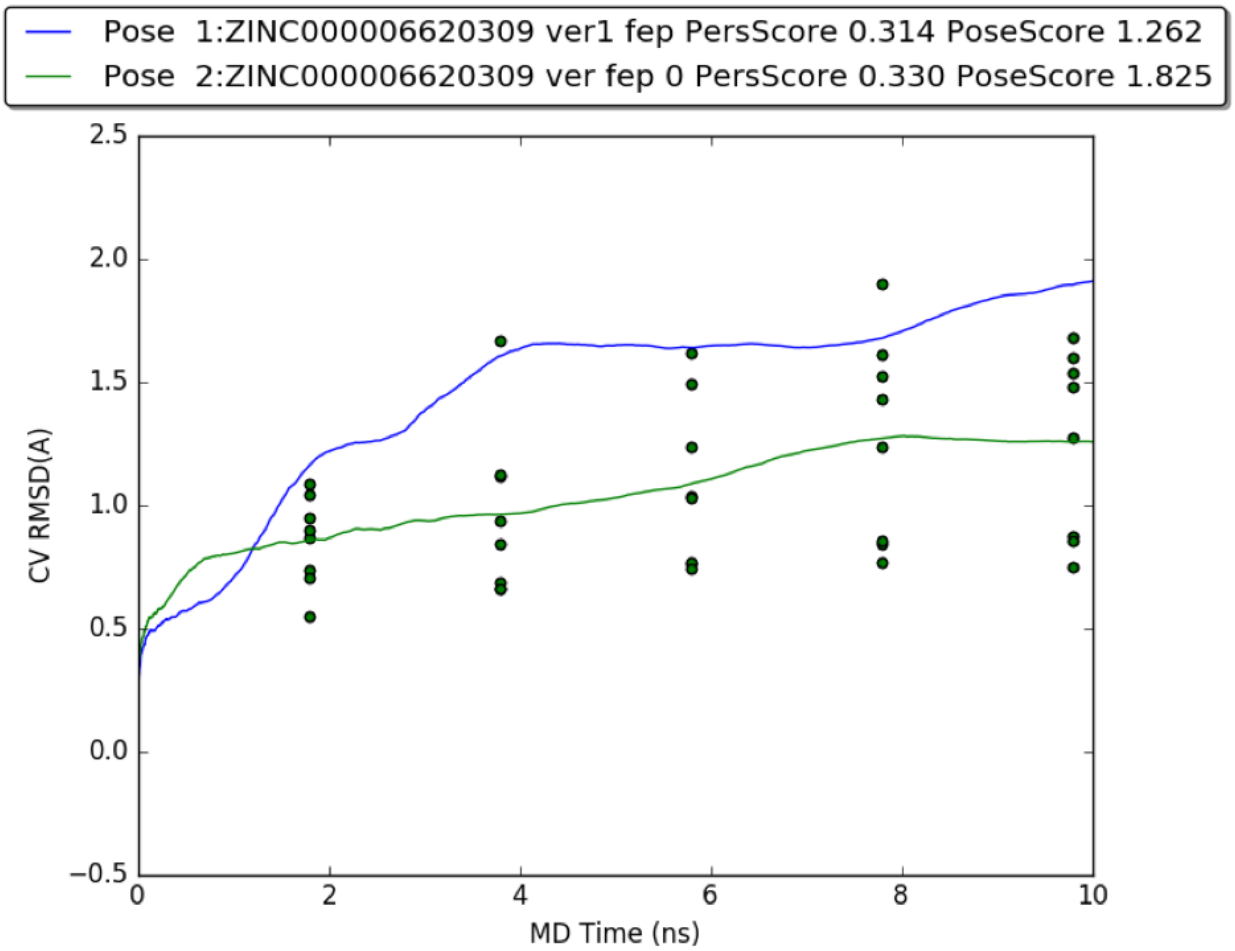


Figure S6: Metadynamics IFD pose prediction results for Lead1. Two docking possess, with rotated by 180 degree aromatic ring, were submit for these calculations and most likely pose was revealed by these runs.

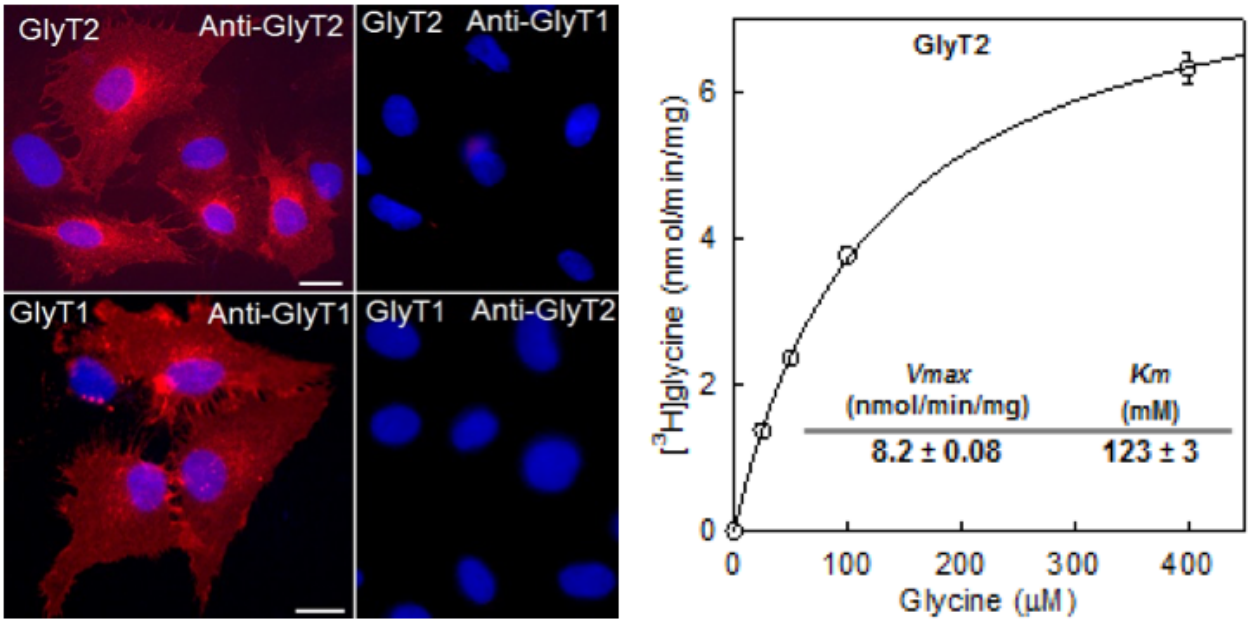
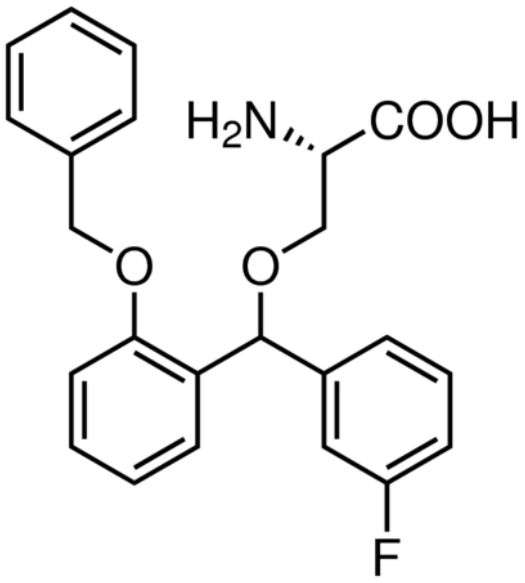
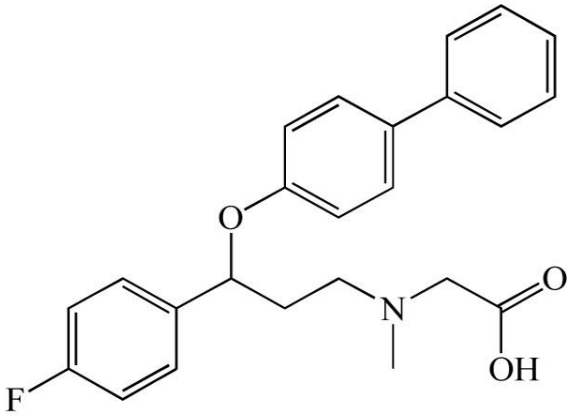


Figure S7: Stable expression of the human GlyT2 and GlyT1b in PAE cells. (A) Cells were fixed with 4% paraformaldehyde followed by staining with anti-GlyT1 or anti-GlyT2 antibodies and CY3-conjugated secondary antibodies. (B) The GlyT2 cells were subjected to uptake assay and the kinetic constants calculated with Sigma Plot 12 software.



(a) ALX1393



(b) ALX5407

Figure S8: Structural representation of (a) ALX1393 and (b) ALX5407. Note that these compounds and discovered herein hits have a COOH group, which interacts with the first NA+ ion, wheres in the irreversible Org-25543 inhibitor it is not present.

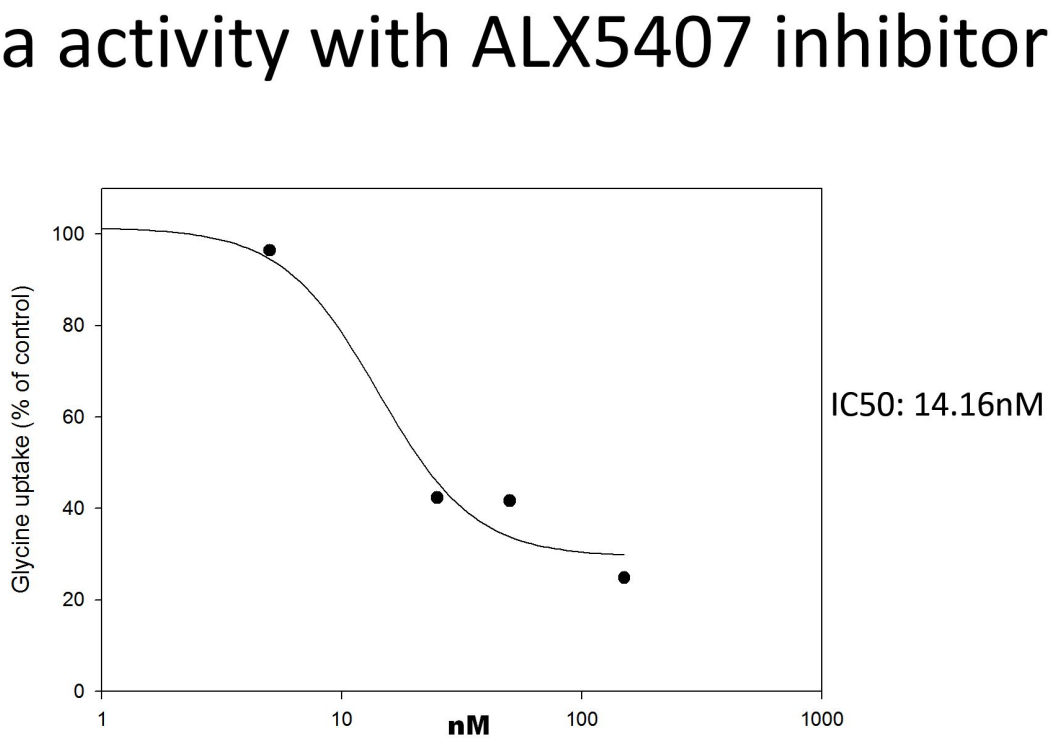


Figure S9: Inhibition curves of GlyT1 activity by ALX5407 inhibitor.We obtained sigmoidal curves and calculated IC50 values via Sigma Plot.

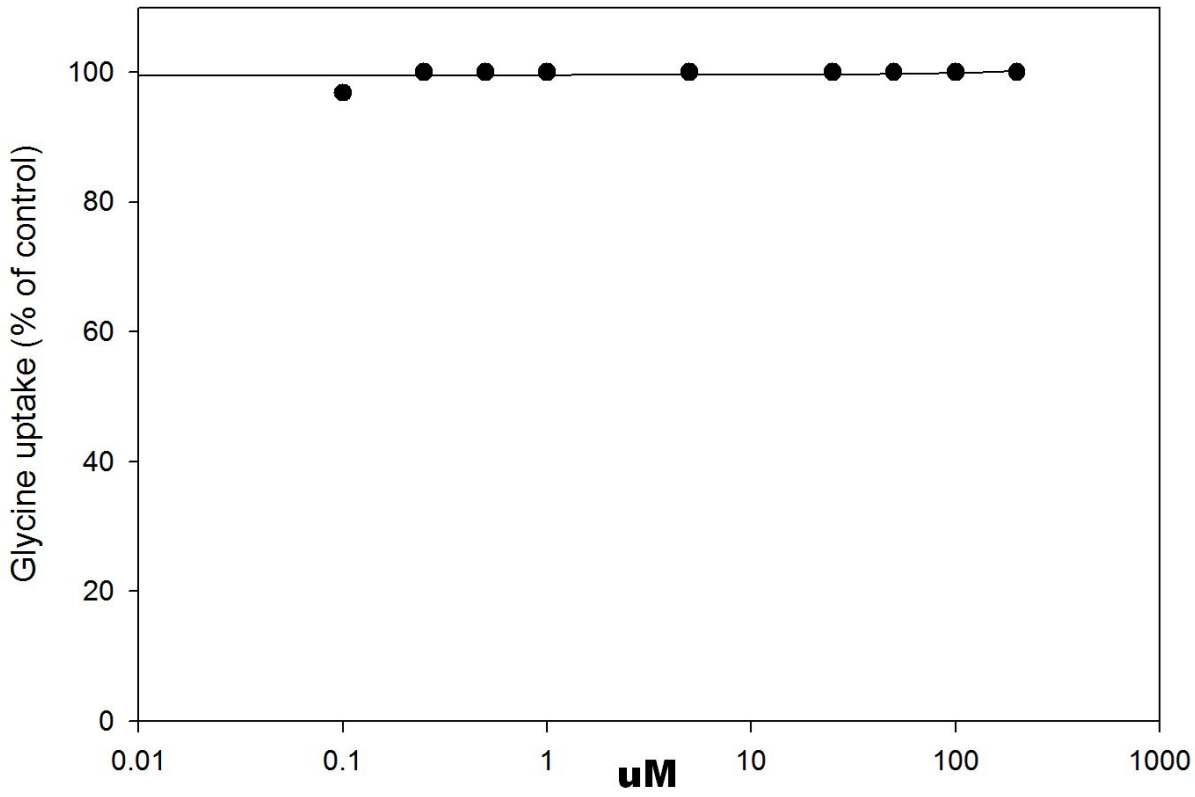


Figure S10: Inhibition of GlyT1 activity by Hit1.

Hit 2 (ZINC6865169)

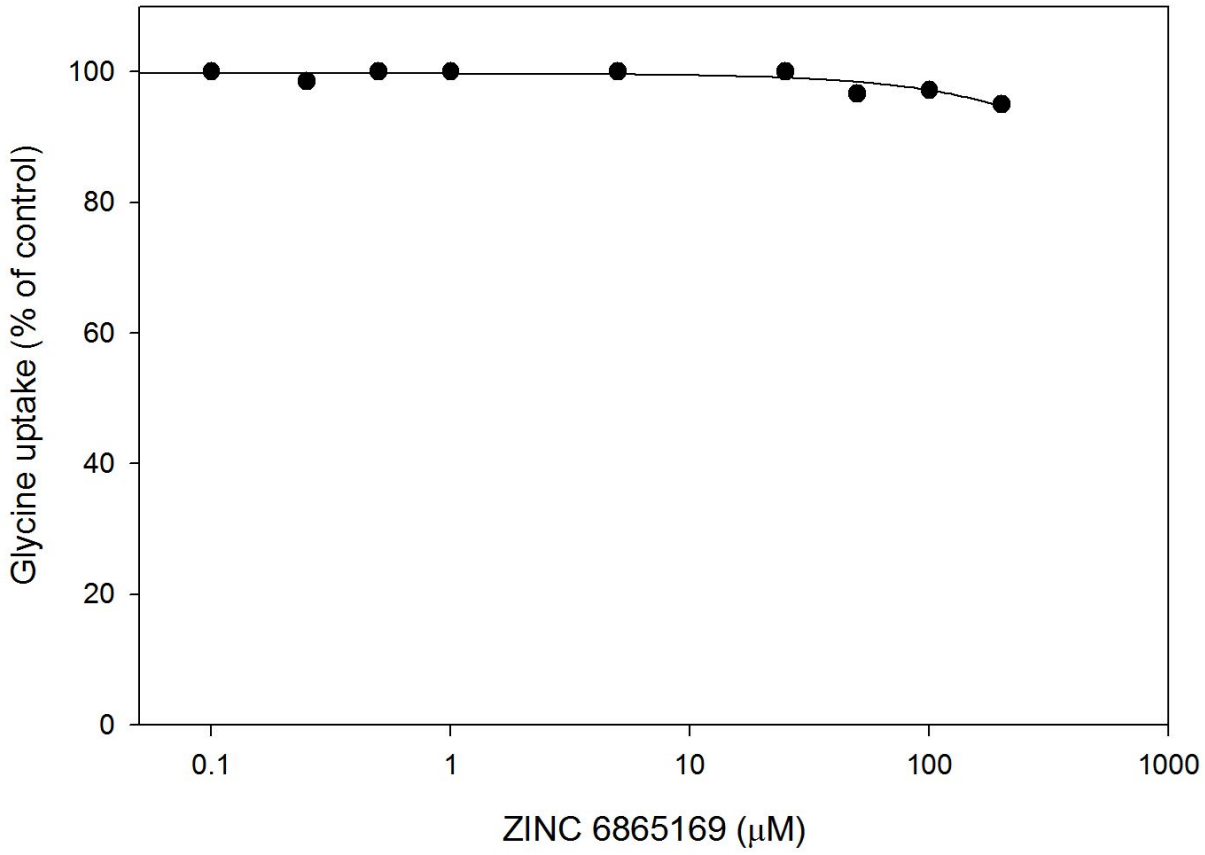


Figure S11: Inhibition of GlyT1 activity by Hit2.

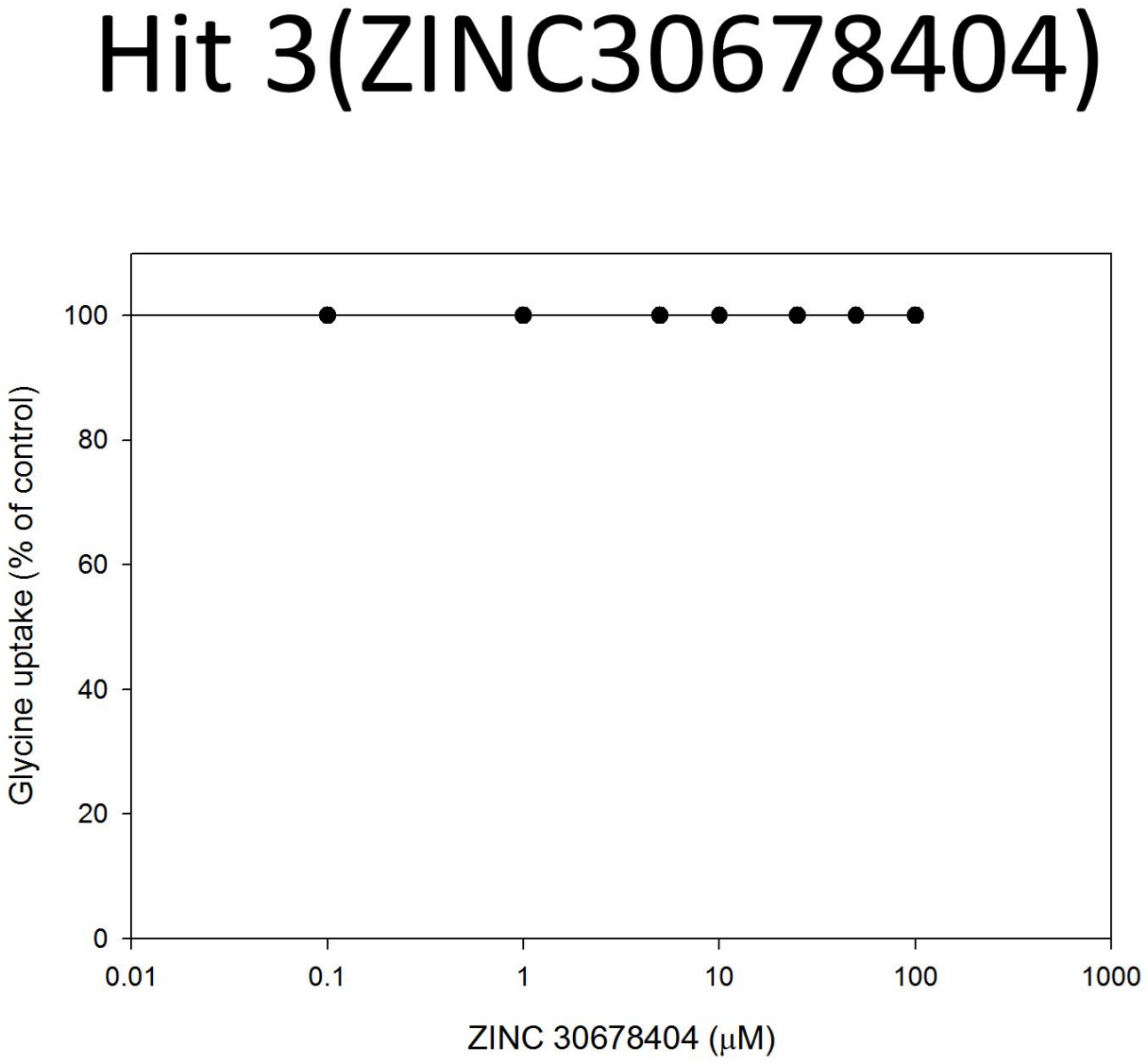


Figure S12: Inhibition of GlyT1 activity by Hit3.

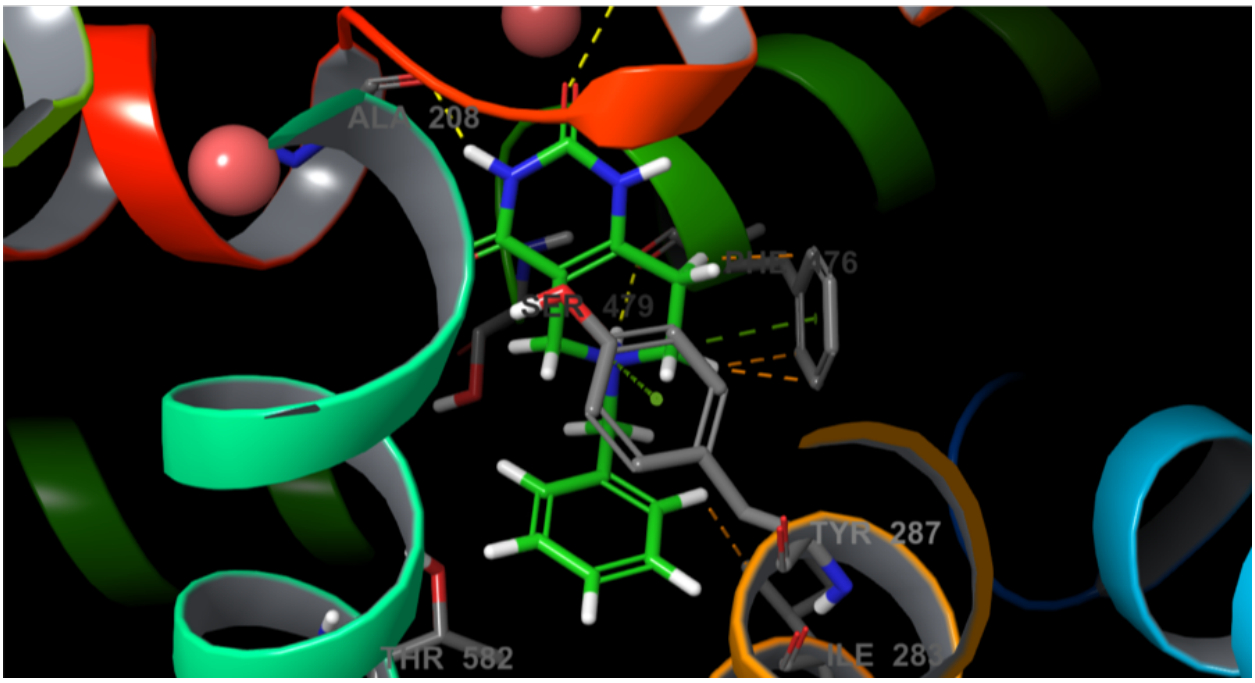


Figure S13: 3D representation of Hit 3-GlyT2 interactions as identified by our MD studies.

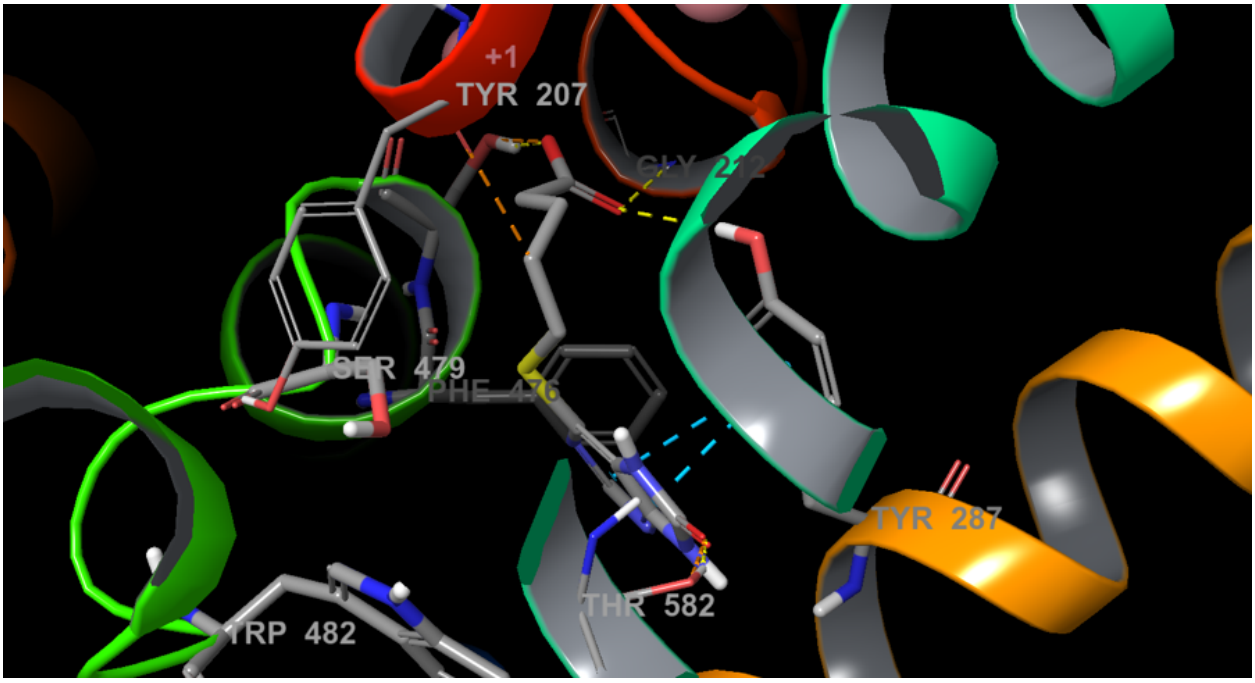


Figure S14: 3D representation of Hit 2-GlyT2 interactions as identified by our MD studies.

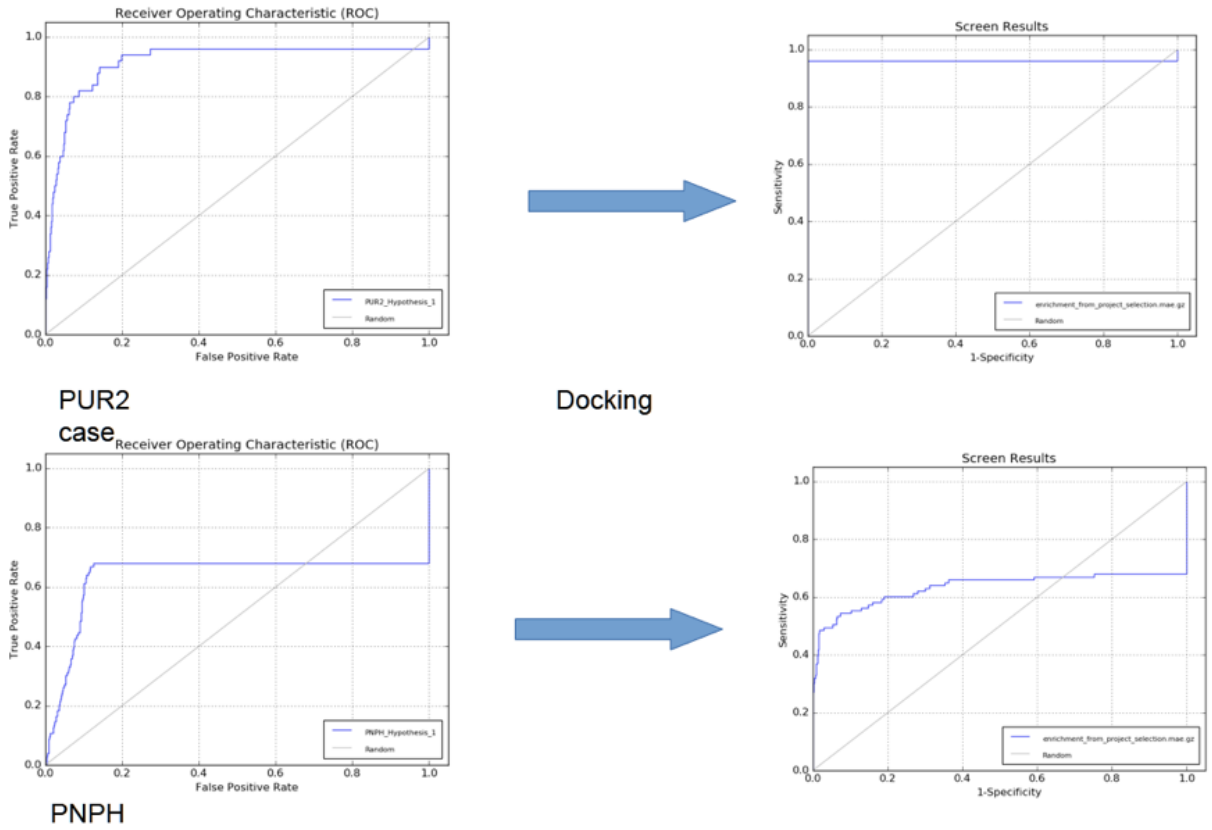


Figure S15: An example of structural based pharmacophore screen results obtained by Phase software (Schrodinger package version 2017-3) on DUD-E. The random selected targets on the figure show a reasonable results.

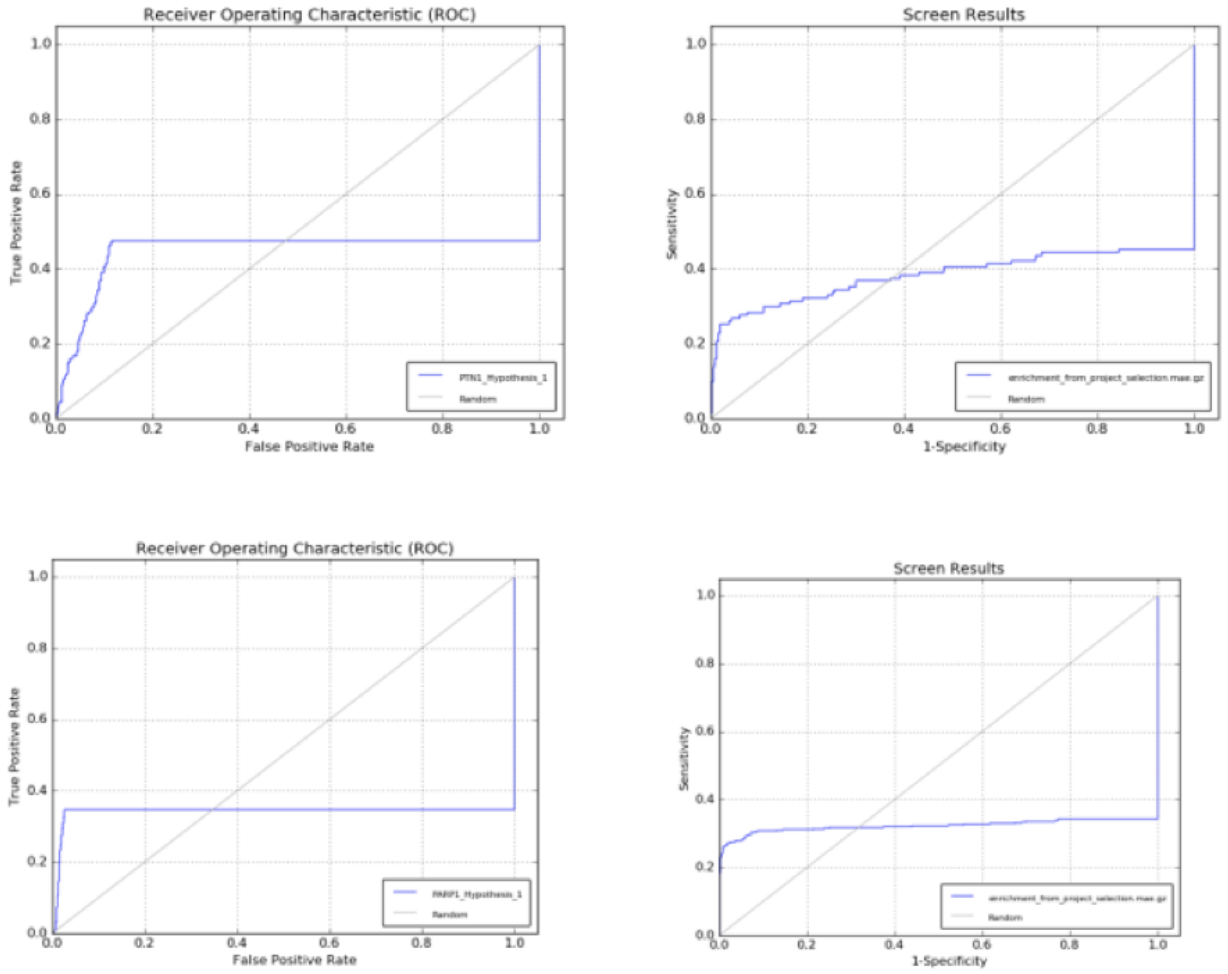


Figure S16: An example of the combination between structural based pharmacophore and docking screen results obtained by Phase and Glide software (Schrodinger package version 2017-3) on DUD-E. The random selected targets on the figure show an impressive result.

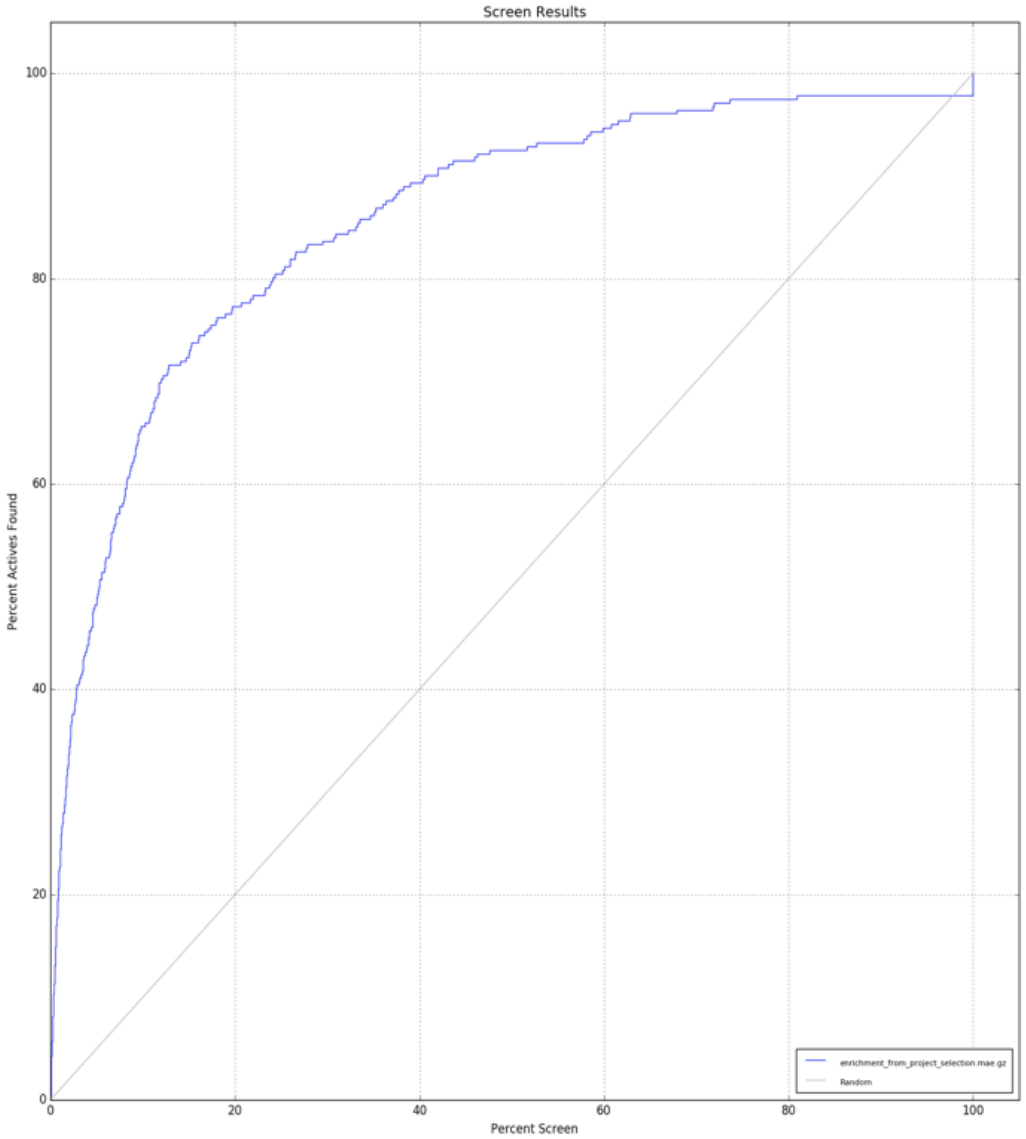


Figure S17: An improved Glide-SP scoring function was detected based on ACE test case

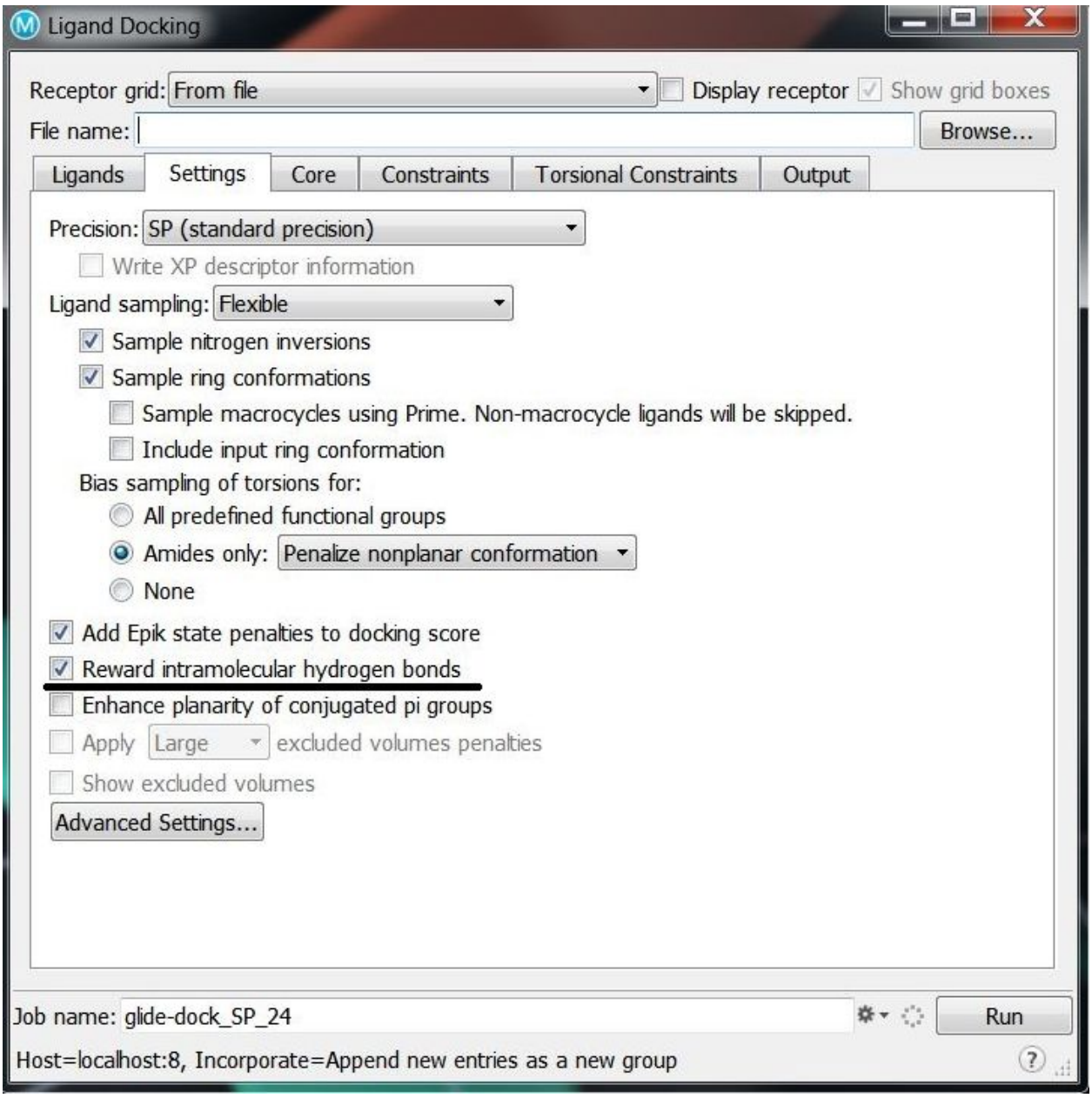


Figure S18: The Glide input settings used in this study

References

[1] Kevin H Wang, Aravind Penmatsa, and Eric Gouaux. Neurotransmitter and psychostimulant recognition by the dopamine transporter. *Nature*, 521(7552):322–327, may 2015.

[2] Jonathan A Coleman, Evan M Green, and Eric Gouaux. X-ray structures and mechanism of the human serotonin transporter. *Nature*, 532(7599):334–339, apr 2016.

[3] David Case and et. al. AMBER 16, 2016.

[4] Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11):1859–1865, aug 2008.

[5] Mikhail A Lomize, Irina D Pogozheva, Hyeon Joo, Henry I Mosberg, and Andrei L Lomize. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic acids research*, 40(Database issue):D370–6, jan 2012.

[6] Filip Fratev, Thomas Steinbrecher, and Svava Ósk Jónsdóttir. Prediction of Accurate Binding Modes Using Combination of Classical and Accelerated Molecular Dynamics and Free-Energy Perturbation Calculations: An Application to Toxicity Studies. *ACS Omega*, 3(4):4357–4371, apr 2018.

[7] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J.C Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, mar 1977.

[8] Henrik G. Petersen. Accuracy and efficiency of the particle mesh Ewald method. *The Journal of Chemical Physics*, 103(9):3668–3679, sep 1995.

[9] Filip F Fratev, Miguel Rivera, and Suman Sirimulla. *ACS SWRM2017 conference October 29 – November 1.* 2017.

[10] Schrödinger LLC. Schrödinger suite.

[11] Woody Sherman, Tyler Day, Matthew P Jacobson, Richard A Friesner, and Ramy Farid. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *Journal of Medicinal Chemistry*, 49(2):534–553, jan 2006.

[12] Anthony J Clark, Pratyush Tiwary, Ken Borrelli, Shulu Feng, Edward B Miller, Robert Abel, Richard A Friesner, and B J Berne. Prediction of Protein–Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations. *Journal of Chemical Theory and Computation*, 12(6):2990–2998, jun 2016.

[13] Edward Harder, Wolfgang Damm, Jon Maple, Chuanjie Wu, Mark Reboul, Jin Yu Xiang, Lingle Wang, Dmitry Lupyan, Markus K. Dahlgren, Jennifer L. Knight, Joseph W. Kaus, David S. Cerutti, Goran Krilov, William L. Jorgensen, Robert Abel, and Richard A. Friesner. OPLS3: A Force Field Providing Broad Coverage of Drug-like

Small Molecules and Proteins. *Journal of Chemical Theory and Computation*, 12(1):281–296, jan 2016.

[14] Kevin J. Bowers, David E. Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In *ACM/IEEE SC 2006 Conference (SC'06)*, pages 43–43. IEEE, nov 2006.

[15] Pu Liu, Byungchan Kim, Richard A Friesner, and B J Berne. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences*, 102(39):13749–13754, sep 2005.

[16] Lingle Wang, Yuqing Deng, Jennifer L. Knight, Yujie Wu, Byungchan Kim, Woody Sherman, John C. Shelley, Teng Lin, and Robert Abel. Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative Binding Affinity Predictions of CDK2 Inhibitors. *Journal of Chemical Theory and Computation*, 9(2):1282–1293, feb 2013.

[17] Susana P. Barrera, Vicente Castrejon-Tellez, Margarita Trinidad, Elisa Robles-Escajeda, Javier Vargas-Medrano, Armando Varela-Ramirez, and Manuel Miranda. PKC-Dependent GlyT1 Ubiquitination Occurs Independent of Phosphorylation: Inespecificity in Lysine Selection for Ubiquitination. *PLOS ONE*, 10(9):e0138897, sep 2015.

[18] A Mingorance-Le Meur, P Ghisdal, B Mullier, P De Ron, P Downey, C Van Der Perren, V Declercq, S Cornelis, M Famelart, J Van Asperen, E Jnoff, and J P Courade. Reversible inhibition of the glycine transporter GlyT2 circumvents acute toxicity while preserving efficacy in the treatment of pain. *British journal of pharmacology*, 170(5):1053–63, nov 2013.