# Essential Dynamics Ensemble Docking for Structure-Based GPCR Drug Discovery

Kyle McKay[#], Nicholas B. Hamilton[#], Jacob M. Remington, Severin T. Schneebeli, Jianing Li*

Department of Chemistry, University of Vermont, Burlington, VT 05405

[#] Equal contributions
*Email: Jianing Li (jianing.li@uvm.edu)

**Abstract**

The lack of biologically relevant protein structures can hinder rational design of small molecules to target G protein-coupled receptors (GPCRs). While ensemble docking using multiple models of the protein target is a promising technique for structure-based drug discovery, model clustering and selection still need further investigations to achieve both high accuracy and efficiency. In this work, we have developed an original ensemble docking approach, which identifies the most relevant conformations based on the essential dynamics of the protein pocket. This approach is applied to the study of small-molecule antagonists for the PAC1 receptor, a class B GPCR and a regulator of stress. As few as four representative PAC1 models are selected from simulations of a homology model and then used to screen three million compounds from the ZINC database and 23 experimentally validated compounds for PAC1 targeting. Our essential dynamics ensemble docking (EDED) approach can effectively reduce the number of false negatives in virtual screening and improve the accuracy to seek potent compounds. Given the cost and difficulties to determine membrane protein structures for all the relevant states, our methodology can be useful for future discovery of small molecules to target more other GPCRs, either with or without experimental structures.

**Key words:** Computer Aided Drug Design, PAC1 Receptor, Antagonist, Virtual Screening, Molecular Dynamics, Principal Component Analysis

## Introduction

Many G protein-coupled receptors (GPCRs) are being investigated as important therapeutic targets, but the success rate of structure-based drug design (SBDD) for GPCRs remains to be further improved.[1-3] One of the primary challenges is that the three-dimensional (3D) structures of most GPCRs have not been fully determined. Even with latest breakthroughs in protein structure prediction like AlphaFold,[4] the available structures may not represent the conformational states needed for accurate SBDD. The receptor (*ADCYAP1R1*, hereafter referred to as PAC1R) of the pituitary adenylate cyclase-activating peptide (PACAP), an emerging therapeutic target for stress-related disorders,[5-9] is a good example. Currently, the full-length PAC1R structures in the Protein Data Bank (PDB) are short isoforms (Uniprot ID: P41586-3),[10-11] but the structures of the most prevalent long isoforms — PAC1null (Uniprot ID: P41586) or PAC1hop (Uniprot ID: P41586-2) — are still unavailable.[12] All the published structures of PAC1R are complexed with peptide agonists and a heterotrimeric G protein complex (Figure 1), and thus do not represent the inactive conformations required for antagonist development. So far, it is thought that over 40% of GPCRs have more than one isoforms,[13] and each GPCR can adopt multiple conformational states which can be stabilized upon interactions with binding partners.[14-15] For accurate SBDD, it is important to employ conformations of the most medically relevant isoform, as it is to this ensemble of 3D pocket structures that the drug must show affinity. Here, we used PAC1R as a model system and investigated how to improve modeling accuracy and to gain predictive power for SBDD with limited 3D structural information, using the method of Essential Dynamics Ensemble Docking (EDED). With the proof of principle, this method can be readily generalized to develop new therapeutic targets to target a wider range of GPCRs.

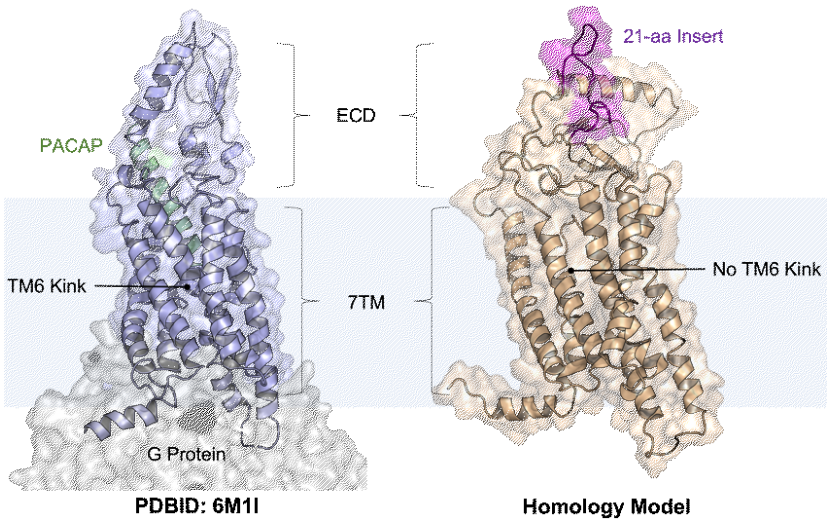



**Figure 1.** Cartoon illustrations of the PACAP-bound PAC1R model (PDBID: 6M1I, PAC1very short) and our homology model (template PDBID: 4L6R, PAC1null, simulation snapshot at 500 ns). The PAC1null isoform is more biomedically relevant than the very short isoform. The PACAP peptide is shown as a helix cartoon (pale green); the 21-amino acid ECD insert (see the sequence in Figure S1) is shown as a flexible coil (purple). This study focused on docking to the peptide-binding pocket.

PAC1R and its endogenous peptide hormone PACAP play an important role in neural development, calcium homeostasis, glucose metabolism, circadian rhythm, thermoregulation, inflammation, feeding behavior, pain modulation, as well as stress and related endocrine

responses.[16-18] For example, increased levels of PACAP in the blood have been reported in women diagnosed with post-traumatic stress disorder,[8] implicating chronic activation of the PAC1R in the disorder. Other studies[7, 19] have suggested that PAC1R activation mediates the adverse emotional consequences of chronic pain via downstream MAPK/ERK activation. Thus, these prior studies indicate that PAC1R antagonism, especially with small-molecule antagonists, represents a new strategy to treat stress, chronic pain, and related disorders.[8] Similar to other class B GPCRs, PAC1R possesses a heptahelical transmembrane domain (7TM) and an extracellular domain (ECD).[1] Most of the neural and peripheral tissues known to date contain the PAC1null or PAC1hop isoforms that includes a 21-amino acid insert in the ECD (Figure 1), which is missing in available PAC1R structures in the PDB.[20] This ECD insert was found highly dynamic in our previous modeling studies,[21-22] but its role in regulating PAC1R remains unknown. While PAC1R antagonists are being developed as potential treatments for stress-related disorders, the agonist-bound cryo-EM structures are not directly applicable to computational design or screening of PAC1R antagonists. GPCRs spontaneously adapt active and inactive signaling states, each of which are characterized by broad conformational ensembles. In a conformational selection view, agonists and antagonists stabilize GPCR conformations of the active and inactive ensembles, respectively.[19, 23] It is now well accepted that to accurately design GPCR ligands as drug candidates, one should use active conformations for agonist design and inactive conformations for antagonist design. With the transition between active and inactive GPCR conformations occurring on the millisecond timescale,[24-27] it is computationally demanding to obtain the inactive PAC1R conformations from the agonist-bound cryo-EM structures via molecular dynamics (MD) simulations. Instead, we seek to use a homology model in this work and test with the EDED method.

Ensemble docking utilizes multiple receptor models for pocket sampling, obtained from clustering the conformations sampled by MD simulations for molecular docking, and displays noted improvement at identifying GPCR ligands when compared to docking against a single experimental structure.[28-38] EDED is distinct from prior ensemble docking approaches, mainly in clustering and selection of receptor models. Global root mean square deviation (RMSD) is convenient to cluster similar structures, but the highly dynamic extracellular and intracellular loops (ECLs and ICLs) of GPCRs can significantly compromise the otherwise good similarity between the 7TM structures. Thus, clustering based on global RMSD can generate many models that, while representative of global changes, are irrelevant to the intricate differences within the local binding pocket of the GPCR. This additional overhead ultimately lowers both the efficiency and accuracy of ensemble docking when using the global RMSD approach for clustering. EDED avoids this issue by focusing on both local similarity and essential dynamics of the binding pocket. Although computational power is more accessible than ever, streamlined workflows which expend computational resources only on worthwhile calculations are always desirable. Herein, we applied EDED to PAC1R with as few as four receptor models, whose results show a reduced false negative rate and a good correlation between the small molecule efficacy and the predicted score. Our results provide the evidence for initial success to develop small-molecule antagonists for PAC1R and pave the way for future structure-based GPCR drug discovery.

**Results and Discussion.**

1. Inactive conformations of PAC1null and key interactions with small molecules

Towards discovery of novel PAC1R antagonists, the inactive state conformational ensemble of PAC1R was estimated using an all-atom MD simulation of a ligand-bound PAC1R model from homology modeling (Figure 1). Our reference ligand is an analog of known PAC1 antagonists[39] that were discovered previously using structure activity relationships. We created

the antagonist-bound model by docking the reference compound into the PAC1R homology model. This complex model was simulated in the POPC membrane for 500 ns, and for the entire length of the simulation the ligand remained bound in roughly the starting conformation (Figure S2). Other features like the closed ECD and straight transmembrane helix six (TM6), as well as short separation between TM3, TM5 and TM6, are consistent with a deactivated structure of a class B GPCR.[40]

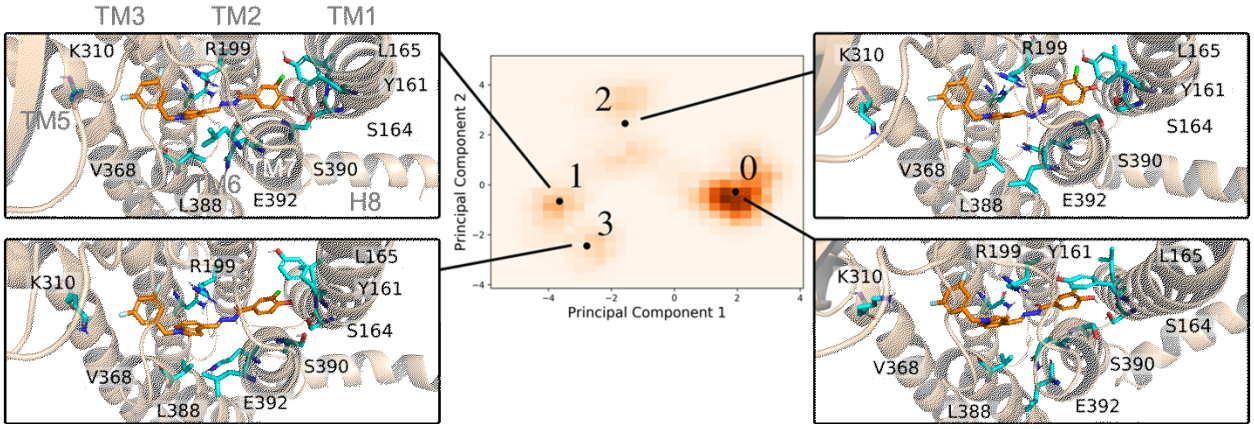



**Figure 2.** Four representative PAC1R conformations using in EDED reveal important changes in the binding pocket. The protein is shown in cartoon and the reference ligand is shown as sticks. The histogram of all trajectory frames projected onto the first two principal components of residues within the ligand-binding pocket of PAC1R. Black dots labelled with numbers from 0 to 3 are the representative structures (S0, S1, S2, and S3) determined by the *K*-means clustering algorithm.

Despite the overall appearance of an inactivated receptor, there were critical changes within the orthosteric pocket during the MD simulations. Using EDED, four members of the inactive conformational ensemble (states S0, S1, S2, and S3 ordered by observed population) were extracted and reveal distinct conformations of the 7TM helices and different side chain orientations within the binding pocket (Figure 2). For one, bending of TM1 was observed to follow S0 < S2 < S3 < S1, where the most populated state (S0) was the most straightened helix. This correlated with local changes to residues Y161, L165, and S164 on TM1, and most significantly the stiffened TM1 in the S0 state enabled both π-stacking (with Y161) and hydrogen bonding (with S164) interactions. On the other hand, displacement of TM7 in the S1 state relative to S0 caused replacement of the hydrogen bond with S164 in the S0 state with a new hydrogen bond with S390. The interactions between the indole on the ligand and V368, L388, and E392 were modulated between the different receptor states with generally tighter interactions in the S1 and S3 states, in comparison with the S0 and S2 states. In addition, changes in TM5 affected the ability of K310 to form the stable interactions with the electron rich substituents on the ligand in states S0 and S3 which were diminished in the S1 and S2 states. Ultimately, this analysis reveals how EDED is able capture the subtle changes in pocket structure that are highly relevant for accurate modeling of ligand-receptor interactions when performing SBDD.

2. Comparison of Docking to a Single Receptor Model and to the Conformational Representatives.

Compared with docking to the ligand-free homology model and ligand-free cryo-EM structure, EDED significantly improved the identification of candidate compounds (Figure 3). The average binding score of the top 350 (approximately 2.5%) of compounds docked to the ligand-free homology model improved from -5.9 to -9.4 kcal/mol when docked against the ensemble.

Likewise, it improved from an average of -5.8 to -9.4 kcal/mol when compared to the PACAP-bound model (PDBID: 6M1I). This gives an average 3.6 kcal/mol improvement in average docking score of the top selected compounds. Additionally, EDED identified six compounds predicted to bind to PAC1R with comparable binding score (-11.2 kcal/mol) as our reference ligand.

To gain physical insight into the improvement of the docking scores, the binding pose(s) of the top compounds from both methodologies were examined. We have previously reported the key role of R199 in PACAP-induced activation of PAC1R.[11] This is further corroborated by strong cation-pi interactions with the residue in our models. Interaction with R199 across all the ensemble conformations became a critical determining factor for which top ensemble docking compounds should be prioritized for synthesis and/or computational optimization. Examining the compounds which have ensemble docking scores close to or better than our reference ligand, this interaction is present for all six top scoring ligands in *at least* one of the docking poses. This is in contrast with the homology model and the PACAP-bound model where only relatively few of the top compounds from this methodology were able to engage in this key interaction. Also, of note are induced fit effects where the MD simulation of our reference ligand in the pocket may affect the binding pocket through subtle shifts in the backbone and the rotation of side chains. In the rigid receptor docking to the homology model, the 7TM helical bundle is closer together, defining a more compact orthosteric pocket. Thus, it is only accessible for small ligands to bind deep into the pocket below R199. In contrast, the conformations in the ensemble docking are more open, better allowing ligands to access the pocket. This can be seen by where most ligands found their best pose. Although both datasets were docked against a grid centered on R199, the ensemble docking results have the majority of top ligands below the residue, low in the pocket. When docked against the homology model, the top ligands are higher in the pocket at lowest in line with R199.

The new ligands examined within the orthosteric pocket showcased the ability of ensemble docking to provide integral confirmations omitted by static modelling, with the ensemble approach providing key ligand poses corresponding to interactions with new side chains revealed in the ensemble. Aside from R199, several key contacts were discovered from study of the top ligands bound to each receptor in the ensemble (Figure S3). These contacts expand the understanding of the orthosteric pocket dynamics and can be exploited in small molecule rational design. In comparison with consistent interactions to the ligand-binding pocket of the homology model, these results suggests that EDED may reveal new crucial ligand-receptor interactions even from a rigid template.

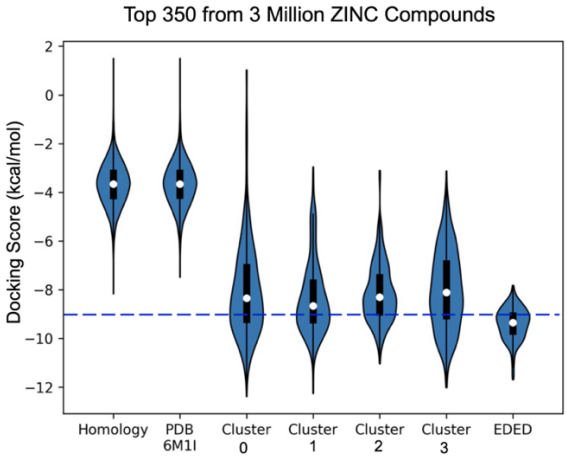



**Figure 3.** Violin plots of the docking score distribution of the top 350 compounds to different receptor models. The dash line shows the -9.0 kcal/mol cutoff used to prioritize compounds for synthesis.

A thermodynamically driven approach to scoring the binding poses of a given compound to multiple receptor structures was used to assess the binding affinity of the docked ligands. This approach quantitatively captures various physical phenomena that are often considered when computing overall docking scores: (i) the relative likelihood of the receptor obtaining the different conformations are explicitly included, and (ii) the binding of the ligand to the receptor changes the energies of the complex differentially in the distinct conformations. Importantly, this model properly handles confounding cases that other approaches, such as a simple direct averaging of different docking score, would not describe well. For instance, for any given ligand, a protein is hypothetically able to adopt on an unlikely conformation ( $\Delta E\_conf_{1,i} \gg 0$ , i.e., much higher score than the structure with lowest score) where the binding of the ligand to the protein could be quite favorable ($-\Delta E_{bind_i}$ approximately equal to 10 kT). Simply including this state in an average of docking scores would treat it as equivalently important as conformations that are far more relevant to the signaling states of the protein. Our approach avoids such errors, by including the energetics of binding in the model, assuring that the overall energy of these rare states is indeed still relatively high and do not contribute significantly to the final score in Eq. 1. In sum, our docking score considers the difference in overall energies of the bound receptor conformations and is appropriate for comparison with a physical experiment that is unlikely to be able to distinguish between different bound conformations (Eq. 1).

$$e^{\Delta E/kT} = e^{(E_{bound} - E_{unbound})/kT} = \frac{P_{unbound}}{P_{bound}} = \frac{\sum_{i=1}^{n} P_{cluster\ i}}{\sum_{i=1}^{n} P_{complex\ i}} \quad \text{(Eq. 1)}$$

Where $E_{bound}$ and $E_{unbound}$ are the energies associated with the ligand being bound or unbound to any receptor conformation, respectively, $P_{bound}$ and $P_{unbound}$ are the total probabilities of the ligand being bound or unbound to any receptor conformation in the ensemble, respectively, $P_{cluster\ i}$ is the probability of a specific receptor conformation (calculated from the MD, see SI for more information), and $P_{complex\ i}$ is the probability of the ligand being bound to that specific ensemble conformation. We note that our model is still more appropriate than equal weighting for cases where one does not trust the relative energies of the different conformations obtained directly from the MD simulations. In such cases setting the $\Delta E\_conf_{1,i}$ to 0 for each conformation (i.e., each conformation is equally likely) reduces Eq. 1 to Eq. 2.

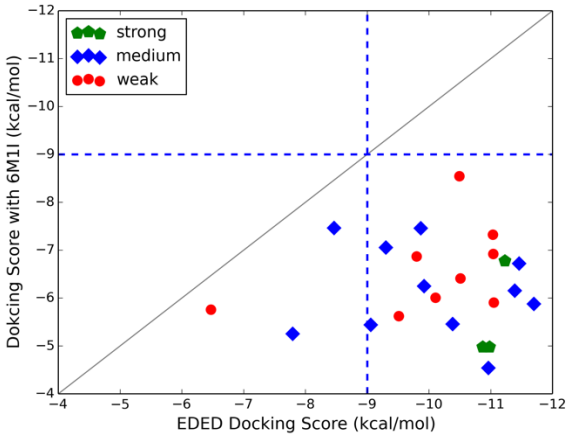
$$\Delta \mathrm{E}_{bind,equal\ weighting} = \ln\left(\frac{n}{\sum_{i=1}^{n} e^{-\Delta E bind_i}}\right) kT \quad \text{(Eq. 2)}$$

Clearly, Eq. 2 is not a simple weighted average of the different binding scores, however to our knowledge this analysis is lacking in the literature.

3. Evaluation of EDED predictions.

Additional to testing EDED with compounds from ZINC, we also tested 23 small-molecule compounds which were classified as strong, moderate, and weak antagonists in PAC1R activity assays (unpublished data from Prof. Victor May). The design of these small molecules was based on previously published work outlining the structure-activity relationship between small molecules and the PAC1 receptor.[39] Ligand-based virtual screening was then performed and yielded the 23 compounds which were experimentally tested. Docking each analog against all four conformations in the ensemble and scoring them as previous described (Eq. 1) shows modest correlation to experimental results (Figure 4). The strong experimental antagonist had the highest predicted binding affinities with an average -10.4 kcal/mol, while the moderate and weak

antagonists both had worse predicted binding affinities -9.8 kcal/mol and -8.5 kcal/mol, respectively.




Experimental Positives: 14
Experimental Negatives: 9

| | EDED | 6M1I |
|---|---|---|
| TP | 12 | 0 |
| TN | 1 | 9 |
| FP | 8 | 0 |
| FN | 2 | 14 |
| FPR | 89% | 0% |
| FNR | 14% | 100% |
| ACC | 57% | 0% |

**Figure 4.** Ensemble weighted glide scores ($\Delta G_{bind}$) of 23 experimentally tested compounds. Compounds with strong, modest, and poor ERK inhibitive activity are depicted in green, blue, and red, respectively. Corresponding colored lines represent the average ensemble weighted glide score for that category. A cutoff of -9 kcal/mol was applied for predicted antagonists to be compared to their experimental results showing either strong or medium inhibition (active) or weak inhibition (inactive).

It is worth noting that our EDED method is best used to identify potential antagonists from a collection of compounds, but the dockings scores (like Glide SP, XP, and our EDED score) to estimate binding energies should be interpreted with caution.[41-43] While we successfully reduced the false negative rate (FNR) with EDED, there is still a high false positive rate (FPR). A delicate balance between ensemble size and the FPR has previously been reported, inspiring us to select a relatively small ensemble for analysis.[44] Our FPR is comparable to prior studies employing both ensemble and static methods for virtual screening.[45-47] Additionally, the experimental assays provided here are a measure of antagonistic ability, and not binding affinity. As quantitative binding assays remain to be performed, it is possible some of the false positives (compounds with poor experimental results but high ensemble docking scores) bind tightly but are not effective antagonists, i.e., they do not stabilize the inactive conformations or prevent cognate ligand binding in other ways. With the extended view provided by EDED, we envision that the chance of obtaining a false negative prediction is likely reduced in our model when compared with static Glide docking. This added width within the sampled energy landscape (from the new side chain confirmations) allows our EDED method to achieve more accurate sampling of potential ligand-receptor interactions, thus increasing the chances of finding a hit compound otherwise overlooked in the static model. Overall, EDED displayed an accuracy of 57% in predicted binding affinity when compared to our experimental results, an increase when compared with Glide's empirical scoring function[48]. Combined with the overall low variance in EDED docking scores for the top 350 compounds analyzed (Figure 3), we believe our methodology represents a robust route for the recognition of small molecules with high receptor affinity.

**Conclusions.**

In conclusion, we have developed and implemented EDED, an ensemble docking inspired methodology for SBDD. By focusing on the essential dynamics of the ligand binding pocket, our method is distinct from many prior studies that built receptor clusters solely based on the root

mean square deviation (RMSD) of the entire protein backbone.[49] Further, the use of clustering within this reduced dimensionality conformational space directly considers the local structural similarity of the ligand-binding pocket. We demonstrate that EDED captures the critical changes in the 3D structure of the binding pocket that are known to correlate strongly with binding affinity of ligands. Our approach is partially based on the assumption that differences in the binding pocket itself (as opposed to the protein as a whole) predominately give rise to the different binding poses and energies that are the goal of any ensemble docking workflow. Using the EDED derived representative structures, we screened a large dataset of compounds and successfully identified novel small molecule antagonists of the PAC1 receptor. However, EDED is not specific to a single GPCR and will likely accelerate the design of small molecule drugs that target other GPCRs with currently unknown conformational states.

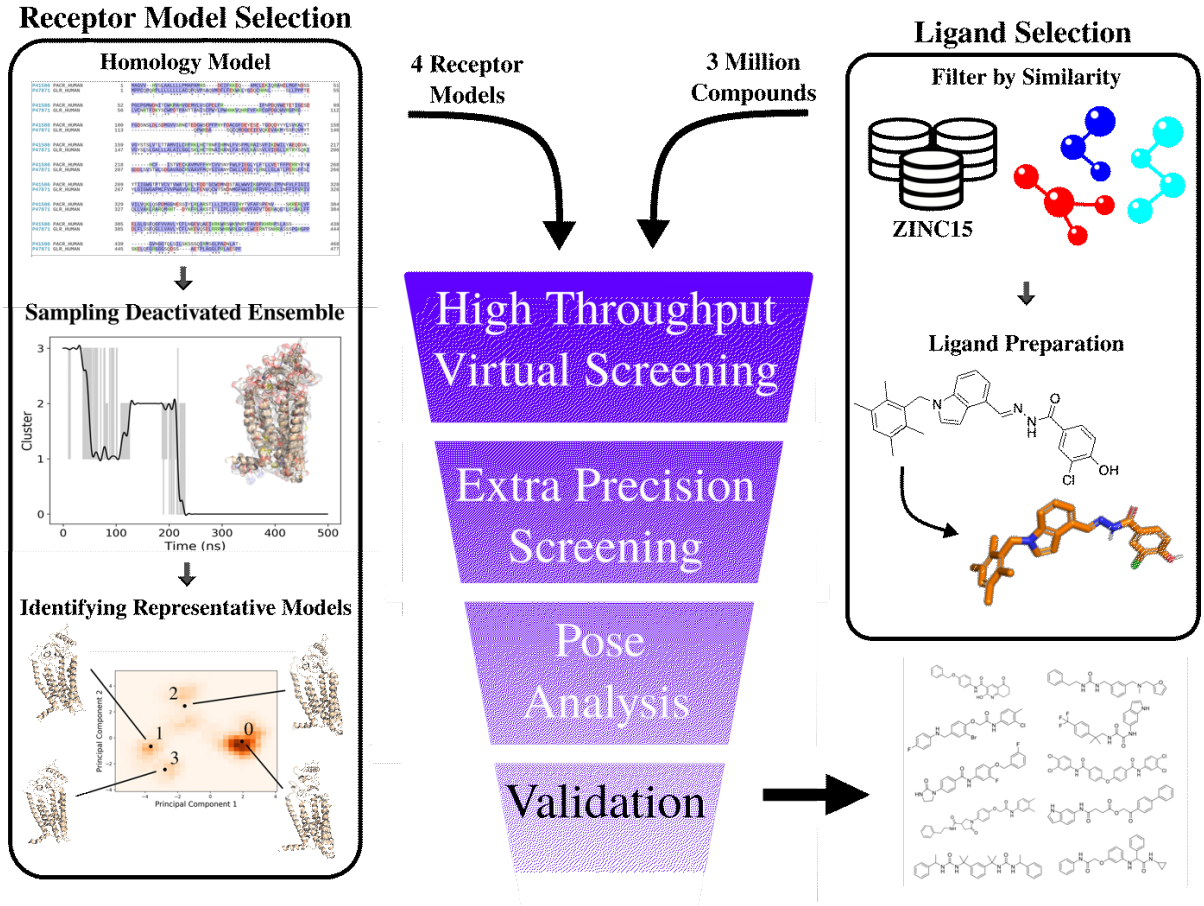



**Figure 5.** Overview of computational workflow for development of PAC1R antagonists. Right Column: selection of input ligands from a structure database (in this example the ZINC15[50] database). Custom filters were used to select raw structures with desirable properties (molecular weight, log*P*, etc.). These structures are then prepared using Schrödinger's *ligprep* software program. Left column): the PAC1null homology model is constructed from the protein's sequence, simulated for 500 ns, and the raw coordinates are analyzed. The representative structures are used in ensemble docking. Hit compounds are selected based on visual inspection of the results.

## Methods and Models.

Receptor Model Preparation in EDED. One key idea of EDED is to obtain chemically relevant receptor models for docking. Instead of using the agonist-bound PAC1R structure, we generated a homology model of inactive PAC1R (with the canonical variant sequence, Uniprot ID: P41586) with a template of the glucagon receptor (PDBID: 4L6R, ~40% similarity).[19] This PAC1R model incorporated the inactive features of class B GPCRs such as a continuous helix along TM6 and a closed ECD. A small-molecule PAC1R antagonist, our reference ligand, was placed in the orthosteric pocket via molecular docking (Glide, Schrödinger Inc.). The complex model was later simulated to sample the inactive conformational ensemble.

Receptor Model Sampling in EDED. To sample inactive conformations for docking, the ligand-bound PAC1R model was simulated with the OPLS3[51] force field in explicit SPC solvent in the NPT ensemble (300K, 1 atm, Martyna-Tuckerman Klein coupling scheme) using classical MD simulations. A POPC membrane was place around the 7TM using the Orientations of Proteins in Membrane (OPM) database.[52] The simulation was performed in the Maestro-Desmond program[53] (GPU version 5.4) with a timestep of 2 fs, recording interval of 4.8 ps, and a total simulation time of 500 ns. The Ewald technique was used for the electrostatic calculations. The van der Waals and short-range electrostatic interactions were cut off at 9 Å. Hydrogen atoms were constrained using the SHAKE algorithm. Two extended simulations were also examined to confirm the ligand poses and receptor confirmations. Once again, a POPC membrane was placed around the 7TM bundle using OPM. NAMD 2.11 was used as the simulation package for these replicates[54]. The CHARMM36 forcefield[55] was used with a TIP3 solvent model in a NPT ensemble (310K, 1 atm) Force switching was utilized at the range of 10-12 Å to approximate the LJ interactions. Langevin piston/Nose-Hoover[56-57] methods were utilized for the pressure control with a piston period of 50fs and a decay time of 25 fs. Langevin coupling of these simulations with a dampening coefficient of 1 $ps^{-1}$ was also utilized. Long range electrostatic interactions were modeled with the particle mesh Ewald method.[58] These simulations were run with a 2 fs timestep and combined for 350ns of data. MD trajectories were analyzed using in-house Python and TCL scripts as well as Visual Molecular Dynamics (VMD).[59]

Receptor Ensemble Selection in EDED. We first aligned the 7TM of PAC1R (residues 156-405) to the homology model to reduce noise due to translational movement. Next, the coordinates of the centers of mass for any residue whose side chain was within 3 Å of any ligand atom in the static model were collected and parsed using in-house designed TCL and python scripts. A dimension reduction based on principal component analysis (PCA) was used to determine which collective motions (termed principal components, PCs) contributed most to variations in the overall conformations of the binding pocket. The first fifteen PCs (accounting for 90% of the cumulative variance) were clustered using a *K*-means clustering algorithm implemented by PyEmma.[22] Based on inspection of the first two PCs (Figure 5), four cluster centers were identified. As these cluster centers are not precise frames within the trajectory but are instead points in the PC space, the cluster centers' PC coordinates were approximately projected back to the original Cartesian coordinates. Frames from the trajectory which had PC values closest to the centers based on a RMSD measurement, were then selected as the ensemble docking receptor structures. This approach allowed a minimum of representative frames to capture the most variance of the binding pocket as opposed to other methodologies which often have many structures. Also, our physics-based approach is transferrable to other GPCRs and expanded clustering. In fact, our focus on the relevant receptor models likely requires less sampling in MD simulations and fewer clusters for subsequent docking, a practical advantage for large-scale screening.

Docking and Scoring of Potential PAC1R Antagonists. Receptor grid models were generated using the three-dimensional structures selected as detailed above with R199 selected as the center of the docking box with an 18-Å cutoff. Docking was carried out using Schrödinger Virtual Screening Workflow[60] (VSW) at three consecutive levels of precision, both for small molecules docked to the static homology model and to the conformation ensemble. Small molecules docked to our PAC1null ensembled were given an overall score, Ensemble $\Delta G_{bind}$, based on Eq. 3.

$$\text{Ensemble } \Delta G_{bind} = \ln\left(\frac{1+\sum_{i=2}^{n} e^{-\Delta E conf_{1,i}}}{\sum_{i=1}^{n} e^{-\Delta E conf_{1,i}-\Delta E bind_i}}\right) kT \quad \text{(Eq. 3)}$$

In Eq. 3, $\Delta E\_conf_{1,i}$ is the difference in energy (in units of kT) between the lowest energy receptor conformation and each subsequent conformation calculated using the clustered trajectory, and $-\Delta E_{bind_i}$ is the corresponding Glide XP docking score to that same conformation. While $\Delta E\_conf_{1,i}$ is representative of the apo receptor free energy, it is worth noting that simulation data used to generate these confirmations included the ligand bound within the pocket.

Docking was carried out against compounds (i) pseudo-randomly selected from the ZINC15[50] database, (ii) as analogs of known antagonists to the static ligand-free homology model, the cryo-EM structure, and the conformational ensemble. In total, a small test set of 10,000 drug-like compounds were selected and download from the ZINC database and docked using Schrödinger's VSW as described previously.

**Acknowledgements**

We thank Drs. Victor May and Matthias Brewer (UVM) for sharing the experimental data and for helpful discussions. The work was mainly supported by the NIH grant R01-GM129431 to JL. S.T.S. was partially supported by the Army Research Office (Grant 71015-CH-YIP). Part of the computational facilities was also supported by an NSF CAREER award (Grant CHE-1848444 awarded to STS).

**Author Contributions**

KM, JR, STS and JL contributed to the conception and design of the study. KM performed the model preparation, simulation setup, docking, and docking analysis. KM and NH performed the simulation analysis. JR derived the EDED scoring function. KM, JR, and JL wrote the first draft of the manuscript. KM, JR, NH, STS and JL wrote sections of the manuscript. All authors contributed to the revision, read, and approved the submitted version.

**References**

1. Odoemelam, C. S.; Percival, B.; Wallis, H.; Chang, M.-W.; Ahmad, Z.; Scholey, D.; Burton, E.; Williams, I. H.; Kamerlin, C. L.; Wilson, P. B., G-Protein coupled receptors: structure and function in drug discovery. *RSC Advances* **2020,** *10* (60), 36337-36348.

2. Wootten, D.; Christopoulos, A.; Marti-Solano, M.; Babu, M. M.; Sexton, P. M., Mechanisms of signalling and biased agonism in G protein-coupled receptors. *Nature Reviews Molecular Cell Biology* **2018,** *19* (10), 638-653.

3. Hauser, A. S.; Attwood, M. M.; Rask-Andersen, M.; Schiöth, H. B.; Gloriam, D. E., Trends in GPCR drug discovery: new agents, targets and indications. *Nature Reviews Drug Discovery* **2017,** *16* (12), 829-842.

4. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021,** *596* (7873), 583-589.

5. Hammack, S. E.; Cheung, J.; Rhodes, K. M.; Schutz, K. C.; Falls, W. A.; Braas, K. M.; May, V., Chronic stress increases pituitary adenylate cyclase-activating peptide (PACAP) and brain-derived neurotrophic factor (BDNF) mRNA expression in the bed nucleus of the stria terminalis (BNST): Roles for PACAP in anxiety-like behavior. *Psychoneuroendocrinology* **2009,** *34* (6), 833-843.

6. Roman, C. W.; Lezak, K. R.; Hartsock, M. J.; Falls, W. A.; Braas, K. M.; Howard, A. B.; Hammack, S. E.; May, V., PAC1 receptor antagonism in the bed nucleus of the stria terminalis (BNST) attenuates the endocrine and behavioral consequences of chronic stress. *Psychoneuroendocrinology* **2014,** *47*, 151-165.

7. Missig, G.; Mei, L.; Vizzard, M. A.; Braas, K. M.; Waschek, J. A.; Ressler, K. J.; Hammack, S. E.; May, V., Parabrachial Pituitary Adenylate Cyclase-Activating Polypeptide Activation of Amygdala Endosomal Extracellular Signal–Regulated Kinase Signaling Regulates the Emotional Component of Pain. *Biological Psychiatry* **2017,** *81* (8), 671-682.

8. Ressler, K. J.; Mercer, K. B.; Bradley, B.; Jovanovic, T.; Mahan, A.; Kerley, K.; Norrholm, S. D.; Kilaru, V.; Smith, A. K.; Myers, A. J.; Ramirez, M.; Engel, A.; Hammack, S. E.; Toufexis, D.; Braas, K. M.; Binder, E. B.; May, V., Post-traumatic stress disorder is associated with PACAP and the PAC1 receptor. *Nature* **2011,** *470* (7335), 492-497.

9. Liao, C.; May, V.; Li, J., PAC1 Receptors: Shapeshifters in Motion. *Journal of Molecular Neuroscience* **2019,** *68* (3), 331-339.

10. Kobayashi, K.; Shihoya, W.; Nishizawa, T.; Kadji, F. M. N.; Aoki, J.; Inoue, A.; Nureki, O., Cryo-EM structure of the human PAC1 receptor coupled to an engineered heterotrimeric G protein. *Nature Structural & Molecular Biology* **2020,** *27* (3), 274-280.

11. Liang, Y.-L.; Belousoff, M. J.; Zhao, P.; Koole, C.; Fletcher, M. M.; Truong, T. T.; Julita, V.; Christopoulos, G.; Xu, H. E.; Zhang, Y.; Khoshouei, M.; Christopoulos, A.; Danev, R.; Sexton, P. M.; Wootten, D., Toward a Structural Understanding of Class B GPCR Peptide Binding and Activation. *Molecular Cell* **2020,** *77* (3), 656-668.e5.

12. Liao, C.; Poujol de Molliens, M.; Schneebeli, S.; Brewer, M.; Song, G.; Chatenet, D.; Braas, K.; May, V.; Li, J., Targeting the PAC1 Receptor for Neurological and Metabolic Disorders. *Current Topics in Medicinal Chemistry* **2019,** *19*.

13. Marti-Solano, M.; Crilly, S. E.; Malinverni, D.; Munk, C.; Harris, M.; Pearce, A.; Quon, T.; Mackenzie, A. E.; Wang, X.; Peng, J.; Tobin, A. B.; Ladds, G.; Milligan, G.; Gloriam, D. E.; Puthenveedu, M. A.; Babu, M. M., Combinatorial expression of GPCR isoforms affects signalling and drug responses. *Nature* **2020,** *587* (7835), 650-656.

14. Vardy, E.; Roth, Bryan L., Conformational Ensembles in GPCR Activation. *Cell* **2013,** *152* (3), 385-386.

15. Li, J.; Jonsson, A. L.; Beuming, T.; Shelley, J. C.; Voth, G. A., Ligand-Dependent Activation and Deactivation of the Human Adenosine A2A Receptor. *Journal of the American Chemical Society* **2013,** *135* (23), 8749-8759.

16. Harmar, A. J.; Fahrenkrug, J.; Gozes, I.; Laburthe, M.; May, V.; Pisegna, J. R.; Vaudry, D.; Vaudry, H.; Waschek, J. A.; Said, S. I., Pharmacology and functions of receptors for vasoactive intestinal peptide and pituitary adenylate cyclase-activating polypeptide: IUPHAR Review 1. *British Journal of Pharmacology* **2012,** *166* (1), 4-17.

17. Bortolato, A.; Doré, A. S.; Hollenstein, K.; Tehan, B. G.; Mason, J. S.; Marshall, F. H., Structure of Class B GPCRs: new horizons for drug discovery. *British Journal of Pharmacology* **2014,** *171* (13), 3132-3145.

18. Culhane, K. J.; Liu, Y.; Cai, Y.; Yan, E. C. Y., Transmembrane signal transduction by peptide hormones via family B G protein-coupled receptors. *Frontiers in Pharmacology* **2015,** *6*.

19. Boehr, D. D.; Nussinov, R.; Wright, P. E., The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology* **2009,** *5* (11), 789-796.

20. May, V.; Parsons, R. L., G Protein-Coupled Receptor Endosomal Signaling and Regulation of Neuronal Excitability and Stress Responses: Signaling Options and Lessons From the PAC1 Receptor. *Journal of Cellular Physiology* **2017,** *232* (4), 698-706.

21. Liao, C.; Zhao, X.; Brewer, M.; May, V.; Li, J., Conformational Transitions of the Pituitary Adenylate Cyclase-Activating Polypeptide Receptor, a Human Class B GPCR. *Scientific Reports* **2017,** *7* (1), 5427.

22. Liao, C.; Remington, J. M.; May, V.; Li, J., Molecular Basis of Class B GPCR Selectivity for the Neuropeptides PACAP and VIP. *Frontiers in Molecular Biosciences* **2021,** *8*.

23. Abrol, R.; Kim, S.-K.; Bray, J. K.; Trzaskowski, B.; Goddard, W. A., Chapter Two - Conformational Ensemble View of G Protein-Coupled Receptors and the Effect of Mutations and Ligand Binding. In *Methods in Enzymology*, Conn, P. M., Ed. Academic Press: 2013; Vol. 520, pp 31-48.

24. Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F., PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation* **2015,** *11* (11), 5525-5542.

25. Heyden, M.; Jardon-Valadez, H. E.; Bondar, A.-N.; Tobias, D. J., GPCR Activation on the Microsecond Timescale in MD Simulations. *Biophysical Journal* **2013,** *104* (2, Supplement 1), 115a.

26. Vilardaga, J.-P., Theme and variations on kinetics of GPCR activation/deactivation. *Journal of receptor and signal transduction research* **2010,** *30* (5), 304-312.

27. Weis, W. I.; Kobilka, B. K., The Molecular Basis of G Protein–Coupled Receptor Activation. *Annual Review of Biochemistry* **2014,** *87* (1), 897-919.

28. Bhattarai, A.; Wang, J.; Miao, Y., Retrospective ensemble docking of allosteric modulators in an adenosine G-protein-coupled receptor. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2020,** *1864* (8), 129615.

29. Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A., Computational Drug Design Accommodating Receptor Flexibility: The Relaxed Complex Scheme. *Journal of the American Chemical Society* **2002,** *124* (20), 5632-5633.

30. Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J. A.; Miao, Y.; Smith, J. C., Ensemble Docking in Drug Discovery. *Biophysical Journal* **2018,** *114* (10), 2271-2278.

31. Acharya, A.; Agarwal, R.; Baker, M. B.; Baudry, J.; Bhowmik, D.; Boehm, S.; Byler, K. G.; Chen, S. Y.; Coates, L.; Cooper, C. J.; Demerdash, O.; Daidone, I.; Eblen, J. D.; Ellingson, S.; Forli, S.; Glaser, J.; Gumbart, J. C.; Gunnels, J.; Hernandez, O.; Irle, S.; Kneller, D. W.; Kovalevsky, A.; Larkin, J.; Lawrence, T. J.; LeGrand, S.; Liu, S. H.; Mitchell, J. C.; Park, G.; Parks, J. M.; Pavlova, A.; Petridis, L.; Poole, D.; Pouchard, L.; Ramanathan, A.; Rogers, D. M.; Santos-Martins, D.; Scheinberg, A.; Sedova, A.; Shen, Y.; Smith, J. C.; Smith, M. D.; Soto, C.; Tsaris, A.; Thavappiragasam, M.; Tillack, A. F.; Vermaas, J. V.; Vuong, V. Q.; Yin, J.; Yoo, S.; Zahran, M.; Zanetti-Polzi, L., Supercomputer-Based Ensemble Docking Drug Discovery Pipeline with Application to Covid-19. *Journal of Chemical Information and Modeling* **2020,** *60* (12), 5832-5852.

32. Huang, S.-Y.; Zou, X., Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins: Structure, Function, and Bioinformatics* **2007,** *66* (2), 399-421.

33. Li, D.; Jiang, K.; Teng, D.; Wu, Z.; Li, W.; Tang, Y.; Wang, R.; Liu, G., Discovery of New Estrogen-Related Receptor α Agonists via a Combination Strategy Based on Shape Screening and Ensemble Docking. *Journal of Chemical Information and Modeling* **2022,** *62* (3), 486-497.

34. Chandak, T.; Mayginnes, J. P.; Mayes, H.; Wong, C. F., Using machine learning to improve ensemble docking for drug discovery. *Proteins: Structure, Function, and Bioinformatics* **2020,** *88* (10), 1263-1270.

35. Li, X.; Zhang, X.-X.; Lin, Y.-X.; Xu, X.-M.; Li, L.; Yang, J.-B., Virtual Screening Based on Ensemble Docking Targeting Wild-Type p53 for Anticancer Drug Discovery. *Chemistry & Biodiversity* **2019,** *16* (7), e1900170.

36. Velazquez, H. A.; Riccardi, D.; Xiao, Z.; Quarles, L. D.; Yates, C. R.; Baudry, J.; Smith, J. C., Ensemble docking to difficult targets in early-stage drug discovery: Methodology and application to fibroblast growth factor 23. *Chemical Biology & Drug Design* **2018,** *91* (2), 491-504.

37. Jukič, M.; Janežič, D.; Bren, U., Ensemble Docking Coupled to Linear Interaction Energy Calculations for Identification of Coronavirus Main Protease (3CLpro) Non-Covalent Small-Molecule Inhibitors. *Molecules* **2020,** *25* (24).

38. Patel, D.; Athar, M.; Jha, P. C., Exploring Ruthenium-Based Organometallic Inhibitors against Plasmodium falciparum Calcium Dependent Kinase 2 (PfCDPK2): A Combined Ensemble Docking, QM/MM and Molecular Dynamics Study. *ChemistrySelect* **2021,** *6* (32), 8189-8199.

39. Beebe, X.; Darczak, D.; Davis-Taber, R. A.; Uchic, M. E.; Scott, V. E.; Jarvis, M. F.; Stewart, A. O., Discovery and SAR of hydrazide antagonists of the pituitary adenylate cyclase-activating polypeptide (PACAP) receptor type 1 (PAC1-R). *Bioorganic & Medicinal Chemistry Letters* **2008,** *18* (6), 2162-2166.

40. Wu, F.; Yang, L.; Hang, K.; Laursen, M.; Wu, L.; Han, G. W.; Ren, Q.; Roed, N. K.; Lin, G.; Hanson, M. A.; Jiang, H.; Wang, M.-W.; Reedtz-Runge, S.; Song, G.; Stevens, R. C., Full-length human GLP-1 receptor structure without orthosteric ligands. *Nature Communications* **2020,** *11* (1), 1272.

41. Pinzi, L.; Rastelli, G., Molecular Docking: Shifting Paradigms in Drug Discovery. *International Journal of Molecular Sciences* **2019,** *20* (18).

42. Pantsar, T.; Poso, A., Binding Affinity via Docking: Fact and Fiction. *Molecules* **2018,** *23* (8).

43. Elokely, K. M.; Doerksen, R. J., Docking Challenge: Protein Sampling and Molecular Docking Performance. *Journal of Chemical Information and Modeling* **2013,** *53* (8), 1934-1945.

44. Mohammadi, S.; Narimani, Z.; Ashouri, M.; Firouzi, R.; Karimi‐Jafari, M. H., Ensemble learning from ensemble docking: revisiting the optimum ensemble size problem. *Scientific Reports* **2022,** *12* (1), 410.

45. Ferreira, R. S.; Simeonov, A.; Jadhav, A.; Eidam, O.; Mott, B. T.; Keiser, M. J.; McKerrow, J. H.; Maloney, D. J.; Irwin, J. J.; Shoichet, B. K., Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. *J Med Chem* **2010,** *53* (13), 4891-905.

46. Deng, N.; Forli, S.; He, P.; Perryman, A.; Wickstrom, L.; Vijayan, R. S. K.; Tiefenbrunn, T.; Stout, D.; Gallicchio, E.; Olson, A. J.; Levy, R. M., Distinguishing Binders from False Positives by Free Energy Calculations: Fragment Screening Against the Flap Site of HIV Protease. *The Journal of Physical Chemistry B* **2015,** *119* (3), 976-988.

47. Hou, X.; Rooklin, D.; Yang, D.; Liang, X.; Li, K.; Lu, J.; Wang, C.; Xiao, P.; Zhang, Y.; Sun, J.-p.; Fang, H., Computational Strategy for Bound State Structure Prediction in Structure-Based Virtual Screening: A Case Study of Protein Tyrosine Phosphatase Receptor Type O Inhibitors. *Journal of Chemical Information and Modeling* **2018,** *58* (11), 2331-2342.

48. Adeshina, Y. O.; Deeds, E. J.; Karanicolas, J., Machine learning classification can reduce false positives in structure-based virtual screening. *Proceedings of the National Academy of Sciences* **2020,** *117* (31), 18477.

49. Kufareva, I.; Abagyan, R., Methods of protein structure comparison. *Methods in molecular biology (Clifton, N.J.)* **2012,** *857*, 231-257.

50. Sterling, T.; Irwin, J. J., ZINC 15 – Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling* **2015,** *55* (11), 2324-2337.

51. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation* **2016,** *12* (1), 281-296.

52. Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L., OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Research* **2012,** *40* (D1), D370-D376.

53. Bowers, K. J.; Chow, D. E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. In *Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters*, SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, 11-17 Nov. 2006; 2006; pp 43-43.

54. Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E., Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys* **2020,** *153* (4), 044130.

55. Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W., CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation* **2016,** *12* (1), 405-413.

56. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **1994,** *101* (5), 4177-4189.

57. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R., Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics* **1995,** *103* (11), 4613-4621.

58. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995,** *103* (19), 8577-8593.

59. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996,** *14* (1), 33-38.

60. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *Journal of Medicinal Chemistry* **2006,** *49* (21), 6177-6196.