CHARMM-GUI *High-Throughput Simulator* for Efficient Evaluation of Protein-Ligand Interactions with Different Force Fields

Hugo Guterres[†1], Sang-Jun Park[†1], Han Zhang[1], Thomas Perone[1], Jongtaek Kim[2], and Wonpil Im[1]*

[1]Departments of Biological Sciences, Chemistry, Bioengineering, and Computer Science and Engineering, Lehigh University, Bethlehem, Pennsylvania 18015, USA

[2]Department of Physics and Chemistry, Korea Air Force Academy, Cheongju 363-849, Korea

[†]Both authors contributed equally to this work

***Corresponding Author:** wonpil@lehigh.edu

Abstract

Molecular docking is one of the most popular computational tools for the hit discovery step in drug design. However, there is ample room for improvement of docking's ability to identify correct binding modes and discriminate active from decoy compounds. Molecular dynamics (MD) simulations of protein-ligand docking structures have been shown to be effective in improving docking results. Here, we present CHARMM-GUI *High-Throughput Simulator* (HTS) that prepares MD simulation systems and inputs for multiple protein-ligand complex structures in a high-throughput manner. HTS supports commonly used MD programs (NAMD, GROMACS, AMBER, OpenMM, GENESIS, Desmond, LAMMPS, and Tinker) along with various force field combinations for protein and ligand, including CHARMM36m, Amber (ff19SB/ff14SB), OPLS-AA/M, CGenFF, GAFF2, and OpenFF. Validation tests using Miller and the directory of useful decoys-enhanced (DUD-E) datasets demonstrate that short MD simulations using HTS-generated systems and simple ligand RMSD calculations consistently outperform docking results. Specifically, MD simulations can better identify correct ligand binding modes among top 10 binding poses as compared to docking scores. In addition, MD simulations can better discriminate active from decoy compounds in the DUD-E dataset than docking scores for both soluble and membrane proteins. We expect that HTS can be a useful tool to facilitate the hit discovery process in drug design by improving docking results.

Keywords: molecular dynamics, docking, protein-ligand interactions, force field

1. Introduction

Over the past three decades, molecular docking has been one of the most widely used computational tools to predict protein-ligand binding interactions and binding energy due to its ease-of-use, high-speed, and scalability.[1, 2] Recently, Luttens *et al* discovered novel and effective inhibitors of SARS-CoV-2 main protease using ultra-large virtual screening docking strategies.[3] Docking calculates ligand-binding modes in a binding pocket and estimates their binding affinities. Many previous studies have shown that the correct ligand-binding modes are present in the docking results, but they are not ranked in the top scores according to their estimated binding affinities.[4-6] Therefore, additional tools are needed to identify the correct binding modes more accurately among the docking results.

Several successfully benchmarked tools to improve docking results and identify the correct ligand binding modes have been proposed over the past few years. Some of the most popular tools are machine learning (ML)-based methods, including Pafnucy and $\Delta_{vina}RF_{20}$.[7, 8] Pafnucy is a method that uses deep neural network (DNN) to estimate binding affinity of protein-ligand complexes. Pafnucy had been tested on a few benchmark sets, including the Astex diverse set, CASF-2013 (comparative assessment of scoring functions), and PDBbind 2016 set, where it consistently outperformed docking scoring functions. $\Delta_{vina}RF_{20}$ is a method that utilizes random forest algorithm to improve the classical docking scoring functions. $\Delta_{vina}RF_{20}$ achieved top performances in several competitions, including CASF-2007, CASF-2013 and CASF-2016. Despite their successful performances, recent evaluations have shown that ML-based methods tend to overfit to the mean values of samples used for training. As a result, when tested against different datasets, ML-based methods have shown poor performances.[9, 10]

Physics-based methods have also been shown by many groups to be effective in improving docking scoring functions.[11] Miller and colleagues recently presented the Schrödinger induced fit docking-molecular dynamics (IFD-MD) method, a reliable docking workflow that accurately identified the correct binding modes in over 90% of their test cases.[12] Their workflow includes ligand-based pharmacophore docking, rigid receptor docking, and explicit solvent MD simulations to solve the induced fit problem of protein-ligand binding. Rastelli and colleagues proposed a similar workflow called binding estimation after refinement (BEAR) that also included MM-PBSA/MM-GBSA (molecular mechanics with the Poisson-Boltzmann or generalized Born surface area continuum solvation) to refine and rescore the protein-ligand complex structures obtained from docking.[4] They demonstrated a significant improvement of docking results with more accurate prediction of the active compound binding modes. Recently, we showed results supporting the role of MD simulations in improving docking results.[13] Using a large set of 56 diverse protein targets and 560 ligands from the directory of useful decoys-enhanced (DUD-E), we demonstrated that MD simulations discriminated active from decoy ligands more accurately compared to docking scores.

The Schrödinger IFD-MD is not widely available to researchers in academia and the BEAR tool does not seem suitable for high-throughput processing of multiple protein-ligand complex structures. CHARMM-GUI is a user-friendly web-based tool that facilitates the preparation of various complex molecular simulation systems.[14-32] CHARMM-GUI has been widely available to researchers in academia and prevalently used for more than a decade. Here, we present CHARMM-GUI *High-Throughput Simulator* (HTS, https://www.charmm-gui.org/input/hts), an intuitive module that automatically generates multiple protein-ligand MD input files at the same time through a simple drag and drop file upload process. Through its interactive features, HTS offers ligand parameterization using 3 different force field (FF) options, including CGenFF, GAFF2, and OpenFF.[33-35] Protein FFs can be selected from CHARMM36m, Amber ff14SB, Amber ff19SB, and OPLS-AA/M.[36-39] In addition, HTS supports various MD programs, including NAMD, GROMACS, AMBER, OpenMM, GENESIS, Desmond, LAMMPS, and Tinker.[40-47] Through

several validation tests using various protein/ligand/water FF combinations, we show that MD simulations effectively discriminate correct ligand-binding modes from the incorrect ones. In addition, MD simulations can better distinguish active from decoy compounds through simple ligand RMSD calculations for both soluble and membrane proteins. These results show that CHARMM-GUI HTS is a reliable tool that can help evaluate and improve protein-ligand interactions using different FFs in a high-throughput manner.

2. Results and Discussion

2.1 Workflow of High-Throughput Simulator

The overall workflow of CHARMM-GUI HTS is shown in **Figure 1**. On the front page of HTS, a box is provided where users can drag and drop a set of protein-ligand complex structures for the same protein. HTS supports both protein in solution and membrane environments, where users can select a system type from a drop-down menu. To facilitate ligand parameterization, HTS provides 3 options. Option 1 is to upload a combination of PDB and SDF files for each protein-ligand complex file, where the SDF file corresponds to the ligand. This option provides additional information for the ligands, including protonation states and connectivity that can be valuable for ligand parameterization. Option 2 is to use ligand SMILES files from the RCSB. For this option to work properly, the associated ligand must exist in the RCSB. Option 3 is to resort to using the ligand's 3-dimensional (3D) PDB coordinates. This choice can be problematic during ligand parameterization due to the absence of connectivity, bond order, and charge information. Therefore, one needs to make sure that the resulting ligand FF parameters match their ligands.

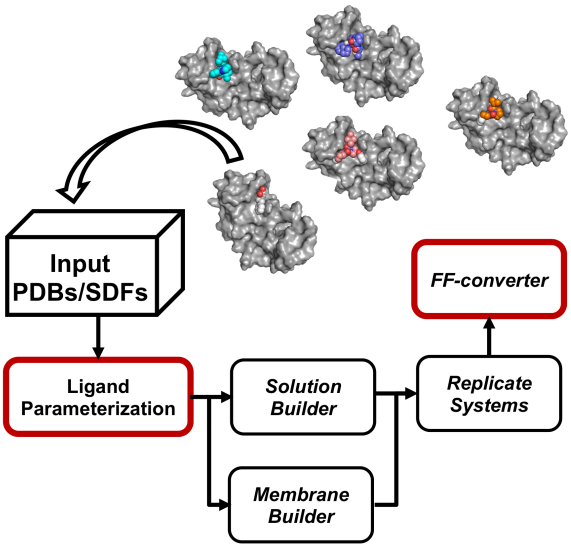

Figure 1. Workflow of CHARMM-GUI *High-Throughput Simulator*. Multiple protein-ligand complex structures can be submitted to HTS to generate MD systems and inputs simultaneously. The steps outlined in red are submitted to our computer server, where the jobs run in parallel.

The following page offers various PDB manipulation options, including adding missing residues and various post-translational modifications (such as disulfide bonds, phosphorylation, ubiquitination, and glycosylation).[17] To handle a large number of uploaded protein-ligand structures efficiently, ligand parameterizations are submitted to our computer server. To quickly

retrieve any HTS job, users can use the associated unique job ID number through the *Job Retriever* tab. To help users with practical applications of HTS jobs, we provide a video demonstration in the CHARMM-GUI website along with example sets of protein-ligand structures for solution and membrane system buildings (https://www.charmm-gui.org/demo/hts).

2.2 Ligand parameterization with 3 different FF options, including CGenFF, GAFF2, and OpenFF

The ligand parameterization page is shown in **Figure 2**. The example case includes 9 different ligands (697, 244, 041, 338, 797, 272, 4NA, 196, and I0G) bound to protein estrogen receptor-β (PDB 1U3R). These complex structures are the top binding pose results from cross docking of the ligands into the binding pocket of the receptor using AutoDock vina.[48] In the first column, each structure is renamed model_1 to model_9 and their corresponding file names are shown in column 2. The ligand names are listed on column 3, where one can display the 3D visualization of any ligand using NGL viewer by clicking the ligand's name (**Figure 2**).[49] Upon inspection, if a ligand is incorrectly parameterized, one can modify and correct structural errors using the structure modification button. This feature opens a 2D sketchpad editor window that is powered by MarvinJS (**Figure 2**).[24] The next 3 columns contain 3 different FFs (CGenFF, GAFF2, and OpenFF) that are used to parameterize each ligand. If a ligand is successfully parameterized, a green check mark is displayed. Otherwise, a red x-mark indicates that a specific ligand parameterization step fails. One way to solve this issue is to modify the ligand using the structure modification button and regenerate parameters for a specific FF using the drop-down menu above the table (**Figure 2**). One can select one of the FF options in the last column for MD system preparation.

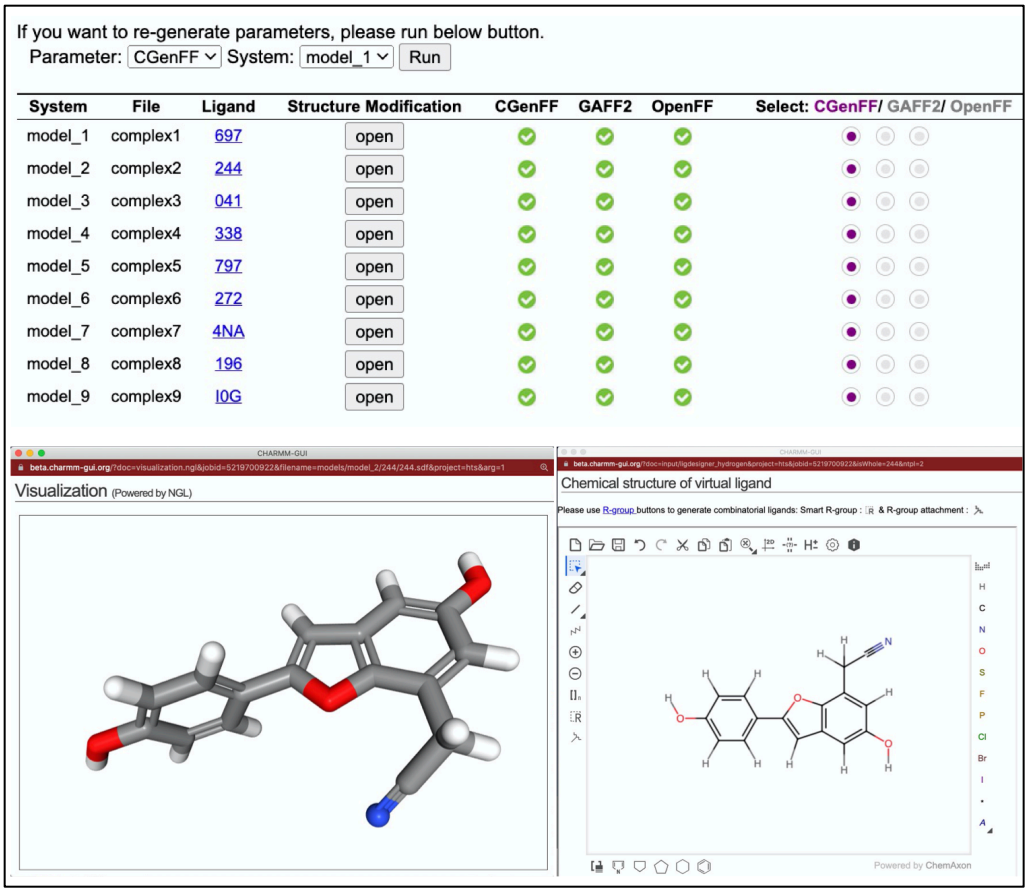

Figure 2. CHARMM-GUI *High-Throughput Simulator* ligand FF parameterization page. The table contains systems of protein-ligand complex structures, where the ligand IDs are shown in column 3. By clicking the ligand ID, a pop-up window (shown on the lower left side) appears, displaying the 3D structure of the ligand in sticks representation. To modify a ligand with any incorrect bond order or charges, one can select the *open* button from the structure modification column. It brings up a 2D sketchpad window (shown on the lower right side) that can be used for modification and redrawing of the existing ligand. When the ligand is successfully parameterized, a green check mark is displayed under the specific FF names (CGenFF, GAFF2, and OpenFF).

2.3 Input generation supports various protein FFs and MD simulation programs

Upon selecting a specific ligand FF, the protein-ligand systems are prepared following the standard procedures in *Solution Builder* or *Membrane Builder*.[14, 15, 18, 19, 21, 50] For solution systems, users have the options to select the system box size using TIP3P water, specify the types of ions, and select an ion placing method for explicit solvent simulations. For membrane bilayer systems, options to orient the protein in the membrane and choices of membrane components are provided. Our workflow generates one system using the protein-ligand complex structure with the biggest ligand and the system is replicated for all other protein-ligand structures (**Figure 1**). Water molecules that overlap with the protein, ligand, and ions are removed. Because ligands can have different charges, each system is then neutralized by deleting ions depending on specific ligand charges in the system. When replacing the biggest ligand with the smaller ones, a small empty space without water might be created in the binding pocket. In most solvent-exposed binding pockets, this space will be hydrated during minimization and equilibration steps of MD simulations. However, users should be careful in cases with buried binding pockets. There are a few different methods to place waters around ligands in buried binding pockets, including Placevent, Rosetta-ECO, grand canonical Monte Carlo (MC) simulation, MC/MD, and Hamiltonian simulated annealing of solvent (HSAS).[51-55] In the near future, we plan to implement one of these methods to properly hydrate the ligands in buried binding pockets. The next step is to generate simulation input files, where this job is submitted to our compute server to handle an expected large number of system preparations efficiently. The result page contains drop-down menu options for protein FFs (CHARMM36m, Amber ff14SB, Amber ff19SB, and OPLS-AA/M) and various simulation programs, including NAMD, GROMACS, AMBER, OpenMM, GENESIS, Desmond, LAMMPS, and Tinker.[36-47] In addition, the hydrogen mass repartitioning (HMR) option is available to further accelerate MD simulations with 4 fs timestep.[27, 56] Note that CHARMM-GUI HTS prepares all the necessary input files to run MD simulations for multiple protein-ligand complex structures, but it does not perform actual MD simulations for users.

2.4 MD simulations using various FF combinations can effectively discriminate good from bad ligand-binding modes

Many studies have shown that although molecular docking can generate ligand-binding modes found in experimental protein-ligand complex structures, the correct binding mode is not often top-scored among a list of predicted binding modes.[4-6] Therefore, additional evaluation method to efficiently select a correct binding mode is still needed. For our first benchmark test, 208 predicted protein-ligand complex structures from docking (based on the Miller dataset in **Table S1**) were evaluated using short MD simulations. Docking results were separated into two categories according to their ligand RMSD relative to the corresponding crystal structure: good ligand-binding modes having ligand RMSD $\leq$ 5 Å and bad modes having ligand RMSD > 5 Å. This is based on our previous work, where we showed that only the good ligand-binding modes could improve toward crystal structure poses during short MD simulations.[13] Out of 208 protein-ligand structures,

147 fall into the category of bad ligand-binding modes and 61 structures are good binding modes (**Table S1**).

The capability to discriminate good from bad ligand-binding modes among its predicted protein-ligand structures can be evaluated by a receiver operating curve, area under the curve (ROC AUC) (**Figure S1**). Our benchmark tests using HTS-based MD simulations show that, compared to the AutoDock vina docking scoring function, this ROC AUC value is consistently improved with various protein/ligand/water FF combinations (**Table 1**). Our assessment using different FF combinations can inform researchers on best practices that can be applied quickly in running MD simulations of protein-ligand complex structures. The different FFs contain various mathematical approximations to reproduce molecular geometry and properties of biomolecular structures. Therefore, their MD simulation results can vary depending on which FF combination is used. Our results show that CHARMM36m/GAFF2/TIP3P FF combination show the best performance in discriminating bad protein-ligand interactions from the good ones with an ROC AUC of 0.84. Other combinations with the CHARMM protein FF show similar results (0.80 with CGenFF and 0.81 with OpenFF), indicating that the GAFF2 ligand FF shows a slightly better discrimination power than CGenFF and OpenFF FFs. Using the Amber FF combinations, we find that ff19SB shows slightly a better performance than ff14SB. Overall, all the different FF combinations can improve the ROC AUC of docking scores from 0.61 to up to 0.84.

Docking scores are rough estimates of binding energies in kcal/mol that are commonly used by researchers for the predicted protein-ligand structures. Similarly, MD simulation trajectories of protein-ligand complex structures can be used to obtain binding energy estimates using MMGBSA/MMPBSA. These binding energies in kcal/mol were calculated for all of our Amber simulation trajectories. The values are used to rescore the docking results. MMGBSA/MMPBSA binding affinity results were evaluated using the ROC AUC to assess their ability in discriminating good from bad docking poses. In **Table 1**, we show that MMGBSA/MMPBSA have higher discriminating power than docking scores, based on their AUC values. MMGBSA/MMPBSA binding affinities show similar AUC values compared to ligand RMSD, where more stable binding poses with lower ligand RMSD consistently have better binding affinity values than unstable poses with high ligand RMSD. Using 6 different FF combinations, our results show that MMPBSA has an average AUC of 0.79, which is slightly better than 0.77 using MMGBSA calculations.

A representative example is shown in **Figure 3** with top 10 ligand-binding modes from docking of ligand Y27 to the receptor Rho kinase. Docking scores incorrectly sorts the top 10 binding poses by highly scoring binding modes with the pyridine moiety (yellow in **Figure 3**) outside of the binding pocket and facing the solvent. The correct binding mode with a ligand RMSD of 2.1 Å from the crystal structure is ranked as pose 9 with a docking score of -6.1 kcal/mol. Interestingly, this binding mode shows the most stable binding throughout MD simulation with an average ligand RMSD of 1.6 Å. It also has the best MMGBSA and MMPBSA binding energy values of -25 and -19.4 kcal/mol, respectively. Consistent with ligand-binding stability during MD simulation, MMGBSA and MMPBSA show the highest binding energies of -15.1 and -8.6 kcal/mol, respectively, for pose 2 that has the highest ligand RMSD of 7.2 Å during MD simulation.

Table 1. Summary of AUC ROC for 208 protein-ligand structures in the Miller dataset showing docking scores from AutoDock vina, ligand RMSD and MMGBSA/ MMPBSA results from MD simulation with various FF combinations

No	FF combinations of protein/ligand/water	ROC AUC			
		Docking	Ligand RMSD	MMGBSA	MMPBSA
1	CHARMM36m/CGenFF/TIP3P		0.80	-	-
2	CHARMM36m/GAFF2/TIP3P		0.84	-	-
3	CHARMM36m/OpenFF/TIP3P		0.81	-	-
4	ff14SB/GAFF2/TIP3P		0.76	0.76	0.76
5	ff19SB/GAFF2/TIP3P	0.61	0.77	0.79	0.79
6	ff19SB/GAFF2/OPC		0.80	0.77	0.81
7	ff14SB/OpenFF/TIP3P		0.78	0.77	0.77
8	ff19SB/OpenFF/TIP3P		0.79	0.77	0.80
9	ff19SB/OpenFF/OPC		0.78	0.78	0.80
	average	*0.61*	*0.79*	*0.77*	*0.79*

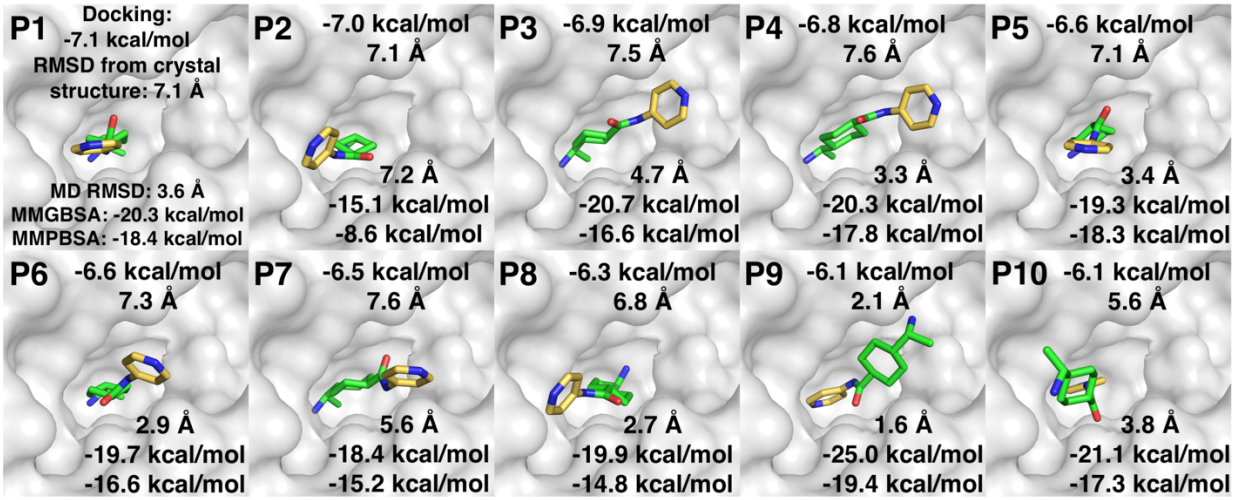

Figure 3. Ligand-binding modes from top 10 AutoDock vina cross-docking results of ligand Y27 (PDB 2ETR) onto the receptor Rho-associated protein kinase (PDB 2ESM). The binding poses were ranked based on the vina docking score from pose 1 (P1) to P10. Their corresponding ligand RMSD from the crystal structure is shown below their docking score. In addition, the ligand RMSD and MMGBSA/MMPBSA binding affinity values from their MD simulations (ff19SB/GAFF2/OPC) are shown for each pose. P9 is highlighted in red square because it has the most similar binding pose compared to the crystal structure with a RMSD of 2.1 Å. In addition, it has the most stable binding in MD simulation with an average ligand RMSD of 1.6 Å and the lowest binding energies of -25.0 kcal/mol (MMGBSA) and -19.4 kcal/mol (MMPBSA).

2.5 MD simulations can distinguish active from decoy binders using ligand RMSD evaluation

In our previous work, we evaluated the performance of MD simulation in discriminating active from decoy ligands from the DUD-E dataset using only 56 targets having a crystal structure with

a resolution of 2 Å or better.[13] For our second benchmark test, we performed the same tests on the rest of protein targets in the dataset. From 41 soluble proteins, we randomly selected 5 active and 5 decoy ligands (**Table S2**).

Docking results from 205 active and 205 decoy ligands on 41 target proteins show an ROC AUC of 0.68 (**Figure 4C**). A lot of overlap between active and decoy ligands' docking scores suggest that it is difficult to distinguish between the two classes of ligands (**Figure 4A**). Here, only one protein-ligand pair with the best docking score is selected as a starting structure for MD simulation. Using HTS, MD systems were prepared for all selected docking structures. MD simulations were conducted for 50 ns and processed by calculating their ligand RMSD to assess ligand-binding stability. The ligand RMSD results significantly improve the ROC AUC to 0.83 (**Figure 4C**). The ligand RMSD histogram distribution shows that active ligands have better binding stability with lower ligand RMSD centering around 4 Å. Whereas, decoy ligands show binding instability with many ligands leaving their binding pockets (**Figure 4B**). Improvements in ROC AUC using ligand RMSD from MD simulations are seen consistently across 5 different protein classes, including kinase, protease, other enzymes, nuclear receptor, and miscellaneous proteins (**Table 2**).

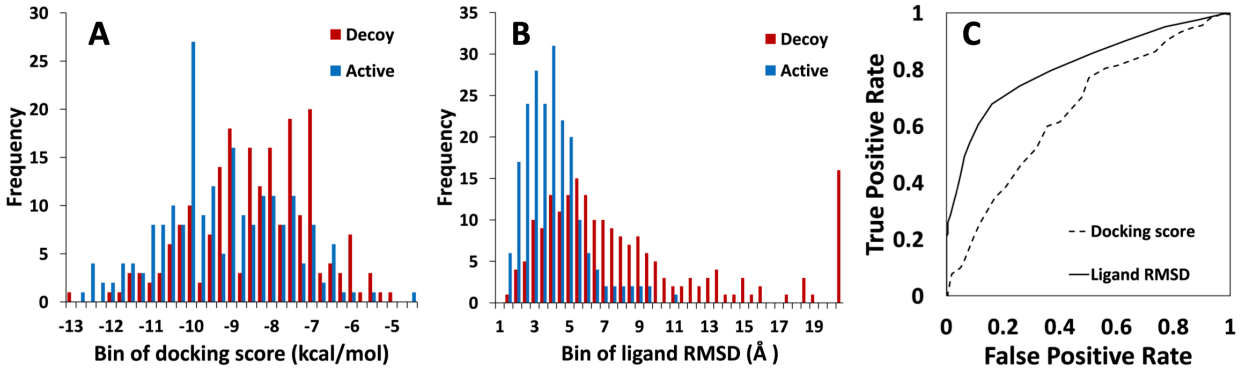

Figure 4. Histogram distributions of active and decoy ligand scores obtained from (A) docking scores and (B) ligand RMSD. (C) ROC plot comparing docking scores with ligand RMDSs from MD simulations of 41 soluble proteins with randomly selected 5 active and 5 decoy ligands for each protein. The AUC values are 0.68 (docking score) and 0.83 (MD ligand RMSD).

Table 2. Summary of AUC ROC for 41 soluble proteins in the DUD-E dataset showing docking scores from AutoDock vina and ligand RMSD results from MD simulation.[†]

		total ligands		AUC	
protein class	total proteins	actives	decoys	docking	MD
total	41	205	205	0.68	**0.83**
kinase	12	60	60	0.73	**0.85**
protease	5	25	25	0.60	**0.87**
other enzymes	18	90	90	0.66	**0.87**
nuclear receptor	5	25	25	0.76	**0.78**
miscellaneous	1	5	5	0.96	0.96

[†]*Improved cases of AUC values using MD simulations are highlighted in bold*

2.6 Membrane protein MD simulations significantly improve docking results

G protein-coupled receptors (GPCRs) are a class of drug targets with a very high pharmacological precedence. GPCRs make up more than one third of all United States Food and Drug Administration (FDA) approved drugs.[57] Therefore, we chose to work with GPCR proteins as benchmark validation for our HTS membrane-bilayer systems. Here, 5 GPCR proteins from the DUD-E dataset were tested with randomly chosen 5 active and 5 decoy ligands.[58] Their docking scores show poor results in distinguishing active from decoy compounds with an ROC AUC of 0.57 (**Figure 5C**). An almost uniform distribution of active and decoy ligands is seen using the docking scoring functions (**Figure 5A**), indicating that docking scores are unable to discriminate real binders from nonbinders to GPCR proteins. Using the membrane-bilayer option in CHARMM-GUI HTS, MD systems for 5 GPCR proteins were prepared. Our results show that ligand RMSD calculations from MD simulations significantly improve the ROC AUC value to 0.85 (**Figure 5C**). Active ligands show better binding stability with lower ligand RMSD that centers around 4 Å, whereas decoy ligands tend to move more with higher ligand RMSD that centers around 7 Å (**Figure 5B**). The binding pockets of GPCR proteins are narrow and partially occluded by the extracellular loop 2 (ECL2).[59] We observed that during the short MD simulations, unstable ligands do not completely leave their binding pockets due to interactions with the ECL2.

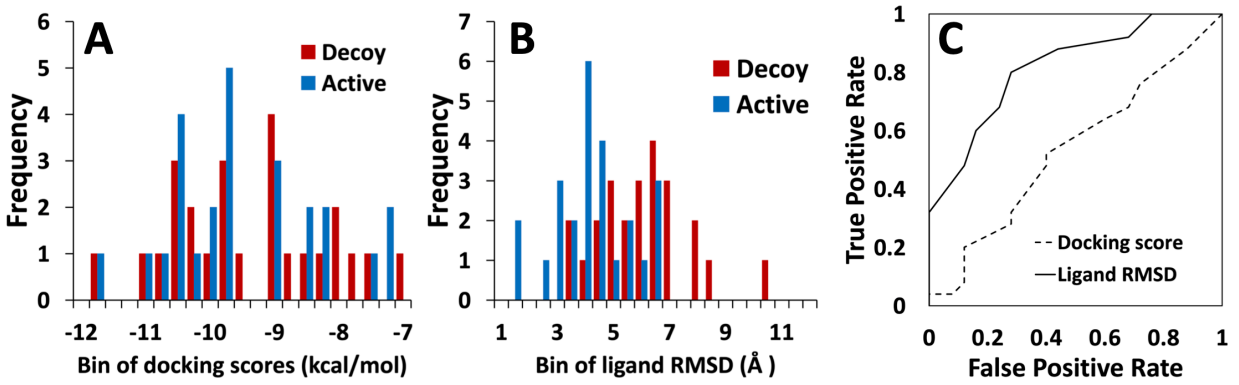

Figure 5. Histogram distributions for GPCR proteins with their active and decoy ligand scores obtained from (A) docking and (B) ligand RMSD. (C) ROC plot comparing docking scoring with ligand RMSD from MD simulations of 5 GPCR proteins in membrane bilayers with randomly selected 5 active and 5 decoy ligands for each protein. The AUC values are 0.57 (docking score) and 0.85 (ligand RMSD).

3. Conclusions

Hit discovery step in structure-based drug design has long been dominated by molecular docking-based methods due to its low cost, high-speed, and scalability. However, docking scoring functions have been shown to be unreliable in distinguishing active from decoy compounds and identifying correct binding modes among multiple poses.[4-6, 13] In this work, we have introduced CHARMM-GUI *High-Throughput Simulator*, a new functional module to facilitate high-throughput preparation of MD simulation systems and inputs for multiple protein-ligand structures simultaneously. We expect that HTS can be used to evaluate and further improve docking results by better discriminating active from decoy compounds and more accurately identify the correct binding modes of any specific active compounds. For example, top 1% of virtual screening results using a large 10,000 ligand library can be evaluated by performing short MD simulations using

CHARMM-GUI HTS to find the real binders. Furthermore, ligand binding modes can be refined by evaluating multiple binding poses using MD simulations.

Our benchmark tests on the Miller and DUD-E targets demonstrate the effectiveness of short MD simulations to evaluate and improve docking results. In the cross-docking test cases using the Miller dataset, we show that MD simulations can better identify correct binding modes among top 10 poses compared to docking scoring. In addition, we confirmed these results using 6 different protein/ligand/water FF combinations. Using the DUD-E dataset, we show that MD simulations can better discriminate active from decoy compounds for 46 targets, including 5 GPCR membrane proteins.

CHARMM-GUI HTS is user-friendly and contains many interactive features to help with ligand parameterization, FF combination options for protein/ligand/water, as well as functionalities to seamlessly prepare protein-ligand systems in solutions and membrane bilayers. HTS effectively circumvents the redundant process of individually preparing multiple protein-ligand complex structures by running the process simultaneously. We expect HTS to facilitate the hit discovery process in drug design by properly evaluating and improving docking results.

4. Methods

4.1 Datasets and computational details

For the benchmark tests, we selected representative 208 protein-ligand structures from the Miller cross-docking test set (**Table S1**) and 460 structures from the DUD-E dataset (**Tables S2-S3**).[12, 58] Using the Miller dataset, we evaluated the performance of MD simulations using various FF combinations in discriminating correct binding modes from the incorrect ones among up to top 10 docking results. Using the DUD-E dataset, we evaluated MD simulation performance in discriminating active from decoy small molecule binders (for 410 soluble and 50 membrane protein-ligand complexes). Because most ligands in the DUD-E dataset are similar, we filtered the selected ligands using their Tanimoto coefficients (Tc) to avoid redundancy. Only ligands with Tc < 0.5 were selected and their Tc distribution is shown in **Figures S2-S3**.

AutoDock vina has been shown to be one of the leading docking methods that is freely available to researchers in academia.[48, 60] Using vina, protein receptors were treated rigid and the ligands were docked into each receptor's binding pocket. Since the benchmark targets are holo protein structures with bound ligands, their native ligand coordinates were used to determine the search space for ligand docking. Each docking was performed using a cubic box search space with 22.5 Å edges. For the Miller dataset, up to top 10 binding poses were collected for each ligand docking. However, 3 targets could not produce 10 binding poses due to the limited size of the binding pocket search space for ligand docking (**Table S1**). For example, target 7, nuclear receptor protein (PDB 1x76), only produced 3 binding poses, because it has a small and closed binding pocket that only allowed for up to 3 poses of the docked ligand. For the DUD-E dataset, only the top scored docking output was selected for each protein-ligand complex.

Using the system and input files generated from CHARMM-GUI HTS, MD simulations were performed for 50 ns for each protein-ligand complex structure. For the Miller dataset, 9 different protein/ligand/water FF combinations were prepared: CHARMM36m/CGenFF/TIP3P, CHARMM36m/GAFF2/TIP3P, CHARMM36m/OpenFF/TIP3P, ff14SB/GAFF2/TIP3P, ff19SB/GAFF2/TIP3P, ff19SB/GAFF2/OPC, ff14SB/OpenFF/TIP3P, ff19SB/OpenFF/TIP3P, and ff19SB/OpenFF/OPC. For the DUD-E dataset, the CHARMM36m/CGenFF/TIP3P FF combination was used. Each soluble protein-ligand structure was solvated in a cubic water box extending at least 10 Å in each direction of the protein. Distance-based ion placements were used

with K^+ and Cl^- ions at 0.15 M to neutralize each system. The non-bonded van der Waals interactions were truncated between 10 and 12 Å using a force-based switching method for the CHARMM FF and at 10 Å for the Amber FFs.[61] Particle-mesh Ewald summation was used for the long-range electrostatic interactions.[62] All simulations were performed using hydrogen mass repartitioning (HMR) to increase the simulation timestep to 4 fs during the production runs.[27, 56] Each system was minimized for 5,000 steps using the steepest descent method followed by 125 ps equilibration in the NVT (constant particle number, volume, and temperature) ensemble. The GPCR membrane systems (50 structures) were placed in a POPC bilayer, solvated in water, and neutralized using K+ and Cl- ions. These systems were minimized and underwent successive equilibration steps following the default CHARMM-GUI *Membrane Builder* equilibration protocol.[15, 50]

The production runs for all systems were performed in the NPT (constant particle number, pressure, and temperature) for 50 ns in 3 replicas using different initial velocities at 303.15 K and 1 bar. Simulations with CHARMM36m protein FFs were all performed using the OpenMM package.[43] Simulations with the Amber ff14SB and ff19SB protein FFs were performed using the Amber package.[42]

Our simulations of proteins in solutions were conducted using 1 GTX 1080TI GPU (graphics processing unit) and 1 CPU (central processing unit). The simulation time for the smallest system with 123 residues was 8.1 hours (OpenMM) or 3.4 hours (Amber) for 50 ns (**Table S4**). The largest system with 728 residues required 27.1 hours (OpenMM) or 11.3 hours (Amber) for 50 ns. MD simulation of membrane systems were conducted using 1 RTX 2080TI GPU (graphics processing unit) and 1 CPU (core processing unit). The smallest system with 313 residues ran for 7.9 hours for 50 ns (OpenMM), while the largest system with 502 residues ran for 10.2 hours for 50 ns (OpenMM).

4.2 RMSD / MMGBSA / MMPBSA calculation details

Binding stability of each ligand in the protein binding pocket was evaluated using ligand RMSD by superimposing all heavy atom coordinates of the protein structure throughout MD trajectory and calculating ligand RMSD using CHARMM for simulations with CHARMM FF and using CPPTRAJ for simulations with Amber FFs.[63, 64] This method was effective in capturing RMSD from each ligand's translation and rotation with respect to the binding pocket during MD simulation. A single ligand RMSD value was obtained for each protein-ligand simulation by calculating the average ligand RMSD throughout the simulation time. Furthermore, an average value is obtained from 3 replicas of the same system.

In MM/PBSA or MM/GBSA, the ligand binding free energy (ΔG_{bind}) was calculated using the equation:

$$\Delta G_{bind} = G_{complex} - G_{receptor} - G_{ligand}$$

Here, ΔG_{bind} can be decomposed into different energy terms:

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S$$
$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{elec} + \Delta E_{vdw}$$
$$\Delta G_{sol} = \Delta G_{PB/GB} + \Delta G_{Surf}$$

ΔE_{MM}, ΔG_{sol}, and -$T\Delta S$ are the changes in the gas phase molecular mechanics (MM) energy, solvation free energy, and conformational entropy upon binding, respectively. The solvation free

energy, ΔG_{sol}, is the sum of the nonpolar energy (ΔG_{Surf}) and polar ($\Delta G_{PB/GB}$) terms. The nonpolar energy was estimated using the solvent-accessible surface area (SASA), while the polar contribution was calculated using GB or PB model. In our PB calculations, the values of 80 and 1 were used for the solvent dielectric constant and protein interior dielectric constant, respectively. The default solvent dielectric constant of 78.5 and protein interior dielectric constant of 1 were applied in the GB calculations. OBC model (igb=5) was applied in GB calculations.[65] The conformational entropy change ($-T\Delta S$) was neglected in this study. The ligand binding free energy, therefore, was the sum of an electrostatic term (ΔE_{elec}), a van der Waals term (ΔE_{vdw}), a nonpolar solvation term, and a GB or PB polar solvation term. A total of 100 snapshots extracted from 50 ns simulations were used to estimate ΔG_{bind}.

5. General statement

Hit discovery is a vital step in drug design to find novel molecules that bind to a biological target. By incorporating a physics-based MD simulation using our HTS tool after molecular docking, we are able to show considerable improvements in docking results to identify correct ligand-binding modes and discriminate active from decoy compounds. Our tool to prepare MD simulation systems and inputs for multiple protein-ligand complex structures in a high-throughput manner is freely available in CHARMM-GUI (https://www.charmm-gui.org/input/hts).

Acknowledgements

This work has been supported by NIH GM126140 and GM138472, and NSF DBI-1660380.

Author Contributions

Hugo Guterres: Data curation; formal analysis; investigation; validation; writing original draft. *Sang-Jun Park*: Methodology; investigation; web server implementation; writing-review and editing. *Han Zhang*: Amber simulations and analysis. *Thomas Perone*: Portions of Miller dataset simulations and analysis. *Jongtaek Kim*: Portions of Miller dataset simulations and analysis. *Wonpil Im*: Conceptualization; funding acquisition; investigation; formal analysis; supervision; writing-review and editing.

Conflict of Interest

W.I. is the co-founder and CEO of MolCube INC.

ORCID

Hugo Guterres, ORCID: 0000-0001-7683-8385

Sang-Jun Park, ORCID: 0000-0002-7307-3724

Han Zhang, ORCID: 0000-0002-5608-9185

Thomas Perone, ORCID: 0000-0001-7941-6954

Jongtaek Kim, ORCID ID: 0000-0002-9821-284X

Wonpil Im, ORCID: 0000-0001-5642-6041

Supporting Information

All supporting information may be found online in the Supporting Information section located at the end of the article.

References

[1] Li, Q., and Shah, S. (2017) Structure-Based Virtual Screening, *Methods Mol Biol 1558*, 111-124.

[2] Lionta, E., Spyrou, G., Vassilatis, D. K., and Cournia, Z. (2014) Structure-based virtual screening for drug discovery: principles, applications and recent advances, *Curr Top Med Chem 14*, 1923-1938.

[3] Luttens, A., Gullberg, H., Abdurakhmanov, E., Vo, D. D., Akaberi, D., Talibov, V. O., Nekhotiaeva, N., Vangeel, L., De Jonghe, S., Jochmans, D., Krambrich, J., Tas, A., Lundgren, B., Gravenfors, Y., Craig, A. J., Atilaw, Y., Sandstrom, A., Moodie, L. W. K., Lundkvist, A., van Hemert, M. J., Neyts, J., Lennerstrand, J., Kihlberg, J., Sandberg, K., Danielson, U. H., and Carlsson, J. (2022) Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses, *J Am Chem Soc 144*, 2905-2920.

[4] Rastelli, G., Degliesposti, G., Del Rio, A., and Sgobba, M. (2009) Binding estimation after refinement, a new automated procedure for the refinement and rescoring of docked ligands in virtual screening, *Chem Biol Drug Des 73*, 283-286.

[5] Ramirez, D., and Caballero, J. (2018) Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data?, *Molecules 23*.

[6] Wagner, J. R., Churas, C. P., Liu, S., Swift, R. V., Chiu, M., Shao, C., Feher, V. A., Burley, S. K., Gilson, M. K., and Amaro, R. E. (2019) Continuous Evaluation of Ligand Protein Predictions: A Weekly Community Challenge for Drug Docking, *Structure 27*, 1326-1335 e1324.

[7] Stepniewska-Dziubinska, M. M., Zielenkiewicz, P., and Siedlecki, P. (2018) Development and evaluation of a deep learning model for protein-ligand binding affinity prediction, *Bioinformatics 34*, 3666-3674.

[8] Wang, C., and Zhang, Y. (2017) Improving scoring-docking-screening powers of protein-ligand scoring functions using random forest, *J Comput Chem 38*, 169-177.

[9] Tran-Nguyen, V. K., Bret, G., and Rognan, D. (2021) True Accuracy of Fast Scoring Functions to Predict High-Throughput Screening Data from Docking Poses: The Simpler the Better, *J Chem Inf Model 61*, 2788-2797.

[10] Chen, L., Cruz, A., Ramsey, S., Dickson, C. J., Duca, J. S., Hornak, V., Koes, D. R., and Kurtzman, T. (2019) Hidden bias in the DUD-E dataset leads to misleading performance of deep learning in structure-based virtual screening, *PLoS One 14*, e0220113.

[11] De Vivo, M., Masetti, M., Bottegoni, G., and Cavalli, A. (2016) Role of Molecular Dynamics and Related Methods in Drug Discovery, *J Med Chem 59*, 4035-4061.

[12] Miller, E. B., Murphy, R. B., Sindhikara, D., Borrelli, K. W., Grisewood, M. J., Ranalli, F., Dixon, S. L., Jerome, S., Boyles, N. A., Day, T., Ghanakota, P., Mondal, S., Rafi, S. B., Troast, D. M., Abel, R., and Friesner, R. A. (2021) Reliable and Accurate Solution to the Induced Fit Docking Problem for Protein-Ligand Binding, *J Chem Theory Comput 17*, 2630-2639.

[13] Guterres, H., and Im, W. (2020) Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations, *J Chem Inf Model 60*, 2189-2198.

[14] Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM, *J Comput Chem 29*, 1859-1865.

[15] Jo, S., Lim, J. B., Klauda, J. B., and Im, W. (2009) CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes, *Biophys J 97*, 50-58.

[16] Cheng, X., Jo, S., Lee, H. S., Klauda, J. B., and Im, W. (2013) CHARMM-GUI micelle builder for pure/mixed micelle and protein/micelle complex systems, *J Chem Inf Model 53*, 2171-2180.

[17] Jo, S., Cheng, X., Islam, S. M., Huang, L., Rui, H., Zhu, A., Lee, H. S., Qi, Y., Han, W., Vanommeslaeghe, K., MacKerell, A. D., Jr., Roux, B., and Im, W. (2014) CHARMM-GUI PDB manipulator for advanced modeling and simulations of proteins containing nonstandard residues, *Adv Protein Chem Struct Biol 96*, 235-265.
[18] Wu, E. L., Cheng, X., Jo, S., Rui, H., Song, K. C., Davila-Contreras, E. M., Qi, Y., Lee, J., Monje-Galvan, V., Venable, R. M., Klauda, J. B., and Im, W. (2014) CHARMM-GUI Membrane Builder toward realistic biological membrane simulations, *J Comput Chem 35*, 1997-2004.
[19] 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., 3rd, MacKerell, A. D., Jr., Klauda, J. B., and Im, W. (2016) CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field, *J Chem Theory Comput 12*, 405-413.
[20] Kim, S., Lee, J., Jo, S., Brooks, C. L., 3rd, Lee, H. S., and Im, W. (2017) CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules, *J Comput Chem 38*, 1879-1886.
[21] Lee, J., Patel, D. S., Stahle, J., Park, S. J., Kern, N. R., Kim, S., Lee, J., Cheng, X., Valvano, M. A., Holst, O., Knirel, Y. A., Qi, Y., Jo, S., Klauda, J. B., Widmalm, G., and Im, W. (2019) CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans, *J Chem Theory Comput 15*, 775-786.
[22] Park, S. J., Lee, J., Qi, Y., Kern, N. R., Lee, H. S., Jo, S., Joung, I., Joo, K., Lee, J., and Im, W. (2019) CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates, *Glycobiology 29*, 320-331.
[23] Qi, Y., Lee, J., Klauda, J. B., and Im, W. (2019) CHARMM-GUI Nanodisc Builder for modeling and simulation of various nanodisc systems, *J Comput Chem 40*, 893-899.
[24] Kim, S., Oshima, H., Zhang, H., Kern, N. R., Re, S., Lee, J., Roux, B., Sugita, Y., Jiang, W., and Im, W. (2020) CHARMM-GUI Free Energy Calculator for Absolute and Relative Ligand Solvation and Binding Free Energy Simulations, *J Chem Theory Comput 16*, 7207-7218.
[25] Lee, J., Hitzenberger, M., Rieger, M., Kern, N. R., Zacharias, M., and Im, W. (2020) CHARMM-GUI supports the Amber force fields, *J Chem Phys 153*, 035103.
[26] Choi, Y. K., Park, S. J., Park, S., Kim, S., Kern, N. R., Lee, J., and Im, W. (2021) CHARMM-GUI Polymer Builder for Modeling and Simulation of Synthetic Polymers, *J Chem Theory Comput 17*, 2431-2443.
[27] Gao, Y., Lee, J., Smith, I. P. S., Lee, H., Kim, S., Qi, Y., Klauda, J. B., Widmalm, G., Khalid, S., and Im, W. (2021) CHARMM-GUI Supports Hydrogen Mass Repartitioning and Different Protonation States of Phosphates in Lipopolysaccharides, *J Chem Inf Model 61*, 831-839.
[28] Guterres, H., Park, S. J., Cao, Y., and Im, W. (2021) CHARMM-GUI Ligand Designer for Template-Based Virtual Ligand Design in a Binding Site, *J Chem Inf Model 61*, 5336-5342.
[29] Guterres, H., Park, S. J., Zhang, H., and Im, W. (2021) CHARMM-GUI LBS Finder & Refiner for Ligand Binding Site Prediction and Refinement, *J Chem Inf Model*.
[30] Park, S., Choi, Y. K., Kim, S., Lee, J., and Im, W. (2021) CHARMM-GUI Membrane Builder for Lipid Nanoparticles with Ionizable Cationic Lipids and PEGylated Lipids, *J Chem Inf Model 61*, 5192-5202.
[31] Zhang, H., Kim, S., Giese, T. J., Lee, T. S., Lee, J., York, D. M., and Im, W. (2021) CHARMM-GUI Free Energy Calculator for Practical Ligand Binding Free Energy Simulations with AMBER, *J Chem Inf Model*.
[32] Choi, Y. K., Kern, N. R., Kim, S., Kanhaiya, K., Afshar, Y., Jeon, S. H., Jo, S., Brooks, B. R., Lee, J., Tadmor, E. B., Heinz, H., and Im, W. (2022) CHARMM-GUI Nanomaterial Modeler

for Modeling and Simulation of Nanomaterial Systems, *J Chem Theory Comput 18*, 479-493.
[33] Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and Mackerell, A. D., Jr. (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, *J Comput Chem 31*, 671-690.
[34] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field, *J Comput Chem 25*, 1157-1174.
[35] Qiu, Y., Smith, D. G. A., Boothroyd, S., Jang, H., Hahn, D. F., Wagner, J., Bannan, C. C., Gokey, T., Lim, V. T., Stern, C. D., Rizzi, A., Tjanaka, B., Tresadern, G., Lucas, X., Shirts, M. R., Gilson, M. K., Chodera, J. D., Bayly, C. I., Mobley, D. L., and Wang, L. P. (2021) Development and Benchmarking of Open Force Field v1.0.0-the Parsley Small-Molecule Force Field, *J Chem Theory Comput 17*, 6262-6280.
[36] Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmuller, H., and MacKerell, A. D., Jr. (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins, *Nat Methods 14*, 71-73.
[37] Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *J Chem Theory Comput 11*, 3696-3713.
[38] Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Migues, A. N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., and Simmerling, C. (2020) ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution, *J Chem Theory Comput 16*, 528-552.
[39] Robertson, M. J., Tirado-Rives, J., and Jorgensen, W. L. (2015) Improved Peptide and Protein Torsional Energetics with the OPLSAA Force Field, *J Chem Theory Comput 11*, 3499-3509.
[40] Phillips, J. C., Hardy, D. J., Maia, J. D. C., Stone, J. E., Ribeiro, J. V., Bernardi, R. C., Buch, R., Fiorin, G., Henin, 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., Kale, L. V., Schulten, K., Chipot, C., and Tajkhorshid, E. (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD, *J Chem Phys 153*, 044130.
[41] Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation, *J Chem Theory Comput 4*, 435-447.
[42] Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *J Comput Chem 26*, 1668-1688.
[43] Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L. P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks, B. R., and Pande, V. S. (2017) OpenMM 7: Rapid development of high performance algorithms for molecular dynamics, *PLoS Comput Biol 13*, e1005659.
[44] Jung, J., Mori, T., Kobayashi, C., Matsunaga, Y., Yoda, T., Feig, M., and Sugita, Y. (2015) GENESIS: a hybrid-parallel and multi-scale molecular dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations, *Wiley Interdiscip Rev Comput Mol Sci 5*, 310-323.
[45] Bowers KJ, S. F., Salmon JK, Shan Y, Shaw DE, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepsis JL, Kolossvary I, Moraes MA. (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters, In *ACM/IEEE conference on supercomputing*, IEEE, Tampa, FL, USA.
[46] Plimpton, S. J. (1995) Fast parallel algorithms for short-range molecular dynamics, *J Comput Phys 117*, 1-19.

[47] Rackers, J. A., Wang, Z., Lu, C., Laury, M. L., Lagardere, L., Schnieders, M. J., Piquemal, J. P., Ren, P., and Ponder, J. W. (2018) Tinker 8: Software Tools for Molecular Design, *J Chem Theory Comput 14*, 5273-5289.
[48] Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J Comput Chem 31*, 455-461.
[49] Rose, A. S., and Hildebrand, P. W. (2015) NGL Viewer: a web application for molecular visualization, *Nucleic Acids Res 43*, W576-579.
[50] Jo, S., Kim, T., and Im, W. (2007) Automated builder and database of protein/membrane complexes for molecular dynamics simulations, *PLoS One 2*, e880.
[51] Sindhikara, D. J., Yoshida, N., and Hirata, F. (2012) Placevent: an algorithm for prediction of explicit solvent atom distribution-application to HIV-1 protease and F-ATP synthase, *J Comput Chem 33*, 1536-1543.
[52] Pavlovicz, R. E., Park, H., and DiMaio, F. (2020) Efficient consideration of coordinated water molecules improves computational protein-protein and protein-ligand docking discrimination, *PLoS Comput Biol 16*, e1008103.
[53] Ross, G. A., Bodnarchuk, M. S., and Essex, J. W. (2015) Water Sites, Networks, And Free Energies with Grand Canonical Monte Carlo, *J Am Chem Soc 137*, 14930-14943.
[54] Ben-Shalom, I. Y., Lin, C., Kurtzman, T., Walker, R. C., and Gilson, M. K. (2019) Simulating Water Exchange to Buried Binding Sites, *J Chem Theory Comput 15*, 2684-2691.
[55] Jiang, W. (2019) Accelerating Convergence of Free Energy Computations with Hamiltonian Simulated Annealing of Solvent (HSAS), *J Chem Theory Comput 15*, 2179-2186.
[56] Hopkins, C. W., Le Grand, S., Walker, R. C., and Roitberg, A. E. (2015) Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning, *J Chem Theory Comput 11*, 1864-1874.
[57] Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schioth, H. B., and Gloriam, D. E. (2017) Trends in GPCR drug discovery: new agents, targets and indications, *Nat Rev Drug Discov 16*, 829-842.
[58] Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking, *J Med Chem 55*, 6582-6594.
[59] Kratochwil, N. A., Gatti-McArthur, S., Hoener, M. C., Lindemann, L., Christ, A. D., Green, L. G., Guba, W., Martin, R. E., Malherbe, P., Porter, R. H., Slack, J. P., Winnig, M., Dehmlow, H., Grether, U., Hertel, C., Narquizian, R., Panousis, C. G., Kolczewski, S., and Steward, L. (2011) G protein-coupled receptor transmembrane binding pockets and their applications in GPCR research and drug discovery: a survey, *Curr Top Med Chem 11*, 1902-1924.
[60] Wang, Z., Sun, H., Yao, X., Li, D., Xu, L., Li, Y., Tian, S., and Hou, T. (2016) Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power, *Phys Chem Chem Phys 18*, 12964-12975.
[61] Steinbach, P. J., and Brooks, B. R. (1994) New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation, *Journal of Computational Chemistry 15*, 667-683.
[62] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A Smooth Particle Mesh Ewald Method, *J Chem Phys 103*, 8577-8593.
[63] Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: the biomolecular simulation program, *J Comput Chem 30*, 1545-1614.

[64] Roe, D. R., and Cheatham, T. E., 3rd. (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *J Chem Theory Comput 9*, 3084-3095.
[65] Su, P. C., Tsai, C. C., Mehboob, S., Hevener, K. E., and Johnson, M. E. (2015) Comparison of radii sets, entropy, QM methods, and sampling on MM-PBSA, MM-GBSA, and QM/MM-GBSA ligand binding energies of F. tularensis enoyl-ACP reductase (FabI), *J Comput Chem 36*, 1859-1873.