



NIH PUBLIC ACCESS

Author Manuscript

*J Chem Inf Model.* Author manuscript; available in PMC 2014 January 28.

Published in final edited form as:

*J Chem Inf Model.* 2013 January 28; 53(1): 267–277. doi:10.1021/ci300505n.

## CHARMM-GUI Ligand Binder for Absolute Binding Free Energy Calculations and Its Application

Sunhwan Jo<sup>1</sup>, Wei Jiang<sup>2</sup>, Hui Sun Lee<sup>1</sup>, Benoît Roux<sup>3,4,\*</sup>, and Wonpil Im<sup>1,\*</sup><sup>1</sup>Department of Molecular Biosciences and Center for Bioinformatics, The University of Kansas, 2030 Becker Drive, Lawrence, KS 66047, USA<sup>2</sup>Leadership Computing Facility, Argonne National Laboratory, 9700 South Cass Avenue, Building 240, Argonne, IL 60439, USA<sup>3</sup>Department of Biochemistry and Molecular Biology, Gordon Center for Integrative Sciences, W323B 929 East 57th Street, Chicago, IL 60637, USA<sup>4</sup>Biosciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Building 202, Argonne, IL 60439, USA

### Abstract

Advanced free energy perturbation molecular dynamics (FEP/MD) simulation methods are available to accurately calculate absolute binding free energies of protein-ligand complexes. However, these methods rely on several sophisticated command scripts implementing various biasing energy restraints to enhance the convergence of the FEP/MD calculations, which must all be handled properly to yield correct results. Here, we present a user-friendly web interface, CHARMM-GUI *Ligand Binder* (<http://www.charmm-gui.org/input/gbinding>), to provide standardized CHARMM input files for calculations of absolute binding free energies using the FEP/MD simulations. A number of features are implemented to conveniently setup the FEP/MD simulations in highly customizable manners, thereby permitting an accelerated throughput of this important class of computations while decreasing the possibility of human errors. The interface and a series of input files generated by the interface are tested with illustrative calculations of absolute binding free energies of three non-polar aromatic ligands to the L99A mutant of T4 lysozyme and three FK506-related ligands to FKBP12. Statistical errors within individual calculations are found to be small (~1 kcal/mol), and the calculated binding free energies generally agree well with the experimental measurements and the previous computational studies (within ~2 kcal/mol). CHARMM-GUI *Ligand Binder* provides a convenient and reliable way to setup the ligand binding free energy calculations and can be applicable to pharmaceutically important protein-ligand systems.

### Keywords

Free Energy Perturbation; Molecular Dynamics; Protein-ligand Interactions; T4 Lysozyme; FKBP12

---

\*Corresponding authors; B. Roux: [roux@uchicago.edu](mailto:roux@uchicago.edu), W. Im: [wonpil@ku.edu](mailto:wonpil@ku.edu).

Supporting Information. Anchoring atom and reference restraint values used for the FEP/MD simulations in this study is available in Table S1. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

## INTRODUCTION

Molecular recognition of small molecules by specific receptor proteins is one of the most important biological processes. Binding of a small molecule to a target protein is required to change the receptor's dynamics and/or conformations allosterically, or to interfere with the protein-protein interactions, aiming at controlling the protein's functions. Specific protein-ligand interactions can also be modulated by other small molecules that have similar physicochemical properties, begging for a quantitative assessment of binding specificity at the molecular level. For this reason, there is great interest in the concept of rational drug design based on atomic structures of receptor proteins or receptor-ligand complexes (reviewed in <sup>1-3</sup>).

Arguably, accurate binding free energy calculations could be used to help develop novel drug molecules and improve the potency of existing ones. Considerable efforts have been made to reliably calculate the receptor-ligand association free energy at the atomic level using various computational methods, such as: linear interaction energy (LIE)<sup>4</sup>; molecular mechanics with Poisson-Boltzmann or generalized Born/surface area (MM-PBSA/GBSA)<sup>5, 6</sup>; or alchemical free energy calculation (reviewed in <sup>7-9</sup>). LIE utilizes the energies calculated from atomic simulations, such as molecular dynamics (MD) or Monte Carlo (MC) simulations, to estimate the relative free energy difference upon changes in van der Waals (vdW) and/or electrostatic contributions from the reference. MM-PBSA/GBSA utilizes the trajectories from MD simulations to capture the dynamical aspect of protein-ligand interactions, but the produced results are sometimes not predictive.<sup>10</sup>

Alchemical free energy calculations, particularly, free energy perturbation molecular dynamics (FEP/MD) simulations, represent the most rigorous computational approach to calculate receptor-ligand binding free energy. FEP/MD simulations have shown to accurately calculate the binding affinities in various biologically important systems.<sup>11-14</sup> The effect from explicit solvent molecules and flexibility of the molecules can be incorporated into the calculation, which is shown to be important in macromolecule-ligand association.<sup>15</sup> To improve the robustness and computational efficiency of FEP/MD simulations, Roux and co-workers have developed a staged FEP/MD simulation protocol based on a step-by-step decomposition of the total reversible work through various computational techniques.<sup>12, 14</sup> First, the absolute binding free energy is decomposed into several intermediate steps in which the ligand-surrounding environment interactions as well as the orientational, translational, and conformational sampling of the ligand are gradually turned "on" (or "off" depending on the setup). The staged protocol effectively and rigorously breaks down the complete free energy calculation into several independent MD simulations, which can be easily distributed over independent computer nodes. Second, the ligand conformational sampling is explicitly taken into account by umbrella sampling of ligand conformations based on root-mean-squared deviation (RMSD) from the bound state of the ligand. Third, Lennard-Jones (LJ) interactions between the ligand and the surrounding environment are decomposed into repulsive and dispersive interactions using the Weeks-Chandler-Andersen (WCA) separation, which improves the convergence of the calculation. Fourth, to reduce the computational cost and complexity, only a small region of interest near the receptor-ligand binding site is explicitly considered using solvent boundary potential methods.<sup>16, 17</sup>

The performance of the staged FEP/MD simulation method has been already tested with various biologically important systems and demonstrated that binding free energy can be calculated accurately with errors less than 2 kcal/mol (compared to experimental values).<sup>12, 14, 18</sup> However, the setup of the staged FEP/MD simulation protocol is not trivial even for expert researchers due to the many restraint potentials and large number of

intermediate windows. In addition, selection of a set of anchoring atoms for orientational/translational restraint potentials could be arbitrary and requires prior knowledge of the ligand system.

In this work, we have generalized the staged FEP/MD simulation protocol<sup>12, 14</sup> and implemented CHARMM-GUI *Ligand Binder* (<http://www.charmm-gui.org/input/gbinding>), a web interface that provides standardized input files for calculations of absolute binding free energies. In addition, *Ligand Binder* has a number of features that can help users to setup the FEP/MD simulations with less human intervention. For example, automatic ligand force field generation, symmetric group detection, system size estimation, and anchoring atom selection/validation are implemented to quickly generate reliable FEP/MD simulation systems and inputs. To test the efficacy of the interface, we have calculated absolute binding free energies of three non-polar small aromatic ligands to the L99A mutant of T4 lysozyme and three FK506-related ligands to FKBP12. The calculated binding free energies generally agree well with the experimental measurements and the previous computational studies. Therefore, *Ligand Binder* provides a convenient and reliable way to setup the ligand binding free energy calculations and can be applicable to pharmaceutically important protein-ligand systems.

## METHODS

### Theoretical Background

The theoretical framework for calculation of absolute binding free energy used in this work is well formulated in the previous studies<sup>12, 14</sup> and is closely related to the double decoupling method.<sup>19, 20</sup> According to this framework, a step-by-step FEP/MD simulation protocol with restraining potentials is carried out to break down the complete alchemical reaction path into a number of intermediate physical states (FEP windows). The ordered step-by-step FEP/MD can be briefly summarized as: (1) switch-on the RMSD restraint potential on the ligand conformation in the binding site, (2) switch-on the relative ligand-receptor positional/rotational restraint potentials (in the presence of the RMSD restraint), (3) switch-off the ligand interactions with the binding site environment (in the presence of the RMSD and positional/rotational restraints), (4) switch-off the positional/orientational restraint potentials (this contribution is evaluated analytically), (5) switch-on the ligand interactions with the bulk solution environment (in the presence of the RMSD restraint), (6) switch-off the conformational RMSD restraint potential on the ligand in bulk solution. The various restraining potentials introduced in the successive FEP windows are used to maintain the position and orientation of the ligand around the “pose” adopted in the bound complex. These restraining potentials help enhance configurational sampling and also serve to correctly handle the decoupled ligand states. In the following sections, we will focus on its generalization and implementation in CHARMM-GUI *Ligand Binder*.

### General FEP/MD setup using CHARMM-GUI *Ligand Binder*

The CHARMM input files generated by *Ligand Binder* are arranged in six distinct steps for clarity and convenience (Table 1). *Ligand Binder* produces both inputs and outputs for *STEP 1* and *STEP 2* for a given protein-ligand complex, but only generates the inputs files of the remaining equilibration and production steps of the FEP/MD simulations because they generally require extensive computational resources. In this section, the role of each step and the corresponding outcomes are discussed in detail. As shown in Table 1, for easy recognition the generated input files have a corresponding step number in their filename followed by the two different tags, “*BULK*” and “*SITE*”, to represent the FEP/MD simulation systems in bulk solvent and in the protein binding site, respectively. Hereinafter, for brevity, the simulation systems with protein/ligand complex and the ligand molecule in

the bulk solvent are referred to as *SITE* and *BULK*, respectively. Various accessory scripts that are compatible with UNIX and portable batch system (PBS) are also provided to perform actual FEP/MD simulations conveniently.

**STEP 1: Reading of a protein-ligand complex into CHARMM**—In this step, the web interface generates inputs to read a protein-ligand complex into the molecular dynamics simulation program CHARMM.<sup>21</sup> A user can upload a PDB format file or provide a PDB entry ID from the RCSB protein databank<sup>22</sup> that contains the protein-ligand complex. Because a molecular force field (FF) is not generally available for many small molecules, the web interface provides several options: (1) if the small molecule's topology and parameter is already included in the CHARMM standard FF, a user simply needs to leave the residue name as it is or rename it to match the residue name in the CHARMM FF; (2) if a user already has a CHARMM FF customized for the user's small molecule, the topology and parameter files can be uploaded and used; (3) if no FF is available, *Ligand Binder* provides an option for automated FF generation based on the CHARMM general FF (CGenFF)<sup>23</sup> or general AMBER force field (GAFF)<sup>24, 25</sup> to generate the ligand CHARMM FF (see below for details). Symmetric groups (currently, planar ring or t-butyl-groups) are algorithmically recognized after the topology generation and used in the later steps to apply flat-bottom dihedral restraints for preventing exchanges between physically identical rotameric states. Simple protein structure modifications (e.g., disulfide bond generation, terminal group changes, phosphorylation, and protonation) can also be made during this step (see ref. <sup>26</sup> for details).

**STEP 2: Solvation of the systems**—Once a protein-ligand structure is successfully initialized in CHARMM, the protein-ligand system is solvated in a water box. The current interface supports a spherical water box for a reduced system with solvent boundary potentials and an orthorhombic water box for the full system with periodic boundary conditions (PBC). While one can generate the FEP/MD simulation inputs for the full system with PBC, only the reduced system is used for the illustrative purposes in this work. The size of the spherical water box ( $R_{\text{wat}}$ ) from the ligand's center of mass (COM) can be specified by a user-defined value. For the user's convenience in determining the size, the extent of the ligand ( $R_{\text{lig}}$ ) is provided on the interface. Based on our experience and extensive testing,  $R_{\text{wat}}$  that is at least ( $R_{\text{lig}} + 1 \text{ \AA}$ ) for *BULK* and ( $R_{\text{lig}} + 5 \text{ \AA}$ ) for *SITE* are recommended for stable simulations. Nonetheless,  $R_{\text{wat}}$  may need to be optimized for specific problems. If a ligand molecule is composed of a small number of atoms, it may not be necessary to calculate the conformational free energy change upon binding. Therefore, the option for the conformational free energy simulation is unchecked by default if  $R_{\text{lig}}$  is smaller than 5 Å, but it can be turned on by the user.

With the given option in this step, *Ligand Binder* generates the input and output files for *STEP 2*, as well as the input files for the rest of the FEP/MD simulations (*STEP 3–6*). The user can download the generated files in an archived file and carry out the FEP/MD simulations on user's machine. It should be noted that although many parts of the input files are similar for different ligands, some system-dependent parameters are inserted during the input generation. Thus, it is advised to re-generate the input files through *Ligand Binder* instead of copying and modifying the input files for different ligands if the user is not confident.

**STEP 3: Setup of the solvent boundary systems**—In this step, the solvent boundary potentials are set up to generate the reduced atomic systems. For *SITE*, the generalized solvent boundary potential (GSBP)<sup>17</sup> is used. *Ligand Binder* uses the following setup for the GSBP system as a default. A spherical inner region is defined by  $R_{\text{wat}}$  from the ligand's COM. All the atoms within the inner region are allowed to move during the equilibration

and FEP/MD simulations. The inner region is extended by 3 Å to define a smooth spherical dielectric cavity ( $R_{\text{sphere}}$ ). The protein atoms in the extended region ( $R_{\text{wat}} < R < R_{\text{sphere}}$ ) as well as the atoms in the inner region linked via 1–3 bonds with the protein atoms in the outer region ( $> R_{\text{sphere}}$ ) are fixed according to a group-based criterion. The water molecules in the outer region are removed and treated implicitly, and the inner region water molecules that overlap with the ligand and the protein are also deleted. The long-range electrostatic influence of the protein atoms in the outer region on the atoms in the inner region is represented in terms of the solvent-shielded static field and the solvent-induced reaction field. The reaction field due to the variations of the charge distribution in the inner region is expressed in terms of a generalized multipolar expansion using 11 spherical harmonic functions by default. The solvent-shielded static field and the reaction field matrix are calculated by the finite-difference Poisson-Boltzmann (PB) method using the PBEQ module<sup>27</sup> in CHARMM, assuming a dielectric constant of 1 inside the protein as well as the inner and extended regions, and 80 otherwise. A non-polar cavity potential is used to keep the water molecules inside the inner region.

For *BULK*, the spherical solvent boundary potential (SSBP)<sup>16</sup> is used together with the spherical water box generated in the previous step. The water box and the ligand molecules are centered at the origin, and water molecules that overlap with the ligand are deleted.

#### STEP 4: Equilibration of the systems

**Equilibration of SITE (steps 4.1 and 4.2):** To overcome the problem that the size of the inner system is invariant during the subsequent simulations, the grand canonical Monte Carlo (GCMC) method is used to hydrate the inner region properly (*step 4.1*).<sup>28</sup> In this step, 20 cycles of MC and MD simulations are performed. Each cycle is comprised of 10,000 MC moves followed by 10,000 MD steps with a time-step of 2 fs. The MC steps include rigid body translation, rotation, and GCMC insertion/deletion of water molecules, and each move type has the equal probability. A harmonic restraint potential with a force constant of 5.0 kcal/(mol·Å<sup>2</sup>) is applied to the protein and the ligand molecule throughout the MD simulations. All bonds involving hydrogen atoms are fixed with the SHAKE algorithm.<sup>29</sup> The simulation is carried out using Langevin dynamics at 300 K with a friction coefficient of 5 ps<sup>-1</sup> assigned to all non-hydrogen atoms. During the simulation, nonbonded interactions within 14.5 Å are explicitly accounted and the rest is treated with an extended electrostatics method. After 20 cycles of GCMC/MD simulations, the protein-ligand complex is equilibrated (*step 4.2*) for 200 ps at 300 K using Langevin dynamics with the same option as in *step 4.1*, but without the positional harmonic restraint.

**Geometric and conformational restraint setup (step 4.3):** In this step, a set of anchoring atoms for the translational and orientational restraints are determined and the detailed procedures are described below. After the anchoring atoms are determined, the last 190-ps equilibration trajectory is used to calculate the average reference distances, angles, and dihedrals for the translational/orientational restraints. If the ligand has a symmetric group, the average dihedral angle around the symmetric group is also calculated. The average ligand conformation during the equilibration simulation is first calculated followed by a short energy minimization. The resulting averaged and minimized ligand structure is used as the reference configuration for the conformational restraint during FEP/MD simulations.

**Equilibration of BULK (step 4.4):** The ligand molecule in bulk solvent is equilibrated for 200 ps at 300 K using the SSBP method and Langevin dynamics with the same option as in *step 4.1*. A weak positional restraint potential with a force constant of 0.5 kcal/(mol·Å<sup>2</sup>) is applied to the ligand's COM to prevent the ligand from drifting away from the origin. The conformational restraint potential using the averaged ligand structure calculated in *step 4.3*

is applied to keep the reference conformation during the equilibration if the conformational free energy calculation option is selected in *STEP 2*.

**STEP 5: FEP/MD and RMSD-umbrella sampling simulations**—In this step, the FEP/MD simulations are carried out to calculate free energies associated with restriction of ligand conformation to the reference conformation ( $\Delta G_c$ ), restriction of ligand orientation and translation ( $\Delta G_{t,r}$ ), and interactions with surrounding environments ( $\Delta G_{int}$ ). In the current scheme, the FEP/MD simulations are divided into a total of 137 independent (FEP/RMSD window) simulations that can be distributed over independent computer nodes. During the simulations, the perturbation energies for each state are collected for the later analysis. All simulations are carried out using Langevin dynamics with the same option as in *step 4.1*. An accessory script (*step5\_jobmanager.com*) is provided to automatically check finished jobs and to (re)submit the FEP/MD and RMSD-umbrella sampling simulations (see Table 1). If necessary, optional GCMC simulation can be performed to adjust the solvation of reduced system during the FEP/MD simulations (the GCMC simulation option has to be turned on in the *STEP 2*).

**Conformational free energy:** To calculate the free energy associated with the restriction of the ligand conformation in the bound state, the conformational free energies ( $\Delta G_c^{site}$  and  $\Delta G_c^{bulk}$ ) are estimated by calculating the potential of mean force (PMF) as a function of ligand RMSD using umbrella sampling simulations.<sup>30</sup> Simulations are carried out using the quadratic biasing potential,  $k_c(\xi - \xi_j)^2$ , where  $k_c$  is the force constant,  $\xi$  is the instantaneous RMSD with respect to the average ligand conformation in the bound state that is calculated from the *SITE* equilibration, and  $\xi_j$  is the RMSD offset value of each window. A total of 21 biasing windows are used with the RMSD offset value from 0.0 to 5.0 Å in steps of 0.25 Å for the ligand in *SITE* and *BULK*. The initial configurations for the 21 umbrella sampling windows are generated using a short initial run (20 ps) with a strong force constant of 500 kcal/(mol·Å<sup>2</sup>). Then, each window is simulated using a force constant of 10 kcal/(mol·Å<sup>2</sup>) for 110 ps and the data from the last 100-ps simulations are collected, which corresponds to one cycle.

During the umbrella sampling simulation for *SITE*, dissociation of the ligand from the binding pocket may occur for some windows with a large RMSD offset value due to large distortion of the ligand conformation. According to the strict step-by-step free energy decomposition scheme adopted in the present work, the translational degrees of freedom of the ligand should not be restricted at this stage. Nevertheless, to avoid the spurious contributions to the PMF from an unbound ligand, a flat-bottomed translational restraint with a small force constant (1.0 kcal/(mol·Å<sup>2</sup>)) is applied. This restraint, which allows free diffusion of the ligand's COM within a spherical region of 2 Å radius centered around the initial position of ligand's COM, is sufficient to prevent the ligand from leaving the binding pocket.

If symmetric groups (e.g., planar ring group or t-butyl group) are present in the ligand, interconversion between the physically equivalent conformations will make the ligand conformational sampling more challenging. Insufficient sampling will eventually lead to an inaccurate estimation of loss of conformational freedom upon ligand binding. To avoid this problem, a steep flat-bottom dihedral restraint potential with a force constant of 500 kcal/(mol·rad<sup>2</sup>) is applied to each of the symmetric unit of the ligand during all the calculations in *BULK* and *SITE*.

**Translational and rotational free energy:** The free energies corresponding to restriction of the ligand's translational and rotational degrees of freedom near the binding site ( $\Delta G_{t,r}^{site}$ ) are

calculated using FEP/MD simulations. The six point positions in the protein ( $P_1$ ,  $P_2$ , and  $P_3$ ) and the ligand ( $L_1$ ,  $L_2$ , and  $L_3$ ) determined in *step 4.3* (see below for details) are used to define the relative position and orientation of the ligand with respect to the protein. A total of three translational (1 distance, 1 angle, and 1 dihedral restraints) and three rotational restraints (1 angle, 2 dihedral restraints) are applied with the reference values determined in *step 4.3*. The translational and rotational restraints are gradually turned on via the linear coupling parameters  $\lambda_{t,r}$  (Table 2) to final force constants of 10 kcal/(mol·Å<sup>2</sup>), 200 kcal/(mol·rad<sup>2</sup>), and 200 kcal/(mol·rad<sup>2</sup>) for the distance, angle, and dihedral restraints, respectively. For each set of coupling parameters, 50-ps simulations are performed and the data from the last 40-ps simulations are collected, which corresponds to one cycle.

**Interactions with surrounding environments:** The contributions from the interactions of the ligand with its surrounding environments ( $\Delta G_{\text{int}}^{\text{site}}$  and  $\Delta G_{\text{int}}^{\text{bulk}}$ ) are calculated with FEP/MD simulations. The Lennard-Jones (LJ) potential is separated into repulsive and dispersive free energies using the WCA separation method.<sup>31</sup> Associated coupling parameters,  $\lambda_{\text{rep}}$ ,  $\lambda_{\text{disp}}$ , and  $\lambda_{\text{elec}}$  are introduced to control the repulsive, dispersive, and electrostatic interactions, respectively (Table 2). For each set of coupling parameters, 110-ps simulations are performed and the data from the last 100-ps simulations are collected, which corresponds to one cycle.

**STEP 6: Calculation of the binding free energy**—In the current staged FEP/MD simulation scheme, the final binding free energy can be expressed as

$$\Delta G_b^\circ = \Delta \Delta G_{\text{int}} + \Delta \Delta G_{t,r} + \Delta \Delta G_c \quad (1)$$

where  $\Delta \Delta G_{\text{int}} = [\Delta G_{\text{int}}^{\text{site}} - \Delta G_{\text{int}}^{\text{bulk}}]$  represents the free energy change due to non-bonded interactions between the ligand and its environment.  $\Delta \Delta G_c = [\Delta G_c^{\text{bulk}} - \Delta G_c^{\text{site}}]$  and  $\Delta \Delta G_{t,r} = [-k_B T \ln F_t C^\circ - k_B T \ln F_r - \Delta G_{t,r}^{\text{site}}]$  represent the free energy changes due to the loss of conformational freedom and the translational/rotational freedom, respectively. The ligand's translational ( $F_t$ ) and rotational ( $F_r$ ) freedoms in *BULK* are calculated analytically.<sup>12,14</sup>

In this step, the data generated in *STEP 5* are aggregated and each term in the standard binding free energy is computed. The weighted histogram analysis method (WHAM)<sup>32, 33</sup> is used to evaluate the PMF as a function of RMSD from the umbrella sampling simulations in *STEP 5*; a Fortran WHAM code is provided. The free energy change due to the loss of conformational freedom ( $\Delta G_c$ ) is calculated by integration of the Boltzmann factor of the RMSD PMF. For *BULK*,  $\Delta G_c^{\text{bulk}}$  is given by

$$\Delta G_c^{\text{bulk}} = \frac{\int d\xi e^{\beta[w_c^{\text{bulk}}(\xi) + k_c \xi^2]}}{\int d\xi e^{-\beta w_c^{\text{bulk}}(\xi)}} \quad (2)$$

where  $w_c^{\text{bulk}}(\xi)$  is the ligand PMF as a function of its RMSD with respect to the reference ligand conformation (i.e., average structure calculated in *step 4.3*). A similar expression is used for  $\Delta G_c^{\text{site}}$ .

The data from the *STEP 5* FEP/MD simulations (for  $\Delta G_{\text{int}}^{\text{site}}$  and  $\Delta G_{\text{int}}^{\text{bulk}}$ ) are unbiased using the WHAM facility in CHARMM. The individual free energy components are calculated for

the ligand in *SITE* and *BULK*, respectively. Once individual free energy components are computed, a table of free energy can be printed using an accessory script (*step6\_table.py*).

### Automatic Generation of Ligand Force Field

*Ligand Binder* provides automatic FF generation options for small molecules using CGenFF<sup>23</sup> or GAFF.<sup>24, 25</sup> The initial structure of a ligand molecule with proper protonation states and explicit hydrogen atoms is required. When the CGenFF option is selected, *Ligand Binder* communicates with the ParamChem server (<http://www.paramchem.org>), which automatically generates the ligand FF based on analogy to known small molecular FF parameters. The ParamChem server assigns the charge and parameter penalty values, which could be used as indicative values for the parameter quality, and it is advisable to check if further optimization is needed before performing the FEP/MD simulations. Although it is not generally recommended due to a compatibility issue, a ligand FF based on GAFF can be generated. When the GAFF option is selected, the topology and parameters of the ligand are generated in the CHARMM format using the Antechamber toolkit.<sup>25</sup> The charges of the ligands are from AM1/BCC,<sup>34</sup> and an explicit charge state can be specified in the case that the ligand molecule is not neutral.

### Automatic Determination of Anchoring Atoms

The translational and orientational restraint potentials applied to the ligand in the binding site are constructed from six anchoring atoms: three point positions in the protein ( $P_1$ ,  $P_2$ , and  $P_3$ ) and three point positions in the ligand ( $L_1$ ,  $L_2$ , and  $L_3$ ) (Figure 1). *Ligand Binder* provides a protocol for the automatic selection of reasonable anchoring points for any protein-ligand complex in *step 4.3*. Although it is referred to as an anchoring “atom”, the average coordinates of a group of atoms are used to define each anchoring position. Specifically,  $L_1$  is defined by the COM of a non-hydrogen ligand atom ( $L_0$ ) closest to the ligand’s COM and the ligand atoms that are bonded to  $L_0$ .  $P_1$  is defined by the receptor backbone heavy atoms of a residue that is closest to the receptor’s COM. If the distance between  $P_1$  and  $L_1$  is too close ( $< 5 \text{ \AA}$ ) or too far ( $> 10 \text{ \AA}$ ),  $P_1$  is randomly reselected from the receptor backbone heavy atoms of a residue within a distance of  $5 - 10 \text{ \AA}$  from  $L_1$ .  $P_2$  is defined by the receptor backbone heavy atoms of a residue that satisfy the condition of  $60^\circ < \angle L_1 P_1 P_2 < 120^\circ$ .  $P_3$  is defined by the receptor backbone heavy atoms of a residue that satisfy the condition of  $60^\circ < \angle P_1 P_2 P_3 < 120^\circ$ .  $L_2$  (and  $L_3$ ) is defined by the COM of a ligand heavy atom and the atoms bonded to the heavy atom with the following conditions;  $30^\circ < \angle P_1 L_1 L_2 < 150^\circ$  and  $30^\circ < \angle L_1 L_2 L_3 < 150^\circ$ . These conditions were empirically established through extensive testing to ensure consistent FEP/MD results.

The translational restraint is defined as  $U_t = 1/2[k_r(r - r_0)^2 + k_\theta(\theta - \theta_0)^2 + k_\phi(\phi - \phi_0)^2]$ , where  $r$  is the distance between  $P_1$  and  $L_1$ ,  $\theta$  is the angle  $\angle P_2 P_1 L_1$ , and  $\phi$  is the dihedral angle  $\angle P_3 P_2 P_1 L_1$ .  $k_r$ ,  $k_\theta$ , and  $k_\phi$  are the force constants for distance, angle, and dihedral angle restraints, respectively, and  $r_0$ ,  $\theta_0$ , and  $\phi_0$  are the reference restraint values taken from the equilibration simulation in *SITE* (*step 4.3*). Similarly, the orientational restraint is defined as  $U_r = 1/2[k_\alpha(\alpha - \alpha_0) + k_\beta(\beta - \beta_0) + k_\gamma(\gamma - \gamma_0)^2]$ , where  $\alpha$  is the angle  $\angle P_1 L_1 L_2$ ,  $\beta$  is the dihedral angle  $\angle P_2 P_1 L_1 L_2$ , and  $\gamma$  is the dihedral angle  $\angle P_1 L_1 L_2 L_3$ . In principle, the six anchoring atoms could be chosen arbitrarily without affecting the outcome of the FEP/MD calculations. However, poorly chosen anchoring atoms may yield geometric restraints that are ineffective to maintain the relative ligand-receptor orientation and cause undesired instabilities during the FEP/MD simulations. For this reason, it is important to select an adequate set of anchoring atoms.



## Computational Details

In order to test a series of input files generated by *Ligand Binder*, illustrative calculations of absolute binding free energies were performed for three non-polar aromatic ligands to the L99A mutant of T4 lysozyme and three FK506-related ligands to FKBP12 (see Figure 2 for the ligand structures) in the reduced simulation systems (Figure 3 and Table 3). The protein-ligand complex structures from the PDB in Table 3 were used to initialize the structures in *Ligand Binder*. The FF for each ligand was automatically generated by CGenFF with ligand structure files obtained from the PDB without modification. As mentioned above, the ParamChem server annotates the charge and parameter penalties as an indicative value for the parameter quality. The three nonpolar ligands used in T4 lysozyme L99A mutant had zero penalties for charge and parameter values. The three FK506-related ligands had a charge penalty of about 100 and a parameter penalty of about 141. Although further optimizations for the entries with a penalty greater than 50 are recommended by the authors of the ParamChem server, such an optimization was not attempted in this work.

All the MD simulations were carried out using CHARMM.<sup>21</sup> All the simulation system setup and FEP/MD calculations were performed using the input files generated by *Ligand Binder* without further modification. For the non-polar ligands in T4 lysozyme L99A mutant, the option for the conformational free energy calculation was not selected due to their small sizes by default. Figure 3 shows a typical reduced protein/ligand complex system for *SITE* and *BULK* using GSBP<sup>17</sup> and SSBP<sup>16</sup> boundary potentials. The number of water molecules is given in Table 3 before and after the hydration of *SITE* using GCMC/MD simulations. Each protein/ligand complex system was equilibrated for 200 ps with Langevin dynamics without any biasing restraints. After equilibration was done for *SITE*, the anchoring atoms for the translational and orientational restraints were automatically chosen and stored in a stream file (*step4.3\_restraint\_geo.prm*) for the FEP/MD simulations. The equilibration trajectory was also used to determine the reference values for the distance ( $r$ ), angle ( $\theta$  and  $\alpha$ ), and dihedral ( $\varphi$ ,  $\beta$ , and  $\gamma$ ) restraints (Figure 1). The list of anchoring atoms used in the simulation as well as the reference values are given in Supporting Information Table S1. The *BULK* systems were also equilibrated for 200 ps with Langevin dynamics with ligand conformation restrained using the average ligand conformation from the *SITE* equilibration as a reference.

For each protein-ligand complex in Table 3, 10 cycles of FEP/MD simulations were carried out for better convergence. The conformational free energy term was not calculated for T4 lysozyme ligands due to small ligand size. Each cycle consisted of 10-ps equilibration and 100-ps production except for the translational/rotational free energy contribution (10-ps equilibration and 50-ps production). Each cycle was started using the last coordinates of the previous cycle with random initial velocities. The free energy values and the errors are presented using the average and the standard deviation of the last five cycles, respectively. Two additional (independent) 10-cycle FEP/MD simulations were performed using different sets of automatically chosen anchoring atoms to illustrate the robustness of our scheme for anchoring atom selections.

## RESULTS AND DISCUSSION

To examine and illustrate the FEP/MD simulation protocol adopted in CHARMM-GUI *Ligand Binder*, the 18 test FEP/MD simulations using six different protein-ligand complexes were performed (Table 3). The calculated standard binding free energies are given in Table 4. In this section, we first discuss the FEP/MD calculation results, and a general discussion on the binding free energy calculation results is then followed.

### Illustration of FEP/MD simulations using CHARMM-GUI *Ligand Binder*

For each ligand, the reduced systems were copied to make three independent simulation systems. The GCMC/MD simulations and the equilibrations were performed independently. In addition, different sets of anchoring atoms were used for each independent simulation. By doing so, the independent simulations serve to estimate the impact of slightly different initial conditions and different selections of anchoring atoms on the final binding free energies (C1, C2, and C3 in Table 4).

The reduced GSBP representation of the binding site is hydrated with GCMC/MD simulations in the presence of the ligand molecule during the equilibration *step 4.1*, and the number of water molecules was fixed during the FEP/MD simulations. Although *Ligand Binder* provides an option to allow the number of water molecules to fluctuate during the FEP/MD simulations, such option was not activated in this study to reduce the computational time. Based on additional tests, the final binding free energies of the test systems in the current study are not significantly affected by allowing the number of water to fluctuate during the FEP/MD simulations, which is in agreement with the observation recently made by Lee et. al.<sup>35</sup> However, rehydration of the binding site could be potentially important to accurately calculate the binding free energy when the size of the reduced system is small or the size of the ligand is large.<sup>28</sup>

In general, the smaller size of T4 lysozyme non-polar ligands had more consistent results for each independent calculation with better convergence compared to larger FKBP ligands. Better convergence of smaller non-polar ligands is expected due to the small size of the ligand and well-defined hydrophobic binding pocket in the receptor. In contrast, FKBP ligands showed relatively larger standard deviations (1–2 kcal/mol) between the cycles of FEP/MD simulations. In addition, the calculated binding free energy of an independent system showed about 1–2 kcal/mol differences.

When compared to the previous computational results<sup>12, 14</sup> and experimental data<sup>36, 37</sup> in Table 5, our results generally agree well within statistical errors, although there are differences for specific systems. For example, the calculated binding free energy of indole is about 2 kcal/mol less favorable than the previous calculation,<sup>12</sup> and the calculated free energy of FKBP ligands in this work showed more statistical errors compared to the previous calculations.<sup>14</sup> Although the current protocol implemented in *Ligand Binder* is based on the previous work done by Roux and co-workers,<sup>12, 14</sup> there are several reasons for such discrepancy. Certainly, longer equilibration and careful adjustments of ligand FF parameters could enhance the accuracy of the calculation results. For instance, a different FF parameter set was used (CHARMM PARAM22<sup>38</sup> FF in ref<sup>12</sup> and Antechamber toolkit<sup>25</sup> in ref<sup>14</sup>) and/or the sampling time was much longer in the previous calculations (about 1–2 ns in ref<sup>14</sup> versus about 100 ps in this work for each cycle). Because the purpose of the current work is to illustrate the standardized and automated protocol for absolute binding free energy calculations, further optimization is not performed. Certainly, some optimization by user (e.g., force field parameters and equilibration/sampling times) could be beneficial.

### Validation of anchoring atoms for orientational restraints

In the current FEP/MD simulation protocol, the restraint potentials are used to maintain the position and orientation of the ligand around the “pose” adopted in the bound complex. These restraints are also used to calculate the contributions from the restriction of orientational and translational degrees of freedom of the ligand molecule upon its binding to a receptor protein. A set of six anchoring atoms (three from the receptor protein and three from the ligand molecule) is selected to define the ligand position and orientation in terms of distance, angle, and dihedral angles (Figure 1).

In principle, the final binding free energy should not be affected by the choice of anchoring atoms. The translational and orientational restraints have been widely adopted in the double decoupling method<sup>12, 13, 20, 39</sup> and the choice of anchoring atoms were shown to have only minimal influences on the final binding free energy.<sup>20</sup> However, it is easy to imagine possible sets of anchoring atoms that are problematic. For example, if a set of anchoring atoms in the ligand yields a co-linear reference angle (e.g.,  $\angle L_1L_2L_3 \sim 180^\circ$ ), the orientation of the ligand relative to the receptor will be poorly defined. Poor choices can be avoided when the anchoring atoms are selected manually, but objective criteria must be implemented when the anchoring atoms are selected by an automated protocol. *Ligand Binder* provides an automated anchoring atom selection protocol, which carries out some basic tests to prevent a poor choice of anchoring atoms.

Table 6 shows the sum of  $\Delta G_{\text{int}}^{\text{site}}$  and  $\Delta\Delta G_{\text{t,r}}$  from independent FEP/MD simulations with different sets of anchoring atoms. Because the orientational and translational restraint potentials were applied during the FEP/MD simulations of  $\Delta G_{\text{int}}^{\text{site}}$  and  $\Delta\Delta G_{\text{t,r}}$ , the individual free energy terms calculated using different sets of anchoring atoms could have different values, but the sum of these terms should remain unchanged regardless of the choice of the anchoring atoms. As shown in the sum of  $\Delta G_{\text{int}}^{\text{site}}$  and  $\Delta\Delta G_{\text{t,r}}$  from the independent FEP/MD simulations, different sets of anchoring atoms automatically chosen by *Ligand Binder* have only minimal impacts on the final free energy, illustrating that the automatic anchoring atom selection in *Ligand Binder* is reliable.

## CONCLUSIONS

In this work, we have developed a web interface, CHARMM-GUI *Ligand Binder* (<http://www.charmm-gui.org/input/gbinding>), to provide standardized CHARMM input files for absolute binding free energy calculations using the FEP/MD simulations. The sophisticated staged FEP/MD simulation protocol based on the step-by-step decomposition of the total reversible work representing the binding process<sup>12, 14</sup> is adopted in *Ligand Binder*. The method is closely related to the double decoupling method,<sup>19, 20</sup> and it has been applied to calculate absolute binding free energies in a number of biologically important systems.<sup>11–14</sup>

*Ligand Binder* has a number of features that help the user to setup the FEP/MD simulations in straightforward manners. First, a ligand FF can be easily generated and incorporated into the simulation inputs. Although a careful parameterization may need to be carried out by the user for improved results, a ligand FF can be generated automatically using either CGenFF<sup>23</sup> or GAFF.<sup>24, 25</sup> In addition, *Ligand Binder* can automatically detect symmetric groups based on the ligand topology (e.g., currently planar ring and t-butyl group) and generate all the necessary input files. Another important feature of *Ligand Binder* is to provide a robust method to select anchoring atoms without knowing the ligand molecule *a priori*. All input files generated by *Ligand Binder* can be downloaded, which allows customization of the FEP/MD simulation protocol if necessary.

The current implementation of *Ligand Binder* supports reduced system representation using SSBP for the ligand in solution, and GSBP for the protein binding site. While these approaches provide useful approximations, it is important to note the situations for which there can be problems. Usage of GSBP, in particular, could result in systematic deviations and inaccuracies of the calculated binding free energies when the receptor protein is highly flexible or undergoes large conformational change upon ligand binding.<sup>35, 40</sup> To alleviate the problem, *Ligand Binder* also supports fully explicit system representation with PBC and comprehensive testing is underway. To enhance sampling and improve convergence, the replica-exchange molecular dynamics (REMD) method is also adopted<sup>41, 42</sup> and can be

activated by the user. As high performance computing resources become increasingly available, state-of-the-art free energy methodologies are expected to eventually migrate toward the full replica-exchange FEP/MD simulations with PBC. Consequently, *Ligand Binder* will support these trends by generating all the necessary input files for the highly scalable simulation programs with built-in parallel/parallel features such as NAMD in the future.

To test the efficacy of *Ligand Binder*, we have performed a set of illustrative absolute binding free energy calculations for three non-polar aromatic ligands to the L99A mutant of T4 lysozyme and three FK506-related ligands to FKBP12. Without any further modification of the current protocol (i.e., generated input files), statistical errors within individual calculations are found to be small (~1–2 kcal/mol), and the calculated binding free energies generally agree well with the experimental measurements and the previous computational studies. It is likely that statistical precision and convergence of the calculations could be improved by carrying out multiple FEP/MD runs or by increasing the sampling times of all the FEP windows. A larger statistical error is observed in FK506-related ligands. Thus, slight modification based on the input files generated by *Ligand Binder* for better sampling would be advisable for medium to large sized ligand molecules.

*Ligand Binder* allows one to quickly setup and perform FEP/MD simulations, thus providing an ideal platform for accurate calculations of the absolute binding free energy. By providing standardized input scripts containing sophisticated commands for various biasing potentials, *Ligand Binder* could facilitate robust calculations of absolute binding free energy without intimate knowledge of ligand molecule or lengthy trial simulations. Such an automated and standardized protocol in alchemical free energy can also be used in evaluation of the binding affinity of a large number of molecules or in a lead optimization process.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

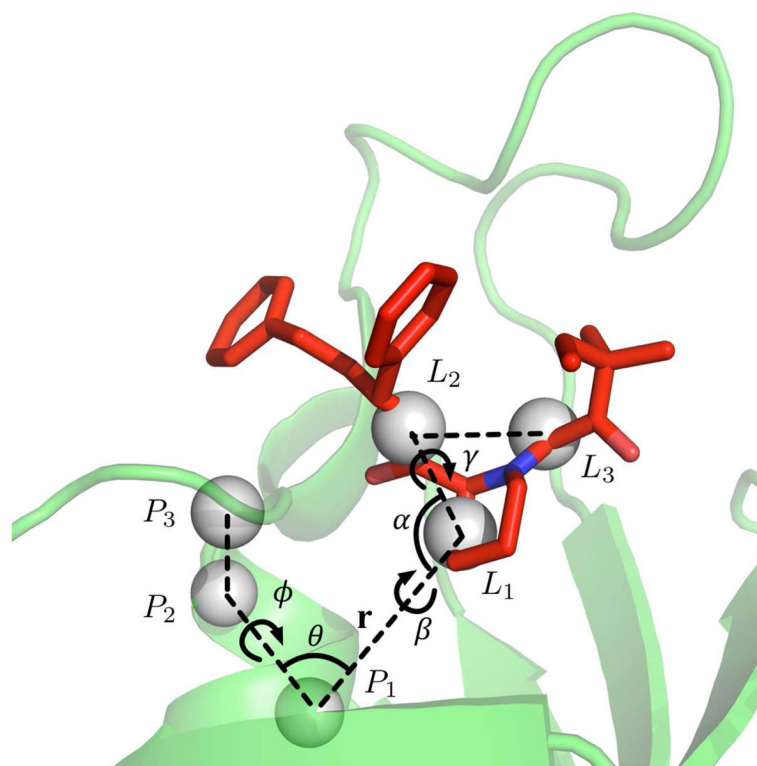
This work was supported by the University of Kansas General Research Fund allocation #2301388-003, NSF ABI-1145987, NIH U54 GM087519-01, TeraGrid/XSEDE resources (TG-MCB070009) (to W.I.), and NSF MCB-0920261 (to B.R.).

## References

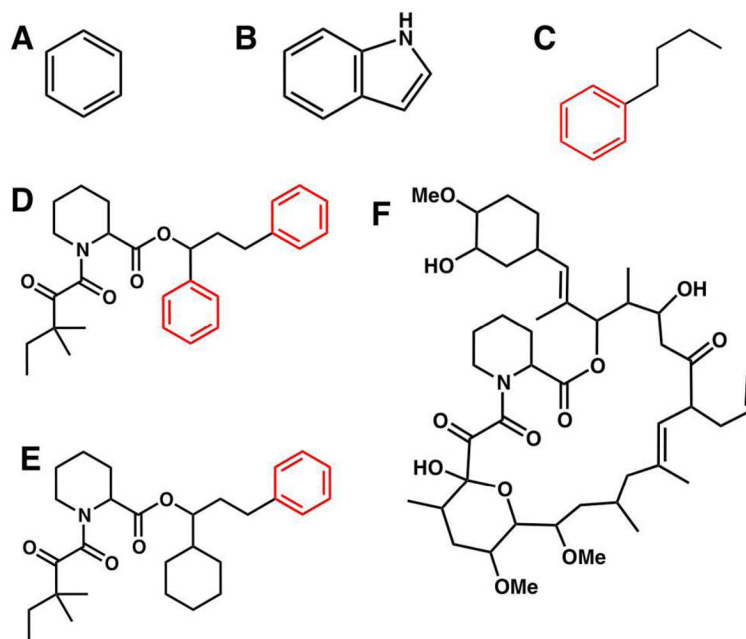
1. Carlson HA, McCammon JA. Accommodating protein flexibility in computational drug design. *Mol Pharmacol.* 2000; 57:213–218. [PubMed: 10648630]
2. Gane PJ, Dean PM. Recent advances in structure-based rational drug design. *Curr Opin Struct Biol.* 2000; 10:401–404. [PubMed: 10981625]
3. Huggins DJ, Sherman W, Tidor B. Rational Approaches to Improving Selectivity in Drug Design. *J Med Chem.* 2012; 55:1424–1444. [PubMed: 22239221]
4. Aqvist J, Medina C, Samuelsson JE. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* 1994; 7:385–391. [PubMed: 8177887]
5. Srinivasan J, Cheatham T, Cieplak P, Kollman P, Case D. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J Am Chem Soc.* 1998; 120:9401–9409.
6. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res.* 2000; 33:889–897. [PubMed: 11123888]

7. Deng Y, Roux B. Computations of standard binding free energies with molecular dynamics simulations. *J Phys Chem B*. 2009; 113:2234–2246. [PubMed: 19146384]
8. Mobley DL, Dill KA. Binding of small-molecule ligands to proteins: “what you see” is not always “what you get”. *Structure*. 2009; 17:489–498. [PubMed: 19368882]
9. Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS. Alchemical free energy methods for drug discovery: progress and challenges. *Curr Opin Struct Biol*. 2011; 21:150–160. [PubMed: 21349700]
10. Pearlman DA. Evaluating the molecular mechanics poisson-boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. *J Med Chem*. 2005; 48:7796–7807. [PubMed: 16302819]
11. Fujitani H, Tanida Y, Ito M, Jayachandran G, Snow CD, Shirts MR, Sorin EJ, Pande VS. Direct calculation of the binding free energies of FKBP ligands. *J Chem Phys*. 2005; 123:084108. [PubMed: 16164283]
12. Deng Y, Roux B. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. *J Chem Theory Comput*. 2006; 2:1255–1273.
13. Mobley DL, Chodera JD, Dill KA. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J Chem Phys*. 2006; 125:084902. [PubMed: 16965052]
14. Wang J, Deng Y, Roux B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys J*. 2006; 91:2798–2814. [PubMed: 16844742]
15. Shoichet BK, Leach AR, Kuntz ID. Ligand solvation in molecular docking. *Proteins*. 1999; 34:4–16. [PubMed: 10336382]
16. Beglov D, Roux B. Finite Representation of an Infinite Bulk System - Solvent Boundary Potential for Computer-Simulations. *J Chem Phys*. 1994; 100:9050–9063.
17. Im W, Bernèche S, Roux B. Generalized solvent boundary potential for computer simulations. *J Chem Phys*. 2001; 114:2924–2937.
18. Ge X, Roux B. Absolute binding free energy calculations of sparsomycin analogs to the bacterial ribosome. *J Phys Chem B*. 2010; 114:9525–9539. [PubMed: 20608691]
19. Gilson MK, Given JA, Bush BL, McCammon JA. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys J*. 1997; 72:1047–1069. [PubMed: 9138555]
20. Boresch S, Tettinger F, Leitgeb M, Karplus M. Absolute binding free energies: A quantitative approach for their calculation. *J Phys Chem B*. 2003; 107:9535–9551.
21. Brooks BR, Brooks CL, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoseck M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. *J Comput Chem*. 2009; 30:1545–1614. [PubMed: 19444816]
22. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res*. 2000; 28:235–242. [PubMed: 10592235]
23. Vanommeslaeghe K, Hatcher ER, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*. 2010; 31:671–690. [PubMed: 19575467]
24. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem*. 2004; 25:1157–1174. [PubMed: 15116359]
25. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graphics Modell*. 2006; 25:247–260.
26. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem*. 2008; 29:1859–1865. [PubMed: 18351591]
27. Im W, Beglov D, Roux B. Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput Phys Commun*. 1998; 111:59–75.
28. Deng Y, Roux B. Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations. *J Chem Phys*. 2008; 128:115103. [PubMed: 18361618]

29. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J Comput Phys.* 1977; 23:327–341.
30. Woo HJ, Roux B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc Natl Acad Sci US A.* 2005; 102:6825–6830.
31. Weeks JD, Chandler D, Andersen HC. Role of Repulsive Forces in Determining Equilibrium Structure of Simple Liquids. *J Chem Phys.* 1971; 54:5237–5247.
32. Kumar S, Bouzida D, Swendsen R, Kollman P, Rosenberg J. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem.* 1992; 13:1011–1021.
33. Souaille M, Roux B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput Phys Commun.* 2001; 135:40–57.
34. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem.* 2002; 23:1623–1641. [PubMed: 12395429]
35. Lee HS, Jo S, Lim H-S, Im W. Application of Binding Free Energy Calculations to Prediction of Binding Modes and Affinities of MDM2 and MDMX Inhibitors. *J Chem Inf Model.* 2012
36. Holt D, Luengo J, Yamashita D, Oh H, Konialian A, Yen H, Rozamus L, Brandt M, Bossard M, Levy M, Eggleston D, Liang J, Schultz L, Stout T, Clardy J. Design, Synthesis, and Kinetic Evaluation of High-Affinity Fkbp Ligands and the X-Ray Crystal-Structures of Their Complexes with Fkbp12. *J Am Chem Soc.* 1993; 115:9925–9938.
37. Morton A, Matthews BW. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry.* 1995; 34:8576–8588. [PubMed: 7612599]
38. MacKerell A, Bashford D, Bellott M, Dunbrack R, Evanseck J, Field M, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau F, Mattos C, Michnick S, Ngo T, Nguyen D, Prodhom B, Reiher W, Roux B, Schlenkrich M, Smith J, Stote R, Straub J, Watanabe M, Wiorcikiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B.* 1998; 102:3586–3616.
39. Hermans J, Wang L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J Am Chem Soc.* 1997; 119:2707–2714.
40. Lau AY, Roux B. The hidden energetics of ligand binding and activation in a glutamate receptor. *Nat Struct Mol Biol.* 2011; 18:283–287. [PubMed: 21317895]
41. Jiang W, Hodoscek M, Roux B. Computation of Absolute Hydration and Binding Free Energy with Free Energy Perturbation Distributed Replica-Exchange Molecular Dynamics. *J Chem Theory Comput.* 2009; 5:2583–2588. [PubMed: 21857812]
42. Woodcock HL 3rd, Hodoscek M, Gilbert AT, Gill PM, Schaefer HF 3rd, Brooks BR. Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations. *J Comput Chem.* 2007; 28:1485–502. [PubMed: 17334987]

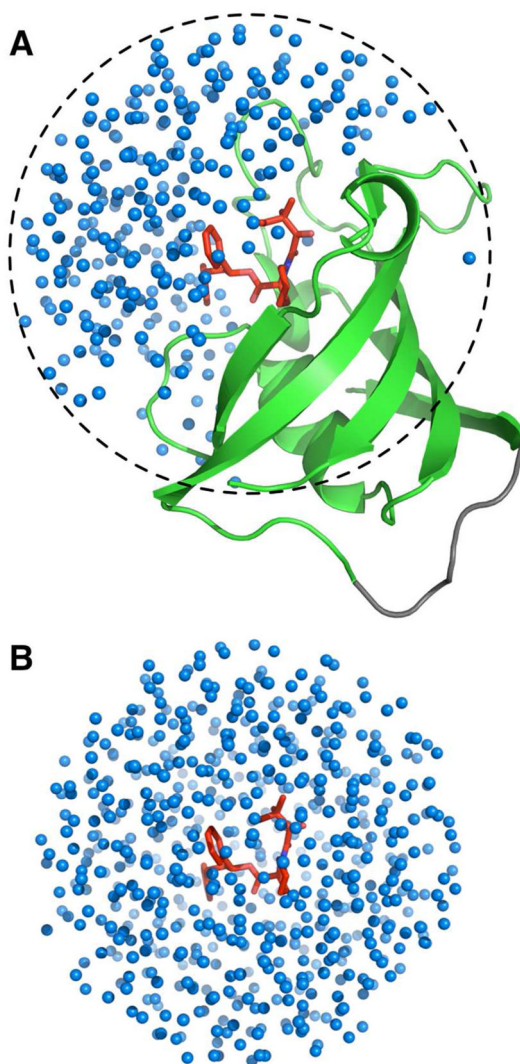


**Figure 1.** Illustration of translational and orientational restraints on SB3 ligand. Anchoring atoms are represented as grey sphere.  $r(P_1-L_1)$ ,  $\theta(\angle P_2P_1L_1)$ ,  $\phi(\angle P_3P_2P_1L_1)$  are defined for the translational restraint, and  $\alpha(\angle P_1L_1L_2)$ ,  $\beta(\angle P_2P_1L_1L_2)$ , and  $\gamma(\angle P_1L_1L_2L_3)$  are defined for the orientational restraint.



**Figure 2.** Structures of ligand molecules used in this work. (A) Benzene, (B) Indole, (C) *n*-butylbenzene, (D) 1,3-diphenyl-1-propyl-1-(3,3-dimethyl-1,2-dioxypentyl)-2-piperidine carboxylate (SB3), (E) 1-cyclohexyl-3-diphenyl-1-propyl-1-(3,3-dimethyl-1,2-dioxypentyl)-2-piperidine carboxylate (SBX), and (F) K506 (FK5). Symmetric groups detected by *Ligand Binder* are colored in red.





**Figure 3.** Reduced simulation systems for (A) *SITE* and (B) *BULK* using GSBP and SSBP boundary potentials.

**Table 1**Important input and output files generated by *Ligand Binder* and during the FEP/MD simulations.

		Filename	Description
STEP1	Input	step1_pdbreader.inp	To read a protein-ligand complex structure
		Output	step1_pdbreader.pdb
		@lig/@lig.rtf	Ligand topology file
		@lig/@lig_g.rtf	Ligand topology file (with atom groups)
		@lig/@lig.prm	Ligand parameter file
	@lig/ndihe.str	Symmetric groups found in the ligand if any	
STEP2	Input	step2.1_site_solvent.inp	To solvate the binding site with a water sphere ( <i>SITE</i> )
		step2.2_bulk_solvent.inp	To solvate the ligand with a water sphere ( <i>BULK</i> )
	Output	step2.1_site_solvent.pdb	Solvated simulation system for <i>SITE</i>
		step2.2_bulk_solvent.pdb	Solvated simulation system for <i>BULK</i>
STEP3	Input	step3.1_gsbp_setup.inp	To setup GSBP simulation system for <i>SITE</i>
		step3.2_ssbp_setup.inp	To setup SSBP simulation system for <i>BULK</i>
		config.py	Contains several (user-defined) configuration variables
		step3_job.pbs	Example PBS script for batch system
	Output	step3.1_gsbp_setup.pdb	<i>SITE</i> simulation system after GSBP setup
		step3.1_gsbp_setup.mij	Calculated multipolar reaction field
		step3.1_gsbp_setup.phix	Calculated external reaction field
		step3.2_ssbp_setup.pdb	<i>BULK</i> simulation system after SSBP setup
STEP4	Input	step4.1_site_gcmc.inp	To adjust the number of water molecules in GSBP
		step4.2_site_equil.inp	To equilibrate <i>SITE</i> system
		step4.3_restraint_setup.inp	To setup anchoring atoms and restraint parameters
		step4.3_restraint_geo.str	To determine anchoring atoms
		step4.3_restraint_geotest.str	To test validity of anchoring atoms
		step4.3_restraint_ref.str	To determine the restraint parameters
		step4.4_bulk_equil.inp	To equilibrate <i>BULK</i> system
		step4_job.pbs	Example PBS script for batch system
	Output	step4.1_site_gcmc.crd	<i>SITE</i> system after adjustment of number of water
		step4.2_site_equil.dcd	Trajectory of <i>SITE</i> system equilibration
		step4.2_site_equil.crd	Last snapshot of equilibrated <i>SITE</i> system
		step4.3_restraint_geo.prm	Anchoring atom definition
		step4.3_restraint_geo.pdb	Anchoring atoms (named as DUM)
	step4.3_restraint_ref.prm	Reference values for restraint potential	
	step4.3_restraint_consihe.str	Symmetric group restraints	
	step4.3_restraint_ligave.pdb	Average structure of bound ligand during equilibration	
	step4.4_bulk_equil.pdb	Last snapshot of equilibrated <i>BULK</i> system	

		<b>Filename</b>	<b>Description</b>
STEP5	Input	step5.1_site_fes.inp	FEP/MD input for <i>SITE</i> system (including PMF input)
		step5.1_site_gconst.str	To setup orientational/translational restraints
		step5.2_bulk_fes.inp	FEP/MD input for <i>BULK</i> system (including PMF input)
		step5_jobmanager.com	Script for submitting batch jobs
		step5_template.pbs	Example PBS script for batch system
STEP6	Input	step6.1_wham_fep.inp	WHAM input for FEP/MD simulations
		step6.1_wham_fep.pbs	Example PBS script for batch system
		step6.2_wham_rmsd.inp	WHAM input for PMF umbrella sampling simulations
		step6.2_wham_rmsd.pbs	Example PBS script for batch system
		step6_wham.com	Script for WHAM calculation
		step6_table.py	Example script to tabulate the binding free energy

**Table 2**

The coupling parameters in the FEP/MD calculations.

$\lambda_{rep}$	0.0	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0					
$\lambda_{disp}$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0				
$\lambda_{elec}$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0				
$\lambda_{r,r}$	0.0	0.025	0.005	0.0075	0.01	0.02	0.04	0.06	0.08	0.1	0.2	0.4	0.6	0.8	1.0

Table 3

Simulation parameters for each FEP/MD simulation system.

Ligand	PDB ID	$R_{\text{lig}}$	$R_{\text{wat}}^{\text{site}}$	$R_{\text{wat}}^{\text{bulk}}$	$n_{\text{wat}}^{\text{site}}$	$n_{\text{wat}}^{\text{bulk}}$
benzene	181L	2.3 Å	14 Å	13 Å	46 (155)	288
indole	185L	3.3 Å	14 Å	14 Å	42 (150)	351
n-butylbenzene	186L	4.1 Å	14 Å	15 Å	45 (152)	439
SB3	1FKG	6.6 Å	14 Å	17 Å	130 (208)	630
SBX	1FKH	6.6 Å	14 Å	17 Å	128 (208)	628
FK5	1FKJ	8.1 Å	14 Å	19 Å	147 (235)	865

$n_{\text{wat}}^{\text{site}}$  and  $n_{\text{wat}}^{\text{bulk}}$  are the numbers of water molecules in *SITE* and *BULK* systems, respectively. The numbers in the parenthesis is the number of water molecules in *SITE* system after 20 cycles of GCMC/MD simulations (averaged over three independent simulations).

Table 4

Binding free energies<sup>§</sup> of T4 lysozyme ligands and FKBP ligands.

Ligand	# <sup>§</sup>	$\Delta G_{\text{int}}^{\text{site}}$	$\Delta G_{\text{int}}^{\text{bulk}}$	$\Delta \Delta G_c$	$\Delta \Delta G_{\text{r}}$	$\Delta G_b$
benzene	C1	-9.9 ± 0.2	0.5 ± 0.1	-	4.5 ± 0.4	-5.9 ± 0.4
	C2	-11.3 ± 0.1	0.4 ± 0.3	-	6.0 ± 0.5	-5.7 ± 0.6
	C3	-8.1 ± 0.2	0.5 ± 0.2	-	2.2 ± 0.3	-6.4 ± 0.4
indole	C1	-16.5 ± 0.1	-6.3 ± 0.2	-	7.6 ± 0.1	-2.7 ± 0.3
	C2	-17.2 ± 0.1	-6.3 ± 0.2	-	8.7 ± 0.1	-2.2 ± 0.2
	C3	-18.0 ± 0.2	-6.2 ± 0.5	-	9.1 ± 0.2	-2.7 ± 0.6
<i>n</i> -butylbenzene	C1	-17.2 ± 0.3	1.9 ± 0.5	-	8.8 ± 0.1	-10.2 ± 0.6
	C2	-18.8 ± 0.2	1.8 ± 0.4	-	9.3 ± 0.1	-11.4 ± 0.4
	C3	-16.7 ± 0.4	1.6 ± 0.3	-	8.5 ± 0.0	-9.8 ± 0.5
SB3	C1	-33.9 ± 0.8	-8.0 ± 1.2	2.8 ± 1.4	9.9 ± 0.0	-13.3 ± 2.0
	C2	-33.5 ± 1.3	-8.5 ± 1.3	3.0 ± 1.1	9.9 ± 0.0	-12.1 ± 2.0
	C3	-34.8 ± 1.6	-9.1 ± 1.6	2.5 ± 1.0	9.2 ± 0.1	-14.0 ± 2.3
SBX	C1	-32.5 ± 1.1	-7.3 ± 0.7	4.9 ± 1.3	9.6 ± 0.1	-10.4 ± 1.8
	C2	-32.5 ± 1.3	-7.8 ± 1.0	5.6 ± 1.3	9.9 ± 0.0	-9.2 ± 2.0
	C3	-33.3 ± 1.7	-6.5 ± 2.1	5.3 ± 1.0	10.0 ± 0.0	-11.5 ± 2.8
FK5	C1	-60.0 ± 1.3	-35.0 ± 1.9	2.3 ± 0.7	9.8 ± 0.0	-12.9 ± 2.4
	C2	-59.4 ± 1.3	-35.3 ± 1.1	3.2 ± 0.9	10.1 ± 0.1	-10.8 ± 1.9
	C3	-59.3 ± 2.0	-34.4 ± 1.3	4.2 ± 1.2	9.9 ± 0.1	-10.7 ± 2.9

<sup>§</sup>Units of free energies are kcal/mol. The errors of all the free energy values are standard deviation of the last five cycles of FEP/MD, each started with the last snapshot of previous cycle and random velocities.

<sup>§</sup>C1, C2, and C3 represent independent FEP/MD simulations for each ligand.

**Table 5**

Comparison of binding free energies of T4 lysozyme ligands and FKBP ligands with experiments<sup>36, 37</sup> and previous calculations.

Ligand	$\Delta G_b$ (kcal/mol)			
	This work <sup>§</sup>	Deng et. al. <sup>12</sup>	Wang et. al. <sup>14</sup>	Experiments <sup>36, 37</sup>
benzene	$-6.0 \pm 0.3$	$-6.0 \pm 0.2$	-	$-5.2 \pm 0.2$
indole	$-2.5 \pm 0.2$	$-4.2 \pm 0.2$	-	$-4.9 \pm 0.1$
<i>n</i> -butylbenzene	$-10.5 \pm 0.5$	$-8.7 \pm 0.4$	-	$-6.7 \pm 0.0$
SB3	$-13.1 \pm 0.8$	-	$-10.3 \pm 0.4$	$-10.9 \pm 0.1$
SBX	$-10.4 \pm 1.0$	-	$-11.7 \pm 1.0$	$-11.1 \pm 0.2$
FK5	$-11.5 \pm 1.0$	-	$-10.1 \pm 1.2$	$-12.7 \pm 0.2$

<sup>§</sup>The free energies for this work are calculated by taking averages from the values in Table 4, and the errors are standard errors from the 3 independent calculations (C1, C2, and C3).

**Table 6**

Comparison of free energy terms<sup>¶</sup> with different selections of the anchoring atoms.

Ligand	# <sup>§</sup>	$\Delta G_{\text{int}}^{\text{site}} + \Delta \Delta G_{\text{t,r}}$
benzene	C1	$-5.4 \pm 0.3$
	C2	$-5.3 \pm 0.4$
	C3	$-5.9 \pm 0.2$
indole	C1	$-9.0 \pm 0.1$
	C2	$-8.5 \pm 0.1$
	C3	$-8.8 \pm 0.2$
<i>n</i> -butylbenzene	C1	$-8.4 \pm 0.2$
	C2	$-9.5 \pm 0.2$
	C3	$-8.2 \pm 0.3$
SB3	C1	$-24.0 \pm 0.5$
	C2	$-23.6 \pm 1.4$
	C3	$-25.7 \pm 1.0$
SBX	C1	$-22.7 \pm 0.7$
	C2	$-22.3 \pm 0.8$
	C3	$-23.0 \pm 1.1$
FK5	C1	$-50.6 \pm 0.8$
	C2	$-49.2 \pm 1.0$
	C3	$-49.4 \pm 1.5$

<sup>¶</sup>Units of free energies are kcal/mol. The errors of all the free energy values are standard deviation of the last five cycles of FEP/MD, each started with the last snapshot of previous cycle and random velocities.

<sup>§</sup>C1, C2, and C3 represent independent FEP/MD simulations for each ligand.