

# Reduced Free Energy Perturbation/Hamiltonian Replica Exchange Molecular Dynamics Method with Unbiased Alchemical Thermodynamic Axis

Wei Jiang<sup>\*,§</sup>, Jonathan Thirman,<sup>§</sup> Sunhwan Jo,<sup>§,£</sup> Benoît Roux<sup>\*,¶</sup>

Computational Science Division, Argonne National Laboratory, 9700 South Cass Avenue,  
Building 240, Argonne, Illinois 60439, USA

Department of Biochemistry and Molecular Biology, Gordon Center for Integrative Science,  
University of Chicago, 929 57th Street, Chicago, Illinois 60637, USA

\*Corresponding authors e-mail: [wjiang@alcf.anl.gov](mailto:wjiang@alcf.anl.gov) [roux@uchicago.edu](mailto:roux@uchicago.edu)

§Argonne National Laboratory

¶University of Chicago

£Currently in SilcsBio, LLC

**Abstract:** Replica-exchange molecular dynamics (REMD) has been proven to efficiently improve the convergence of free energy perturbation (FEP) calculations involving considerable reorganization of their surrounding. We previously introduced the FEP/( $\lambda$ ,H)-REMD algorithm for ligand binding, in which replicas along the alchemical thermodynamic coupling axis  $\lambda$  were expanded as a series of Hamiltonian boosted replicas along a second axis to form a two-dimensional (2D) replica-exchange exchange map [Jiang, W.; Roux, B., *J. Chem. Theory Comput.* **2010**, 6 (9), 2559-2565]. Aiming to achieve a similar performance at a lower computational cost, we propose here a modified version of this algorithm in which only the end-states along the alchemical axis are augmented by boosted replicas. The reduced FEP/( $\lambda$ ,H)-REMD method with one-dimensional (1D) unbiased alchemical thermodynamic coupling axis  $\lambda$  is implemented on the basis of generic multiple copy algorithm (MCA) module of the biomolecular simulation program NAMD. The flexible MCA framework of NAMD enables a user to design customized replica-exchange patterns through Tcl scripting in the context of a highly parallelized simulation program without touching the source code. Two Hamiltonian tempering boosting scheme were examined with the new algorithm: a first one based on potential energy rescaling of a pre-identified “solute”, and a second one via the introduction of flattening torsional free energy barriers. As two illustrative examples with reliable experiment data, the absolute binding free energies of *p*-xylene and *n*-butylbenzene to the nonpolar cavity of the L99A mutant of T4 lysozyme were calculated. The tests demonstrate that the new protocol efficiently enhances the sampling of torsional motions for backbone and side chains around the binding pocket and accelerates the convergence of the free energy computations.

## Introduction

A treatment of binding affinity of small ligands to macromolecules is of central importance in biology and pharmacology, and accurate computational prediction methods could be of great practical value.<sup>1-7</sup> Free energy perturbation molecular dynamics (FEP/MD) with explicit solvent molecules provide one of the most fundamental routes for computing the binding affinities of small molecules to macromolecules.<sup>8-12</sup> Relative alchemical free energy calculations can be a very practical guide in computer aided drug design,<sup>13-15</sup> while absolute binding free energy calculations, despite its higher computing cost, can provide deep insight into molecular recognition mechanisms.<sup>16-18</sup> In molecular recognition problems, a critical issue is achieving a sufficient sampling when there is large structural reorganization – either within the ligand or the protein – accompanying the formation of the bound complex. The incomplete configurational sampling gives rise to the so-called “Hamiltonian lagging” problem identified in the early days of free energy simulations,<sup>19</sup> resulting in computed binding free energies that are dependent on the starting protein or ligand configuration and of limited practical use.

In the past decade, intense efforts were spent on the development of computational strategies to help sample more effectively the slow structural reorganization coupled to the alchemical transformation in absolute ligand binding free energy calculations. One strategy exploited by Roux and co-workers to improve convergence is by combining the alchemical FEP calculations with umbrella sampling potential of mean force (PMF) simulations along some pre-identified degree of freedom.<sup>8-9, 20-21</sup> Dill and co-workers used a similar strategy by calculating the PMF along the  $\chi_1$  dihedral to treat the effect of a side chain rotamer on ligand binding in T4 Lysozyme.<sup>12</sup> An important drawback of such a PMF-based strategy is the need to pre-identify the specific degree of freedom thought to be relevant. This can be difficult in practice. For this reason, it has been necessary to seek alternative approaches allowing one to boost the sampling of multiple degrees of freedom. One strategy introduced by Jiang and Roux<sup>22</sup> is the FEP/( $\lambda$ ,H)-REMD scheme for ligand binding, in which replicas along the alchemical thermodynamic coupling axis  $\lambda$  were expanded as a series of Hamiltonian-tempering boosted replicas along a second axis to form a two-dimensional (2D) replica-exchange exchange map. Free energy surface flattening or “boosting” potentials along the  $\chi_1$  dihedral angles of the side chains lining the binding pocket were applied to all alchemical  $\lambda$ -windows to cancel out the intrinsic energy barrier opposing relevant

conformational transitions that are difficult to sample. The FEP/ $(\lambda, H)$ -REMD scheme naturally retains the proper thermodynamic Boltzmann sampling of the system along the alchemical axis while increasing the inter-conversion rates between metastable conformations of the apo and holo states via the Hamiltonian boosting axis. To reduce the computational cost associated with the extensive 2D replica-exchange, Wang et al<sup>14</sup> proposed a more compact protocol for relative free energy calculation based on a one-dimensional (1D) chain of replicas corresponding to a progression along the alchemical  $\lambda$ -axis together with a Replica Exchange Solute Tempering (REST2) boosting scheme based on a rescaling of the potential energy of a pre-identified “solute”. Subsequently, Levy et al<sup>23</sup> presented a variant of REST2 based on torsional boosting potentials.

One possible challenge with combining the alchemical and boosting axis along a single 1D chain of replica is that the large bias of the Hamiltonian tempering must cancel out between the end-points in order to yield meaningful FEP results. Such problem is avoided in the extensive 2D FEP/ $(\lambda, H)$ -REMD that keeps the alchemical and boosting separate.<sup>22</sup> Aiming to achieve a similar performance of this method but at a lower computational cost, we propose here a reduced 1D version in which the boosted replicas are attached only to the end-states along the alchemical  $\lambda$ -axis, thereby retaining an unbiased progression along the alchemical thermodynamic axis itself. In the original 2D version of FEP/ $(\lambda, H)$ -REMD, each system with a given thermodynamic coupling factor  $\lambda$  is further coupled with a set of replicas evolving on a biased energy surface with boosting potentials used to accelerate the inter-conversion among different rotameric states in the neighborhood of the binding site. Exchanges were allowed to occur alternatively along the two axes corresponding to the thermodynamic coupling parameter  $\lambda$  and the boosting potential, in an extended dual array of coupled  $\lambda$ - and H-REMD simulations. Our previous FEP/ $(\lambda, H)$ -REMD method represents an extensive 2D communication patterns requiring a large number of CPUs, which is undesirable if one wishes to reduce the computational cost of high throughput FEP simulations. A reduced FEP/ $(\lambda, H)$ -REMD method with one-dimensional (1D) unbiased alchemical thermodynamic coupling axis  $\lambda$ . A reduced 1D FEP/ $(\lambda, H)$ -REMD scheme, with high frequency  $\lambda$  exchanges along the thermodynamic axis, can nonetheless capture much of the enhanced sampling of the original 2D framework because neighboring  $\lambda$ -windows quickly relay accelerated configurations from boost replicas. Recent work indicates that with high frequency  $\lambda$  exchanges, any given replica thoroughly samples the space accessible to the ensemble in a timescale of hundreds of picoseconds.<sup>24</sup>

Therefore, boost replicas only need to be added to the end (apo and/or holo) states in binding free energy calculations to provide accelerated configurations for all alchemical windows during high frequency  $\lambda$  exchanges.

We illustrate the reduced 1D FEP/ $(\lambda, H)$ -REMD scheme to calculate the absolute binding free energies of large aromatic molecules to the nonpolar cavity of the L99A mutant of T4 lysozyme. The nonpolar binding site has been studied extensively experimentally and computationally by various free energy methods and docking.<sup>9, 22, 25-26</sup> It is accepted that the binding site undergoes a significant structural reorganization upon the binding insertion of large ligands. The F-helix (residue 107-115) is directly implied in this reorganization, via a rotameric transition of the Val111 side chain from a trans conformation for the ligand-free apo state to a gauche conformation for the bound state state. Recent experimental measurement and simulation study further categorizes structural reorganization of the F-helix into three discrete conformations – closed, intermediate, and open – in response to different ligand size.<sup>25</sup> Among those aromatic ligands that have reliable crystal structures, p-xylene and n-butylbenzene induce discrete rotameric state changes of Val111. N-butylbenzene also induces measurable backbone reorganization that makes an intermediate state relative to benzene bound state. These two binding complexes have accurate binding free energy values, which are indispensable for reliable methodology evaluation. It is demonstrated that the reduced FEP/ $(\lambda, H)$ -REMD protocol significantly accelerates sampling of target binding site and convergence of FEP calculations with much less computing resource than previous version. In the following sections, FEP/ $(\lambda, H)$ -REMD refers to alchemical FEP with boosting potentials while FEP/ $\lambda$ -REMD refers to general alchemical  $\lambda$ -exchange without boosting potential.

## Computational Details

### A. Generic Multiple Partition Implementation of Soft-Core FEP/H-REMD Simulation Protocol

In the FEP staging simulation protocol,<sup>9, 27</sup> the potential energy is expressed in terms of three thermodynamic coupling parameters  $\lambda_{LJ}, \lambda_{elec}, \lambda_{rstr} \in [0, 1]$ , used to control the Lennard-Jones (LJ), the electrostatic and restraining potential contributions,

$$U(\lambda_{LJ}, \lambda_{elec}, \lambda_{rstr}) = U_0 + \lambda_{LJ}U_{LJ} + \lambda_{elec}U_{elec} + \lambda_{rstr}U_{rstr} \quad (1)$$

where  $U_0$  is the potential of the system with the non-interacting (decoupled) ligand,  $U_{\text{LJ}}$  is the soft-core form of Lennard-Jones potential of soft-core form,  $U_{\text{elec}}$  is the electrostatic interaction, and  $U_{\text{rstr}}$  is the restraining potential. Relying on the multiple partition module of charm++/NAMD,<sup>24</sup> a multi-stage FEP simulation protocol can be implemented in a single parallel/parallel job. The free energy corresponding to the process of completely inserting the ligand into the binding site,

$$U(\lambda_{\text{LJ}}=0, \lambda_{\text{elec}}=0, \lambda_{\text{rstr}}=1) \rightarrow U(\lambda_{\text{LJ}}=1, \lambda_{\text{elec}}=1, \lambda_{\text{rstr}}=0) \quad (2)$$

is decomposed in 3 stages,  $U(\lambda_{\text{LJ}}=0, \lambda_{\text{elec}}=0, \lambda_{\text{rstr}}=1) \rightarrow U(\lambda_{\text{LJ}}=1, \lambda_{\text{elec}}=0, \lambda_{\text{rstr}}=1) \rightarrow U(\lambda_{\text{LJ}}=1, \lambda_{\text{elec}}=1, \lambda_{\text{rstr}}=1) \rightarrow U(\lambda_{\text{LJ}}=1, \lambda_{\text{elec}}=1, \lambda_{\text{rstr}}=0)$ . In principle, the last stage corresponding to the decoupling of the restraining potential could be included as part of the replica-exchange alchemical process, but it is more advantageous to execute this stage as a separate task to maintain the load balance as the replicas with fully interacting ligand ( $\lambda_{\text{LJ}}=1$  and  $\lambda_{\text{elec}}=1$ ) run faster than those with partial alchemical coupling ( $\lambda_{\text{LJ}}<1$  and  $\lambda_{\text{elec}}<1$ ). To help achieve a significant sampling enhancement, a 1D chain of  $M$  replicas with different strength of “boosting” biasing potentials is appended to the restrained apo ( $\lambda_{\text{LJ}}=0, \lambda_{\text{elec}}=0, \lambda_{\text{rstr}}=1$ ) and holo state ( $\lambda_{\text{LJ}}=1, \lambda_{\text{elec}}=1, \lambda_{\text{rstr}}=1$ ) of the FEP/ $\lambda$ -REMD calculation.<sup>28</sup> The boosting parameter  $b$  scales the biasing potential (the system is not biased when  $b=1$ , and maximally boosted when  $b=0$ ). In the Tcl script for refined FEP/H-REMD protocol, each replica parameter set has now two coupling parameters,  $\lambda_i$  and  $b_i$ , with  $b_i=1$  for all the replicas between the end states of the alchemical transformation corresponding to Eq. (2). High frequency replica-exchange are performed along this one-dimensional (1D) chain of states according to the conventional Metropolis Monte Carlo exchange criterion.<sup>29</sup>

$$P(\lambda_i, b_i \rightarrow \lambda_j, b_j) = \min \left\{ 1, e^{-[U(\lambda_i, b_i, \mathbf{r}_i) + U(\lambda_j, b_j, \mathbf{r}_j) - U(\lambda_i, b_i, \mathbf{r}_j) - U(\lambda_j, b_j, \mathbf{r}_i)] / k_B T} \right\} \quad (3)$$

where  $U$  denotes the potential energy of all the underlying replica, and  $\lambda_i$  and  $\lambda_j$  denote the window parameters and  $b_i$  and  $b_j$  denote the boosting parameters.

## B. Boosting Potentials for Slow Degrees of Freedom

Effective boosting potentials can be constructed with Potential of Mean Force (PMF) calculations of mini peptides,<sup>22, 30</sup> or utilize those of other accelerated MD methods.<sup>31-32</sup> However, such free energy surface flattening

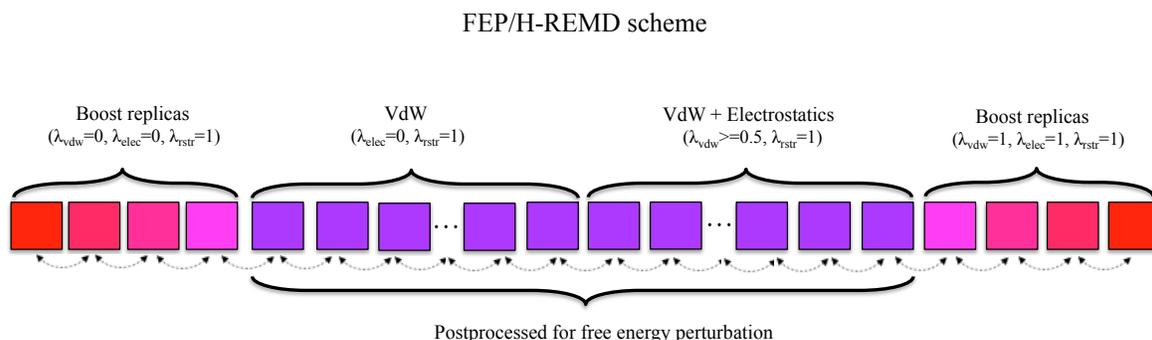
potentials might suffer from transferability problems. Alternatively, one can employ transferable boost potentials employed in Replica Exchange Solute Tempering (REST2),<sup>33-34</sup> which is already implemented in multiple MD packages. It needs to be noted that FEP formalism is grounded in an unbiased integral (summation) along the alchemical reaction path, requiring a  $\lambda$ -window with boost potential to be reweighted for a strict free energy calculation. However, a reweighting process of accelerated simulation is nontrivial and significant analysis is required,<sup>35-36</sup> especially true for absolute free energy calculation where immensely complex energy barriers lie between apo and holo state. Similar with our previous protocol, in the free energy postprocessing phase, only  $\lambda$ -windows with zero bias are taken into account to avoid reweighting complexities. In the following, two Hamiltonian tempering scheme are examined chosen for their availability in the software NAMD: potential energy rescaling (REST2-like) and torsional flattening potential. The whole residue Val111 with/without neighboring residues are selected as accelerated region. The sampling of rotameric states of Val111 is quantified and analyzed, and binding free energies are calculated.

Potential energy rescaling of selected degrees of freedom has ideal transferability and has been implemented in NAMD. All of the replicas are run at the same temperature but the potential energy for each replica is scaled differently:

$$E_m^{\text{REST2}}(X) = \frac{\beta_m}{\beta_0} E_{ss}(X) + \sqrt{\frac{\beta_m}{\beta_0}} E_{sw}(X) + E_{ww}(X) \quad (4)$$

The parameter swaps of these boosted replicas form a REST2 protocol,<sup>33-34</sup> where the subscript “s” denotes selected degrees of freedom while the subscript “e” denotes the remaining environment atoms. Essentially the charge and LJ parameter of each atom of hot region are rescaled by a factor of  $\sqrt{\beta_m / \beta_0}$  and related bonded terms are rescaled correspondingly. Alternatively, flattening potentials for torsional motions of backbones and side chains can be fitted through PMF calculations of mini peptides.<sup>22, 30</sup> Such a free energy based flattening potential contains significant entropy contributions and is expected to be more efficient than potential energy rescaling/barrier flattening. In practice, the initial structure of a FEP simulation can be either an experimentally available crystal structure or a docked structure. Along the alchemical reaction path, the first FEP window demands the largest timescale to finish structural relaxation from holo to apo state. On the other side, the last  $\lambda$ -window could still need a large timescale to overcome hidden barriers. Indeed, in a docked structure (holo state),

multiple torsional degrees of freedom at binding site might be problematic and need special sampling acceleration. In order to draw a general solution for orthogonal sampling enhancement, boosting replicas are applied to both the first and the last FEP window, as illustrated in Figure 1.



**Figure 1:** Implementation of the reduced FEP/ $(\lambda, H)$ -REMD method with 1D unbiased alchemical thermodynamic coupling implemented within the charm++ multiple partition module. Each square box represents an FEP/MD simulation with its own trajectory. A branch of four boosting-biasing replica (red) is attached to each of the two end FEP windows along the reversible work process, forming an extended thermodynamic axis. The biasing strength of boosting replicas linearly increases outward along the thermodynamic axis, illustrated with varying chroma of red color. The possible attempted moves, indicated by the dashed-line arrows, are only allowed between neighboring replicas. It needs to be noted that during the postprocessing phase only the outputs generated from the normal FEP windows (blue) are processed.

### C. FEP/MD Simulations

The generic multi-partition module of NAMD developed to support a wide range of Multiple Copy Algorithms (MCA)<sup>24</sup> allows the user to arbitrarily design customized replica-exchange patterns via the Tcl scripting interface without touching the source code. The mixed exchanges with alchemical  $\lambda$  and boosting parameter  $b_i$  can be straightforwardly implemented through Tcl scripting, reducing the original 2D replica exchange pattern to a 1D chain. All the FEP simulations for binding site were carried out on the IBM Blue Gene/Q cluster Mira of the Leadership Computing Facility (LCF) at Argonne National Laboratory (ANL). The simulations were carried out in a high performance mode with version 2.10 of the NAMD program,<sup>37</sup> which was modified and extended for the present study. The binding site free energy simulations were carried out with periodic boundary conditions at constant pressure. The initial T4L/L99A binding systems were constructed from the

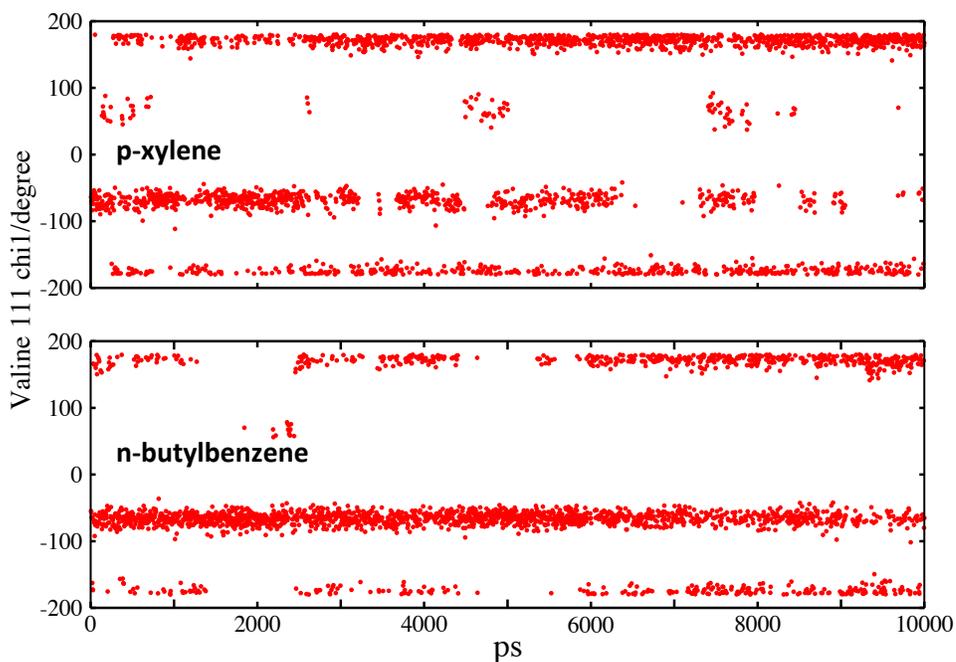
crystallographic structure (PDB 186L and 187L) with CHARMM-GUI Ligand Binder.<sup>38-39</sup> The disulfide bond was retained and all crystallographic ions and waters were kept. Counter ions were randomly positioned by CHARMM-GUI within solvent. TIP3P water model was used along with RATTLE algorithm for geometry restraint. The systems were propagated with a 2 fs time step using Langevin dynamics (collision frequency 2/ps) at a temperature of 300K. Constant NPT ensemble was maintained with Langevin piston pressure control.<sup>40</sup> For the binding free energy calculations, 200 ps production runs were performed for the binding site with a replica-exchange frequency of 1/80 steps. Generalized CHARMM force field parameters<sup>41</sup> of *p*-xylene and *n*-butylbenzene were constructed through CHARMM-GUI Ligand Binder.

In the FEP/ $\lambda$ -REMD simulation, 24  $\lambda$ -windows were employed with general soft-core potentials<sup>42-43</sup> applied to LJ interactions to avoid ‘endpoint catastrophe’. In a testing run of 200 ps with FEP/ $\lambda$ -REMD, the exchange acceptance ratio distributions among the 23 neighboring replica pairs range from 50% - 70%. The electrostatic interaction is turned on at  $\lambda$  value 0.5. In addition to the alchemical  $\lambda$ -axis, four boosting replicas are coupled with the apo state and holo state respectively, as shown in Figure 1. The four boosting replicas of each end state adopt identical  $\lambda$  values with the host end state and linearly increasing biasing strengths. In REST2, the rescaling strength is from 0.8125 to 0.25 while the torsional flattening strength is from 0.25 to 1.0. The flattening potential of torsional motions for side chains and backbone is from our previous work<sup>22</sup> and ref [22]. In all calculations, the exchange attempt history and corresponding potential energy evaluation of Metropolis Monte Carlo trial move of each replica energies were collected during the production run, and post-processed using both the Simple Overlap Sampling (SOS)<sup>44</sup> and WHAM.<sup>45</sup> To avoid reweighting complexities of boosting replicas, the SOS and WHAM post-processing are only applied to the 24 FEP windows. To monitor the convergence of the binding site calculation, 50 consecutive FEP calculations ( $50 \times 200$  ps) were performed for each system, each starting from the configuration saved at the end of the previous run. The data generated during the last 25 FEP calculations are used for binding free energy calculation and data analysis.

## Results and Discussion

Both crystallographic studies and computations indicate that the side chain of Val111 of T4L/L99A changes its rotameric states when moderately large ligands bind to the nonpolar cavity.<sup>26</sup> The Val111 residue is located in

the middle of the F-helix and has most direct steric contact with the nonpolar tail of aromatic molecule. In the case of *p*-xylene and *n*-butylbenzene, the side chain of Val111 experiences rotameric conversion from *gauche* to *trans* state, to avoid a steric clash with the ligand. In absolute binding free energy calculations, it is convenient to utilize the *holo* configuration as the starting set of coordinates is provided either by the X-ray crystallographic structure of the bound complex or by the output of *in silico* docking. It should be noted that intentional employment of a ‘bad’ initial structure, such as apo configuration for all  $\lambda$ -windows, can make large uncertainties for a quick evaluation of a sampling enhancement method.<sup>25</sup> In practice, unpredictably long MD trajectories have to be generated to remove any trace of a bad initial configuration, especially for large ligands that possess multiple native binding modes. However, one must ensure that the sampling underlying the FEP calculation does reversibly cover the relevant set of thermodynamic states. As shown in Figure 2, within 3 ns

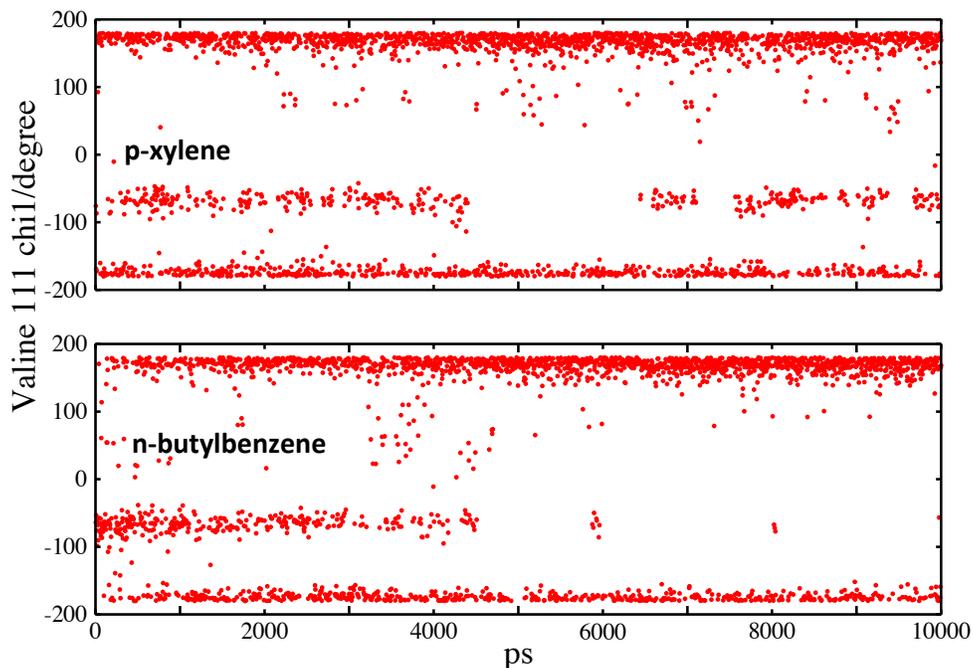


**Figure 2:** Populations of rotameric state of Val111 in the first FEP window, obtained with FEP/ $\lambda$ -REMD. The two binding complexes exhibit distinctly different *holo*-apo state transition behavior in both time scale and populations.

timescale, in the first FEP window sparse spontaneous dihedral transition is observed in the 1D FEP/ $\lambda$ -REMD simulations. The side chain remains kinetically “trapped” in its *holo* rotameric state, with  $\chi_1$  around  $-60^\circ$ . However, the two binding complexes exhibit distinctly different *holo*-apo transition time scale and rotameric populations. In the first FEP window, the time series of  $\chi_1$  for the *p*-xylene simulation exhibits gradually

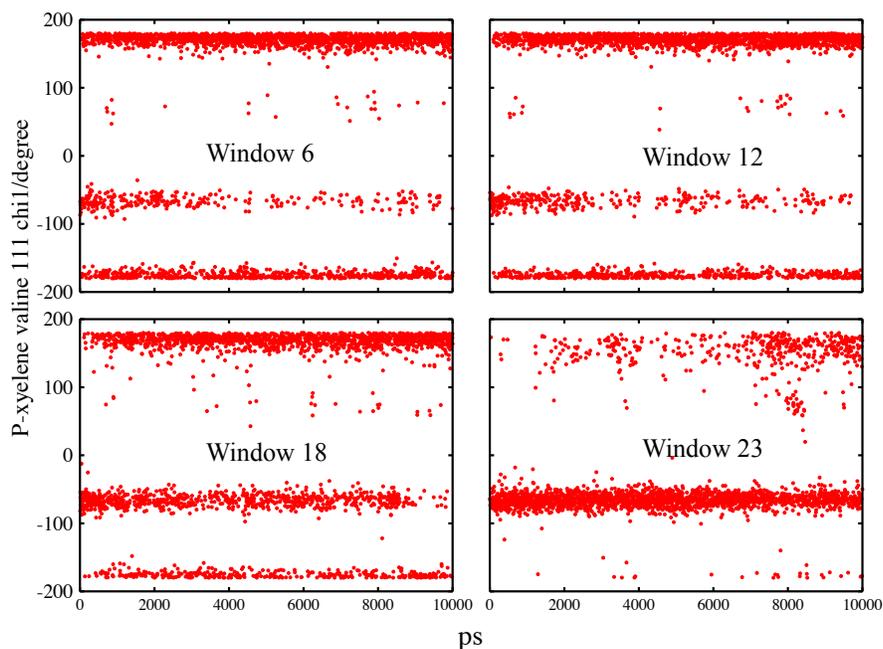
increasing number of *holo-apo* state transitions after time scale 3 ns while that for n-butylbenzene retains a low transition rate during the whole simulation time scale, indicating a higher energy barrier due to concerted backbone reorganization of the F-helix. In standard FEP/ $\lambda$ -REMD simulations, the binding free energy for p-xylene and n-butylbenzene is  $-6.1 \pm 0.3$  and  $-8.1 \pm 0.4$  kcal/mol, respectively. Those values are too favorable compared with experiment results,  $-4.7$  and  $-6.7$  kcal/mol.<sup>26</sup>

To address the issue of a kinetically trapped degree of freedom, the reduced 1D FEP/( $\lambda$ ,H)-REMD framework is introduced, with the boosting potential applied exclusively to the  $\chi_1$  degree of freedom of Val111, the most problematic residue. In REST2, potential energy was rescaled for all atoms of Val111 while the torsional flattening potential was only applied to the  $\chi_1$  dihedral angle. Four biasing replicas on the first and last window were used to guarantee an exchange acceptance ratio of at least 25% between the replicas with adjacent values of the boosting parameter  $b$ . Calculated binding free energies of  $-5.1 \pm 0.4$  and  $-5.2 \pm 0.5$  kcal/mol are obtained for p-xylene using with REST2 and torsional flattening potential, respectively. Similarly, binding free energies of  $-7.4 \pm 0.6$  and  $-7.1 \pm 0.5$  kcal/mol are obtained for n-butylbenzene with REST2 and torsional flattening potential respectively, in good agreement with the experimental value. It should also be noted that a high acceptance ratio of about 60% was obtained among all boosting replica pairs with a torsional flattening potential, benefiting from much fewer degrees of freedom accelerated than REST2. This observation indicates that well fitted flattening potential targeting judiciously chosen degrees of freedom may be a promising avenue to enhance the sampling of proteins binding sites. In practice a calculated free energy in good agreement with the experiment value is, by itself, merely one of several possible performance metrics, as it could be fortuitous within the complex of sampling efficiency, force field quality and simulation methodology. A deeper insight can be obtained by considering the time evolution of the  $\chi_1$  dihedral of Val111 of the two binding complexes, as illustrated in Figure 3. In the FEP/(H,  $\lambda$ )-REMD simulations, for p-xylene binding complex, the  $\chi_1$  of Val111 rapidly starts to make transitions within  $\sim 100$  ps toward  $180^\circ$  while for n-butylbenzene the same rotameric state transition occurs at time scale  $\sim 2$  ns, corresponding to the dominant rotamer for the *apo* state.

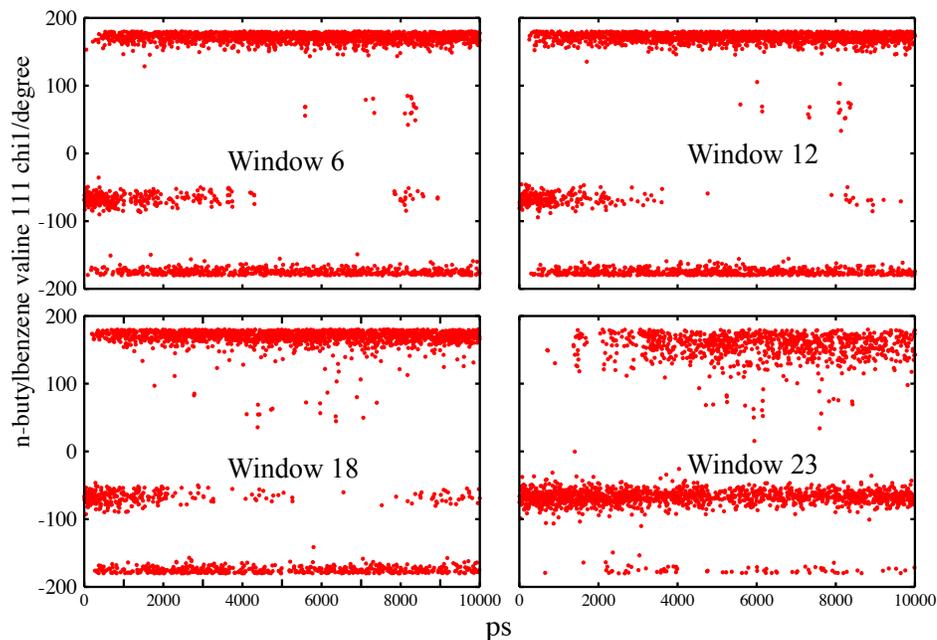


**Figure 3:** Enhanced sampling of rotameric state of Val111 in the first FEP window, obtained with FEP/H-REMD. Note that the n-butylbenzene exhibits apparently higher population of dominant state at production time scale.

Moreover, the time-evolution of  $\chi_1$  for the *holo* state (window #24) fluctuates predominantly around  $-60^\circ$ , with some excursions to other values. These observations are in agreement with our previous calculations with FEP/REST2 method.<sup>34</sup>



**Figure 4:** Population progression of rotameric state of Val111 along thermodynamic axis, p-xylene binding complex.



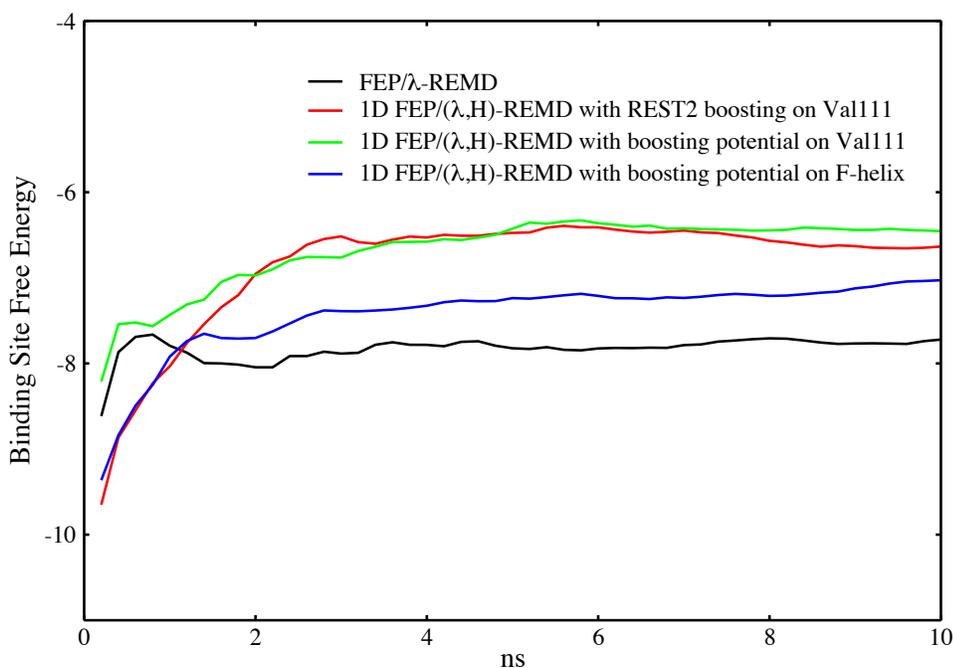
**Figure 5:** Population progression of rotameric state of Val111 along thermodynamic axis, n-butylbenzene binding complex.

In Figure 4 and Figure 5, it can be seen that along the alchemical coupling axis ( $\lambda$ ), the population of rotamers changes progressively from the appropriate distribution of the *apo* and *holo* states. For the first FEP window of p-xyelene binding complex, the average populations for *trans*, *gauche*<sup>+</sup>, and *gauche*<sup>-</sup> are 0.86, 0.03, and 0.11, respectively. These results are in good accord with the values estimated from the PMF of Dill and co-workers, and our previous results with a 2D version of FEP/(H,  $\lambda$ )-REMD.<sup>46</sup> For the *holo* state, the average populations of the *trans*, *gauche*<sup>+</sup>, and *gauche*<sup>-</sup> rotamers are 0.27, 0.03, and 0.70, respectively,, in good accord with the values estimated from the PMF of Dill and co-workers and our previous result. A similar distribution of rotameric populations is obtained with REST2 (shown in Supplementary Information). For both the *apo* and *holo* state, the probability of the dominant state is reproduced within a few percent. Differences in the population of minor states with our previous study could be attributed to the use of a reduced simulation system with the GSBP solvent boundary potential.

The ultimate aim of a reduced 1D FEP/(H,  $\lambda$ )-REMD framework is to enable efficient binding free energy calculations without prior knowledge of the slow degrees of freedom, such as the  $\chi_1$  of Val111 in T4L/L99A. In a new round of FEP/(H,  $\lambda$ )-REMD calculations for p-xyelene, we applied indiscriminately a biasing boosting potential to all protein side chains of the F-helix. This new simulation yield a calculated

binding free energy of  $-5.4 \pm 0.5$  kcal/mol, essentially the same value that is obtained when the dihedral of Val111 was targeted directly. Correspondingly, for the first FEP window, the average populations of *trans*, *gauche*<sup>+</sup>, and *gauche*<sup>-</sup> rotameric states are 0.89, 0.01, and 0.10, respectively, in agreement with the single side chain tempering. However, with the FEP/(H,  $\lambda$ )-REMD the populations of the first FEP window of n-butylbenzene remains inconsistent with that of p-xylene, with a  $\sim 10\%$  higher population of dominant state. Intrigued by the discrepancy, we decided to expand the Hamiltonian tempering to the backbone of the entire F-helix and repeat the calculation of the binding free energy of n-butylbenzene. In this case, parallel tempering with torsional flattening potential for  $\chi_1$  of Val111 and backbones is straightforwardly applied to the whole F-helix. With a larger region boosted by Hamiltonian tempering, a more favorable binding affinity  $-7.4 \pm 0.6$  kcal/mol is obtained, in agreement with experiment. Quantitative analysis indicates that with backbone tempering of the F-helix the minor states *gauche*<sup>+/-</sup> become  $\sim 10\%$  higher population (relative to the single  $\chi_1$  tempering) in the first FEP window, resulting in an average populations for the rotameric states (0.86, 0.02, 0.12) in agreement with those for p-xylene. In contrast, Hamiltonian tempering based on the REST2 scheme applied to the entire F-helix causes rapid folding-unfolding transition of the helix. Such large distortion of the potential energy surface exaggerated binding site free energy and therefore no REST2 result is shown here. The failure of REST2 on the whole F-helix can be attributed to its ‘parameter rescaling’ nature. Differing with other general barrier flattening methods, REST2 rescales force field parameters of a selected region. As a result, the fundamental hydrophobic/hydrophilic balance of the selected tempered region might be altered if the force field parameters are rescaled aggressively. In practice, the application of Hamiltonian tempering based on a REST2 scheme, including the selection of the tempered region, must to be done very carefully. These problems point to the need for further development of improved REST2-like schemes. In the case of n-butylbenzene, the binding complex does not experience a large backbone reorganization and the F-helix retains a normal helix state.<sup>25</sup> Due to normal conformational fluctuations in an MD simulation and degeneracy of RMSD, in practice it is unreliable to use RMSD variations to judge sampling efficiency of n-butylbenzene binding site. In contrast, rotameric conversion exhibits well-defined discrete states whose populations are coupled with neighboring backbone motions, and therefore is more suitable to investigate sampling enhancement for current test cases. Perhaps the clearest assessment of the sampling performance achieved with the different methods is obtained by monitoring

the progression of the calculated free energy as a function of simulation time. In **Figure 6** is shown the cumulative average of the calculated free energy for n-butylbenzene located in the T4L binding site as a function of the sampling time from the different methods. It is observed that the cumulative free energy from the calculations based on the reduced 1D FEP/(H,  $\lambda$ )-REMD with REST2 (red curve) or with boosting potentials on Val111 (green curve) converge rapidly within about 2 ns per window, whereas the calculation based on the straight FEP/( $\lambda$ )-REMD (black curve) remains highly biased even after 10 ns per window. The calculations based on the reduced 1D method with Hamiltonian tempering boosting potentials applied to the entire F-helix also displays signs of poor convergence, most likely because the ambitious tempering distorts the F-helix structure.



**Figure 6:** Cumulative average of the calculated free energy of n-butylbenzene in the T4L binding site as a function of the sampling time using different methods.

## Conclusion

A FEP/( $\lambda$ ,H)-REMD simulation scheme, reduced to a 1D chain with an economical set of boosted replicas attached to the *apo* and *holo* end-states, was proposed to enhance the sampling of target structural reorganization in binding free energy calculations. In this modified scheme, a chain of boosted replica evolving on a biased energy surface was appended only to *apo* and *holo* end-states along the thermodynamic coupling transformation

for maximum efficiency. The benefits from the REMD scheme relies on a high frequency exchange attempt to speed up the configurational sampling in the extended ensemble. No reweighting procedure is needed as only  $\lambda$ -windows with no boosting are post-processed for the free energy calculation. In principle sampling of any residue lining the binding pocket can benefit from parallel tempering with potential energy or free energy surface flattening potentials. Application of 1D FEP/ $(\lambda, H)$ -REMD shows that the sampling of torsional motions surrounding the nonpolar cavity of T4L/L99A is significantly enhanced and that the binding free energy for large ligands such as *p*-xylene and n-butylbenzene can be calculated accurately by starting from the *holo* protein configuration. Ultimately, it is important to note that the selection of boosted degrees of freedom as well as the strength of the Hamiltonian tempering is very important, as it can manifestly affects the sampling efficiency. A computational FEP strategy based on torsional flattening potentials benefits from high acceptance of replica-exchange while retaining the fundamental physicochemical properties of accelerated region. On the other hand, a strategy based on the REST2 scheme benefits from high transferability and convenience. But the tempered region must be monitored carefully to avoid unwanted large distortions. Development of improved REST2-like schemes remains of high interest.

## **Acknowledgment**

This research is funded by NIH grant P41-GM104601 for NAMD code development and Advanced Scientific Computing Research (ASCR) program, Office of Science of the U.S. Department of Energy. Jonathan Thirman is supported by postdoctoral fellowship through Aurora Early Science Program, Office of Science of Department of Energy. This research used the ALCF resource at ANL, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357.

## **Supporting Information**

The NAMD Tcl scripts for the FEP/ $(H, \lambda)$ -REMD (both REST2 and torsional flattening potential) applications are provided in Supporting Information. It needs to be noted that REST2 already was officially released in NAMD source tree, while torsional flattening potential is to be officially released later. NAMD source code can be downloaded via <http://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=NAMD>. In

the NAMD source tree, the directory lib/replica contains other relevant FEP scripts and test examples. Replica utility scripts, including replica sorting, visualization and WHAM post-processing, are also provided. Rotameric state populations obtained with FEP/(H,  $\lambda$ )-REMD and REST2 boosting are provided in Supporting Information.

## References

1. Jorgensen, W., Efficient drug lead discovery and optimization. *Accounts of Chemical Research* **2009**, *42* (6), 724.
2. Jorgensen, W. L., The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813.
3. Gilson, M.; Given, J.; Bush, B.; McCammon, J., The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys J* **1997**, *72* (3), 1047-1069.
4. Gilson, M.; Zhou, H., Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct* **2007**, *36*, 21-42.
5. Mobley, D.; Dill, K., Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". *Structure* **2009**, *17* (4), 489-498.
6. Jorgensen, W. L., Efficient drug lead discovery and optimization. *Acc. Chem. Res.* **2009**, *42* (6), 724-733.
7. Chodera, J. D.; Mobley, D.; Shirts, M.; Dixon, R.; Branson, K.; Pande, V. S., Alchemical free energy methods for drug discovery: Progress and challenges. *Curr Opin Struct Biol.* **2011**, *21* (2), 150-160.
8. Souaille, M.; Roux, B., Extension to the weighted histogram analysis method: Combining umbrella sampling with free energy calculations. *Computer Physics Communications* **2001**, *135* (2001), 40-57.
9. Deng, Y.; Roux, B., Calculation of standard binding free energies: Aromatic molecules in the t4 lysozyme I99a mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255-1273.
10. Deng, Y.; Roux, B., Hydration of amino acid side chains: Nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. *J. Phys. Chem.* **2004**, *108*, 16567-16576.
11. Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A., Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.* **2007**, *371*, 1118-1134.
12. Mobley, D. L.; Vhodera, J. D.; Dill, K. A., Confine-and-release method: obtaining correct binding free energies in the presence of protein conformational change. *J. Chem. Theory Comput.* **2007**, *3*, 1231-1235.
13. Wang, L.; Wu, Y.; Deng, Y.; Berne, B.; Friester, R.; Abel, R., Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **2015**, *137*, 2695-2703.
14. Wang, L.; Berne, B. J.; Friesner, R. A., On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities. *PNAS* **2012**, *109* (6), 1937-1942.
15. Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A., Advancing drug discovery through enhanced free energy calculations. *Acc. Chem. Res.*, **2017**, *50* (7), 1625-1632.
16. Payne, C. M.; Jiang, W.; Shirts, M. R.; Himmel, M. E.; Crowley, M.; Beckham, G. T., Glycoside hydrolase processivity is directly related to oligosaccharide binding free energy. *J. Am. Chem. Soc.* *135* (50), 18831-18839.
17. Boulos, S. P.; Davis, T. A.; Yang, J. A.; Lohse, S. E.; Alkilany, A. M.; Holland, L. A.; Murphy, C. J., Nanoparticle-protein interactions: A thermodynamic and kinetic study of the adsorption of bovine serum albumin to gold nanoparticle surfaces. *Langmuir* **2013**, *29* (48), 14984-14996.
18. Comer, J.; Chen, R.; Poblete, H.; Vergara-Jaque, A.; Riviere, J. E., Predicting adsorption affinities of small molecules on carbon nanotubes using molecular dynamics simulation. *ACS Nano* **2015**, *9* (12), 11761-11774.
19. Pearlman, D.; Kollman, P., The lag between the Hamiltonian and the system configuration in free energy perturbation calculations. *Journal of Chemical Physics* **1989** *91* (12), 7831-7839.
20. Woo, H.; Roux, B., Calculation of absolute protein-ligand binding free energy from computer simulations. *PNAS* **2005**, *102* (19), 6825-6830.

21. Wang, J.; Deng, Y.; Roux, B., Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* **2006**, *91*, 2798-2814.
22. Jiang, W.; Roux, B., Free energy perturbation hamiltonian replica-exchange molecular dynamics (fep/h-remd) for absolute ligand binding free energy calculations *J. Chem. Theory Comput.* **2010**, *6* (9), 2559-2565.
23. Menten, A.; Deng, N.-j.; Vijayan, R. S.; Xia, J.; Gallicchio, E.; Levy, R. M., Binding energy distribution analysis method: Hamiltonian replica exchange with torsional flattening for binding mode prediction and binding free energy estimation. *J. of Chem. Theory Computat.* **2016**, *12* (5), 2459-2470.
24. Jiang, W.; Phillips, J.; Huang, L.; Fajer, M.; Meng, Y.; Gumbart, J.; Luo, Y.; Schulten, K.; Roux, B., Generalized scalable multiple copy algorithms for molecular dynamics simulations in namd. *Comput Phys Commun* **2014**, *185* (3), 908-916.
25. Lim, N. M.; Wang, L.; Abel, R.; Mobley, D., Sensitivity in binding free energies due to protein reorganization. *J. Chem. Theory Comput.* **2016**, *12* (9), 4620-4631.
26. Morton, A.; Matthews, B. W., Specificity of ligand binding in a buried nonpolar cavity of t4 lyozyme: Linkage of dynamics and structural plasticity. *Biochemistry* **1995**, *34*, 8576-8588.
27. Mobley, D. L.; Chodera, J. D.; Dill, K. A., On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **2006**, *125* (8), 084902.
28. Jiang, W.; Hodoseck, 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.
29. Murata, K.; Sugita, Y.; Okamoto, Y., Free energy calculations for DNA base stacking by replica-exchange umbrella sampling *Chemical Physics Letters* **2004**, *385*, 1-7.
30. Kannan, S.; Zacharis, M., Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 697-706.
31. Hamelberg, D.; Mongan, J.; McCammon, J. A., Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules *J. Chem. Phys.* **2004**, *120*, 11919-11929.
32. Miao, Y.; Feixas, F.; Eun, C.; McCammon, J. A., Accelerated molecular dynamics simulations of protein folding. *J. Comput. Chem.* **2015**, *36* (20), 1536-1549.
33. Wang, L.; Friesner, R.; Berne, B. J., Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **2011**, *115* (30), 9431-9438.
34. Jo, S.; Jiang, W., A generic implementation of replica exchange with solute tempering (REST2) algorithm in NAMD for complex biophysical simulations. *Comput Phys Commun* **2015**, *197*, 304-311.
35. Fajer, M.; Hamelberg, D.; McCammon, J. A., Replica-exchange accelerated molecular dynamics (REXAMD) applied to thermodynamic interconversion. *J. Chem. Theory Comput.* **2008**, *4*, 1565-1569.
36. Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A., Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *J. Chem. Theory Comput.* **2014**, *10* (7), 2677-2689.
37. Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781-1802.
38. Jo, S.; Jiang, W.; Lee, H. S.; Roux, B.; Im, W., Charmm-gui ligand binder for absolute binding free energy calculations and its application. *J. Chem. Inf. Model.* **2013**, *53* (1), 267-277.
39. Jo, S.; Cheng, X.; Lee, J.; Kim, S.; Park, S.; Patel, D.; Beaven, A.; Lee, K.; Rui, H.; Im, W., CHARMM-GUI 10 years for biomolecular modeling and simulation. *J Comput Chem* **2017**, *38* (15), 1114-1124.

40. SE Feller; Y Zhang; RW Pastor; Brooks, B., Constant pressure molecular dynamics simulation: The langevin piston method. *J. Chem. Phys.* **1995**, *103* (11).
41. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; A. D. MacKerell, J., Charmm general force field (cgenff): A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31* (4), 671-690.
42. Beutler, T. C.; Mark, A. E.; Schaik, R. C. v.; Gerber, P. R.; Gunsteren, W. F. v., Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett* **1994**, *222*, 529-539.
43. M., Z.; Straatsma, T. P.; A., M. J., Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994**, *100*, 9025-9031.
44. Lu, N.; Kofke, D.; Woolf, T., Improving the efficiency and reliability of free energy perturbation calculations using overlap sampling methods. *J Comput Chem* **2004**, *25* (28-39).
45. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M., THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011-1021.
46. Im, W.; Berneche, S.; Roux, B., Generalized solvent boundary potential for computer simulations. *J. Chem. Phys.* **2000**, *114*, 2924-2937.

# TOC Graphic

