JCTC Journal of Chemical Theory and Computation

Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations

Wei Jiang[†] and Benoît Roux*[,†,‡]

Biosciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Building 240, Argonne, Illinois 60439, and Department of Biochemistry and Molecular Biology, Gordon Center for Integrative Science, University of Chicago, 929 57th Street, Chicago, Illinois 60637

Received April 1, 2010

Abstract: Free Energy Perturbation with Replica Exchange Molecular Dynamics (FEP/REMD) offers a powerful strategy to improve the convergence of free energy computations. In particular, it has been shown previously that a FEP/REMD scheme allowing random moves within an extended replica ensemble of thermodynamic coupling parameters "λ" can improve the statistical convergence in calculations of absolute binding free energy of ligands to proteins [*J. Chem. Theory Comput.* **2009**, *5*, 2583]. In the present study, FEP/REMD is extended and combined with an accelerated MD simulations method based on Hamiltonian replica-exchange MD (H-REMD) to overcome the additional problems arising from the existence of kinetically trapped conformations within the protein receptor. In the combined strategy, each system with a given thermodynamic coupling factor λ in the extended ensemble is further coupled with a set of replicas evolving on a biased energy surface with boosting potentials used to accelerate the interconversion among different rotameric states of the side chains in the neighborhood of the binding site. Exchanges are allowed to occur alternatively along the axes corresponding to the thermodynamic coupling parameter λ and the boosting potential, in an extended dual array of coupled λ- and H-REMD simulations. The method is implemented on the basis of new extensions to the REPDSTR module of the biomolecular simulation program CHARMM. As an illustrative example, the absolute binding free energy of *p*-xylene to the nonpolar cavity of the L99A mutant of the T4 lysozyme was calculated. The tests demonstrate that the dual λ-REMD and H-REMD simulation scheme greatly accelerates the configurational sampling of the rotameric states of the side chains around the binding pocket, thereby improving the convergence of the FEP computations.

Introduction

Free energy perturbation molecular dynamics (FEP/MD) simulations with explicit solvent molecules provide one of the most fundamental routes for computing the binding affinities of small compounds to proteins.[1,2] In practice, a critical issue with FEP/MD simulations is to achieve a sufficient sampling of all the relevant degrees of freedom. Problems can arise with large structural reorganizations either in the ligand or in the protein upon formation of the bound complex because sampling those is typically beyond the reach of straight brute-force FEP/MD simulations. More specifically, when there are large energy barriers separating the relevant conformational states, the ligand or the protein may remain kinetically trapped in the starting configuration for a very long time during FEP/MD simulations, and alternate conformations are never visited. The incomplete configurational sampling results in computed binding free energies that are dependent on the starting protein or ligand configuration, which are of limited significance and practical use.

The structural changes observed upon the binding of aromatic molecules to a nonpolar cavity engineered in the L99A mutant of the T4 lysozyme (T4L) provide a good illustration of the type of problems that can arise from insufficient sampling (Figure 1). For the bound complexes involving small and medium-sized ligands (e.g., benzene, toluene, benzofurane, and indole), the protein structure is essentially identical to the ligand-free (*apo*) conformation. For those ligands, the calculated absolute binding free energies are well converged, regardless of whether the FEP/MD simulations are started from the *holo* or the *apo* state.[1,3,4] Difficulties arise in the case of larger ligands (e.g., indene, *n*-butylbenzene, isobutylbenzene, *o*-xylene, and *p*-xylene). In this case, the side chain of Val111, which is in direct contact with the bound ligand, changes its rotameric states from a *trans* conformation ($\chi_1 = 180°$) for the ligand-free apo to a *gauche* conformation ($\chi_1 = -60°$) for the bound state with large ligands. The intrinsic energy barrier around the χ_1 torsion of the valine (~5 kcal/mol) is sufficient to prevent the side chain from reorienting on the time scale of typical FEP/MD simula-

* Corresponding author e-mail: roux@uchicago.edu.
[†] Argonne National Laboratory.
[‡] University of Chicago.

10.1021/ct1001768
Published on Web 07/30/2010

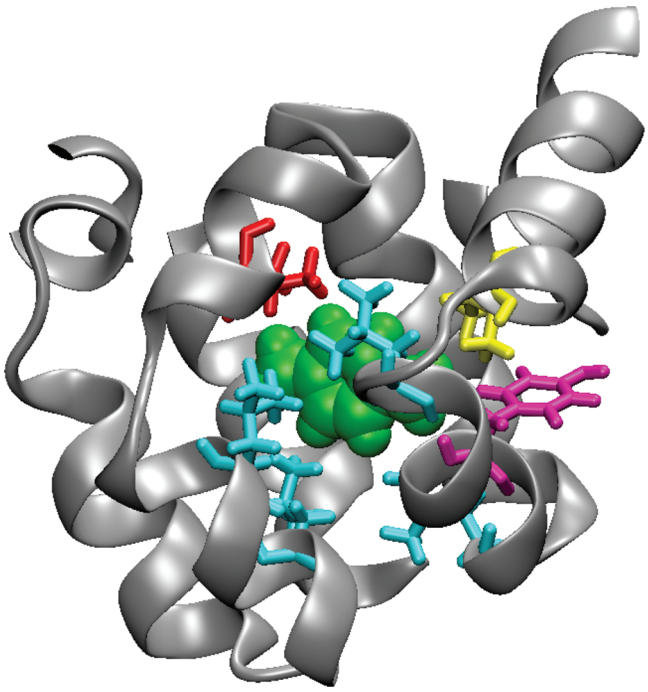

Figure 1. The artificially engineered nonpolar cavity of the L99A mutant of the T4 lysozyme (T4L/L99A) with *p*-xylene bound. Highlighted are seven protein side chains within 6 Å of the ligand (PDB 187L). Red color: valine 111. Blue color: leucine 84, 91, 118, and 121. Purple color: tyrosine 88. Yellow color: isoleucine 78.

tions. As a consequence, a FEP/MD calculation started from the *holo* state with the Val111 in the *gauche* state remains kinetically trapped while the ligand is alchemically decoupled, yielding a calculated binding free energy that is too favorable by 2−3 kcal/mol.[1] Alternatively, a FEP/MD calculation started from the *apo* state is too unfavorable by 2 kcal/mol.[4] As discussed in detail by Mobley et al.,[4] the lack of consistency between the two series of FEP calculations directly reflects the incomplete configurational averaging caused by the slow relaxation of the kinetically trapped degrees of freedom.

An elegant and powerful approach to enhance the sampling of the slowly varying degrees of freedom is to introduce a restraining potential serving as a "guide" to help reduce the size of the configurational space that needs to be explored during the free energy calculation. In practice, this first requires the identification of a key order parameter, ξ, associated with the slowly varying structural feature. Then, the potential of mean force (PMF) along this order parameter, $W(\xi)$, must be calculated via umbrella sampling biased simulations, and standard alchemical FEP/MD calculations are carried out in the presence of a biasing potential restricting the dynamics along ξ over a small range. Finally, unbiased thermodynamic averages for the entire association/dissociation process can be obtained by carrying out the explicit numerical integration of the probability distributions involving the Boltzmann factor of the PMF, $\exp[-W(\xi)/k_BT]$.[1,2,4−6] The free energy difference is evaluated as the reversible work for switching on a conformational restraint in one end-point state and switching it off in the other, according to a so-called "confine-and-release" cycle.[4] One might refer to this entire procedure as a "deliberate" PMF-based sampling strategy.

One important drawback from a deliberate PMF-based sampling strategy is that it relies on the prior identification of one or a few key degrees of freedom that one intends to control via umbrella sampling simulations. In the general case, it may not always be easy to determine which degrees of freedom might be slowly varying. A possible route to resolve the situation could be to extend the PMF-based strategy to multiple order parameters, but in practice, this does require carrying out umbrella sampling simulations for the entire multidimensional subspace. Thus, a deliberate PMF-based sampling strategy becomes rapidly unwieldy and inapplicable in the general situation where there could be structural rearrangements involving many elements. An alternative approach might be to simulate a conveniently chosen artificial reference state with soft cores as in the enhanced sampling−one-step perturbation method (ES-OS),[7] although it is unclear if this could treat multiple side chains surrounding a protein binding site simultaneously. A general treatment of structural relaxation remains a major challenge for applications in computer-assisted drug lead discovery,[8] where the main source of structural information may be ligand docking.

An alternative to a deliberate PMF-based sampling strategy is to exploit the concept of accelerated MD to increase the interconversion rates between metastable states.[9−12] The central element of accelerated MD consists in introducing a "boosting" potential that biases the energy surface to cancel out the intrinsic energy barriers opposing the relevant transitions that one wishes to sample. To retain the proper thermodynamic Boltzmann sampling of the system, the accelerated simulation can be combined with a parallel tempering Hamiltonian-REMD (H-REMD) scheme.[13,14] While this approach also requires the prior identification of the relevant subspace corresponding to the slowly varying degrees of freedom, the method is considerably less computationally expensive than the need to perform umbrella sampling simulations over multiple degrees of freedom as with a PMF-based strategy. An adequate sampling of the relevant subspace is expected to be, in most case, computationally affordable via a H-REMD scheme. In particular, as exemplified by the isomerization of Val111 in the L99A mutant of T4L discussed above, transitions of side chains and/or backbones in the neighborhood of the binding pocket clearly dominate the structural relaxation of the protein receptor in ligand binding free energy computations. More generally, the total number of side chains in the neighborhood of a binding pocket is fairly limited and it is likely that their dynamical transitions could be accelerated with FEP/H-REMD simulations.

In a previous communication, free energy perturbation (FEP) with a staged reversible thermodynamic work protocol designed for the calculation of absolute ligand binding affinities was combined with a distributed replica exchange MD (λ-REMD) simulation scheme.[15] It was shown that this FEP/λ-REMD scheme could improve the statistical convergence of FEP calculations by allowing random Monte Carlo moves in an extended ensemble of thermodynamic coupling parameter λ. The important concept of replica exchange in binding free energy calculations was first introduced by Woods and co-workers.[16] However, a straightforward FEP/λ-REMD algorithm is insufficient to accelerate the sampling of kinetically trapped degrees of freedom such as the isomerization of Val111 in the L99A mutant of T4L. The exchanges along the thermodynamic

coupling λ can help to mix the side chain rotamers of the protein in the *apo* and *holo* states, but transitions occur rarely due to the intrinsic dihedral energy barriers. In the present work, we extend those ideas to propose a rapid and robust framework for free energy computations combing the concept of λ-REMD simulations within the staged FEP, and the concept of accelerated MD with boosting potentials via H-REMD. To achieve the proper sampling enhancement in the relevant subspace, we combine λ-REMD with H-REMD. Random moves are allowed within an extended set of replicas biased by different values of the boosting factor "*b*" controlling the amplitude of a biasing potential according to a H-REMD scheme. The implementation is based on the MPI level parallel/parallel mode made possible by the Distributed Replica (REPDSTR) technique[17,18] of the program CHARMM,[19] in which each λ window of FEP is treated as an independent replica with its private I/O. With REPDSTR, it is straightforward to introduce a set of auxiliary boosting replicas for each λ window. This yields a dual REMD protocol for FEP calculations, with replica-exchange along two axes (2D) corresponding to the thermodynamic coupling parameter λ and a second axis corresponding to the boosting factor *b*. The entire array of REMD simulations can be executed as a single job via REPDSTR. It is shown that the dual FEP simulation scheme combining λ-REMD and H-REMD significantly accelerates the configurational sampling of the protein in FEP calculations. The method is illustrated with the calculation of the absolute binding free energy of *p*-xylene to the nonpolar cavity of T4L/L99A.

Computational Details

A. REPDSTR Implementation of Staging Simulation Protocol. In the FEP staging simulation protocol, the potential energy is expressed in terms of four coupling (window) parameters[1,2,20]

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

where U_0 is the potential of the system with the noninteracting (decoupled) ligand; λ_{rep}, λ_{dis}, λ_{elec}, and $\lambda_{rstr} \in [0,1]$ are the thermodynamic coupling parameters; U_{rep} and U_{dis} are the shifted Weeks−Chandler−Anderson[21] (WCA) repulsive and dispersive components of the Lennard-Jones potential; U_{elec} is the electrostatic contribution; and U_{rstr} is the restraining potential.

With the updated REPDSTR module of CHARMM,[19] the four-stage FEP simulation protocol can be implemented in a single parallel/parallel MPI job. Figure 2a shows the REPDSTR implementation of the updated FEP/REMD scheme, which is able to support the complete insertion process of the ligand into the binding pocket. The free energy corresponding to the process of inserting the ligand into the binding site is

$$U(\lambda_{rep}=0,\lambda_{dis}=0,\lambda_{elec}=0,\lambda_{rstr}=1) \rightarrow U(\lambda_{rep}=1,\lambda_{dis}=1,\lambda_{elec}=1,\lambda_{rstr}=0) \quad (1a)$$

To achieve a significant sampling enhancement, *M* additional replicas with "boosting" biasing potentials are introduced (b = 0, 1/*M*, 2/*M*, ..., 1) for each λ value of the FEP/REMD calculation. The boosting parameter *b* scales the biasing potential (the system is not biased when $b = 0$, and the biasing potential cancels out the intrinsic dihedral PMF of the side chain when $b = 1$). Figure 2b shows the FEP/REMD scheme with a biased Hamiltonian. Replica exchanges are attempted alternatively in λ space and *b* space, forming a two-dimensional (2D) REMD framework.

The replica-exchange algorithm follows the conventional Metropolis Monte Carlo exchange criterion:

$$P(\lambda_i \rightarrow \lambda_j) = \min\{1, e^{(-[U(\lambda_i,b_0,\mathbf{r}_{i,0})+U(\lambda_j,b_0,\mathbf{r}_{j,0})-U(\lambda_i,b_0,\mathbf{r}_{j,0})-U(\lambda_j,b_0,\mathbf{r}_{i,0})])/k_BT}\} \quad (2)$$

$$P(b_k \rightarrow b_l) = \min\{1, e^{-[U(\lambda_i,b_k,\mathbf{r}_{i,k})+U(\lambda_j,b_l,\mathbf{r}_{j,l})-U(\lambda_i,b_k,\mathbf{r}_{i,l})-U(\lambda_j,b_l,\mathbf{r}_{j,k})]/k_BT}\} \quad (3)$$

where *U* denotes the potential energy of the underlying replica, λ_i and λ_j denote the staging parameters, and b_k and b_l denote the boosting parameters.

B. Biasing Potentials for Residue χ_1 Dihedral Angle. Effective biasing boosting potentials can be obtained by fitting the potential of mean force (PMF) of individual dihedral degrees of freedom calculated on small peptides, in the spirit of the work of Kannan and Zacharis.[10] In the present application, biasing potentials for the side chain dihedral angles χ_1 were constructed by calculating a PMF in the gas phase. The biasing potentials were determined for valine, leucine, isoleucine, and tyrosine residues using umbrella sampling simulations of one-residue peptides. In the ligand binding calculation, these types of residues are distributed within a distance of 6 Å away from the center of mass of the ligand. A series of quadratic umbrella potentials with a force constant of 100 kcal mol^{-1} rad^{-2} and distributed every 5° was used. The angle χ_1 is defined as the C−CA−CB−CG dihedral, consistent with the CHARMM force field.[22] During the umbrella sampling simulations, the motions of the backbones were restrained by harmonic potentials around the conformation observed in the protein. The umbrella sampling simulations were unbiased with the weighted histogram analysis method (WHAM).[23] The resulting PMF along the dihedral angles was then fitted to a cosine Fourier series of the form

$$V(\phi) = \sum_{n=1}^{3} K_n(1 + \cos(n(\phi - \phi_{min}))) \quad (4)$$

The fitting parameters K_n are given in Table 1. Figure 3 shows the PMF and the result of the fit (the black curve is the PMF, and the red curve is the fitted cosine Fourier series). An appropriate boosting potential can easily be constructed by inverting the sign of the PMF $V(\phi)$ in eq 4 via the CONS DIHE command of CHARMM,[19] thereby canceling the potential barrier between the rotamers of a side chain for any selected residues.

C. MD Simulations. All the FEP/REMD simulations for the binding site were carried out on the IBM Blue Gene/P cluster Intrepid of the Argonne Leadership Computing Facility (ALCF) at Argonne National Laboratory (ANL). The simulations were carried out in a high performance mode using version c36a2 of the CHARMM program,[19] which was

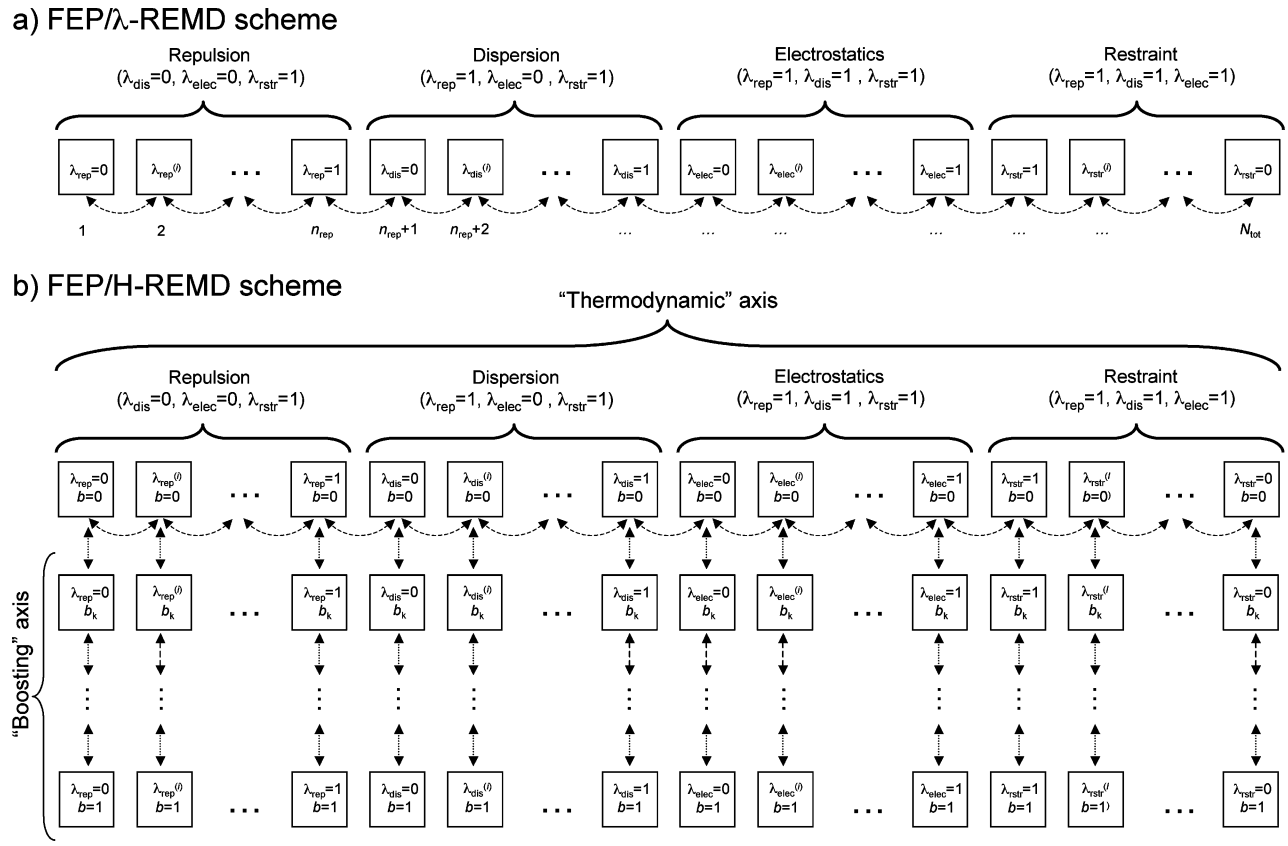

Figure 2. REPDSTR implementation of replica-exchange FEP simulation protocol in the context of the reversible work staging process for ligand binding free energy computations. Each square box represents an atomic simulation with its own I/O. Panel a illustrates the FEP/REMD scheme along the axis of the reversible thermodynamic work with coupling parameter λ ("λ-swap" moves).[15] The curly dashed-line arrows indicate the possible attempted exchanges, which are allowed only between neighboring replicas along the reversible staging process. At each cycle, the trial exchanges alternate between odd and even numbered replicas (ranked from 1 to N_{tot}), where even exchange means between windows 0 and 1, 2 and 3, 4 and 5, etc. and odd exchange means between windows 1 and 2, 3 and 4, 5 and 6, etc. Panel b illustrates the FEP/H-REMD scheme, where a vertical branch of M boosting-biasing replicas is attached to each of the windows along the reversible work process shown in panel a. The possible attempted moves, indicated by the dashed-line arrows, are again only allowed between neighboring replicas. In the FEP/H-REMD scheme, replica exchange is alternatively attempted along the axis of the reversible thermodynamic work with coupling parameter λ (curly horizontal arrows), and along the biasing axis with boosting parameter b (straight vertical arrows). Each exchange cycle consists of four stages: even and odd *local* exchanges between the biasing replicas within a host FEP window and even and odd *global* exchanges between those FEP windows with $b = 0$.

Table 1. Fitting Parameters of χ_1PMF with Cosinus-Fourier Series

residue	K_3 (kcal/mol)	ϕ_{min} (deg)	n	K_2 (kcal/mol)	ϕ_{min} (deg)	n	K_1 (kcal/mol)	ϕ_{min} (deg)	n
Val	2.9873	118.86°	3	−0.7662	78.96	2	0.7360	25.18	1
Ile	2.9407	119.07°	3	−0.7178	96.73	2	0.9651	31.86	1
Leu	2.4730	117.59	3	−0.7547	68.32	2	1.4487	−17.81	1
Tyr	2.3712	113.03	3	−0.7047	77.13	2	1.2107	−6.354	1

modified and extended for the present study. The hydration computations were performed on the IBM quads computing cluster KBT at ANL. The binding site free energy simulations were carried out on a reduced model of a solvated T4L/L99A system with the generalized solvent boundary potential (GSBP).[24] The initial T4L/L99A system was constructed from the crystallographic structure (PDB 187L) as described previously.[1] The hydration free energy computations of the isolated ligands were carried out under PBC conditions at constant pressure. The systems were propagated with a 2 fs time step using Langevin dynamics at a temperature of 298.15 K. For the binding free energy calculation of *p*-xylene, 100 ps production runs were performed for the binding site with a replica-exchange frequency of 1/100 steps. $M = 7$ boosting replicas were used with H-REMD. The force field parameters and initial structure of *p*-xylene were taken from our previous study for the sake of consistency.[1]

In all calculations, the energies were collected during the production run and postprocessed using WHAM.[23] For the binding site simulations, the WHAM postprocessing is only applied to replicas with zero biasing potential. To monitor the convergence of the binding site calculation, 20 independent FEP calculations (20 × 100 ps) were performed consecutively for each system, each starting from the configuration saved at the

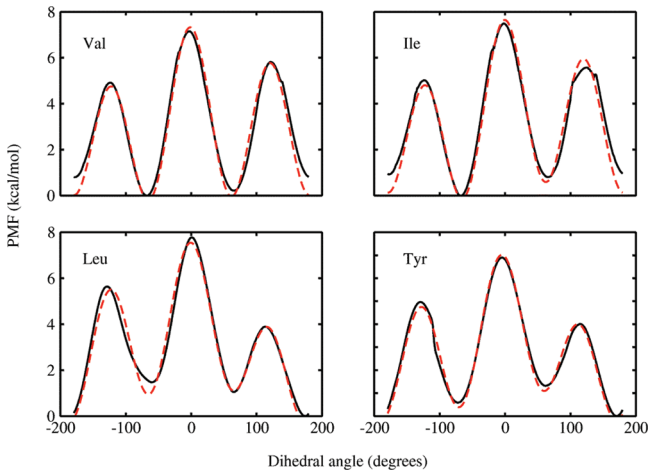

Figure 3. Fitting of χ_1 potential of mean force with linear superposition of three dihedral potential functions. Solid curve, PMF obtained with umbrellas sampling in gas phase; dashed curve, fitting curves.

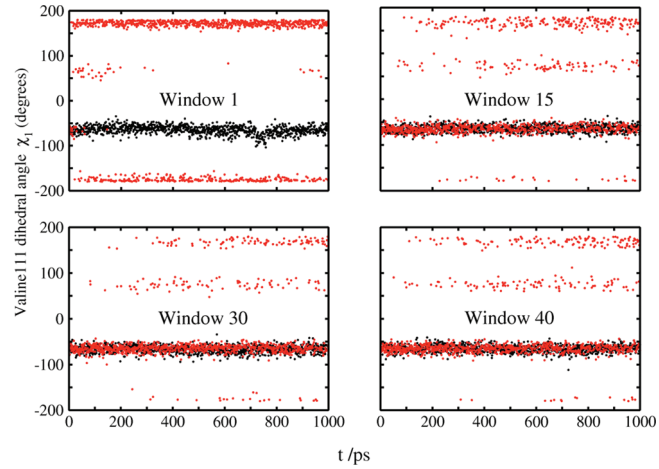

Figure 4. Enhanced sampling of rotameric states of Valine111. Black dots, values obtained with FEP/REMD; red dots, obtained with FEP/H-REMD. The investigated dihedral is χ_1, C−CA−CB−CG2 dihedral in the CHARMM force field. Window #1 turns off all ligand−receptor interactions and therefore corresponds to the *apo* state. Window #40 turns on all ligand−receptor interactions and corresponds to the *apo* state. With an increasing window index from 1 to 40, the thermodynamic state changes progressively from *apo* to *holo*.

end of the previous run. The last 10 runs were used to produce the final averaged result and calculate the standard deviations.

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.[1,25] In the case of *p*-xylene, the side chain of Val111 rotates by approximately 140°, to a χ_1 value of −35°, to avoid a steric clash with the ligand (Val111 is depicted in red in Figure 1). It is often convenient to utilize the *holo* configuration as the starting set of coordinates in absolute binding free energy calculations since this is provided either by the X-ray crystallographic structure of the bound complex or by the output of *in silico* docking. However, one must ensure that FEP/MD simulations do reversibly cover the relevant set of thermodynamic states. In principle, an ideal sampling should be able to reflect any conformational changes within the protein between the two end-point *holo* and *apo* states along the thermodynamic decoupling simulations. However, as shown in Figure 4, no spontaneous dihedral transition is observed in simple FEP/REMD simulations. The side chain remains kinetically "trapped" in its *holo* rotameric state, with χ_1 around −60°. This is consistent with the observation reported by Dill and co-workers; the dihedral of the Val111 is unable to cross spontaneously the energy barrier during the FEP simulations.[4] In previous calculation performed by Deng and Roux with FEP/MD simulations, a 300 ps production run started from the *holo* state resulted in a binding free energy of −9.06 kcal/mol, which is considerably overestimated when compared to the experimental value of −4.7 kcal/mol. A straight FEP/REMD scheme, by itself, improves the value to −6.4 kcal/mol (Table 2). However, the result remains too favorable compared to the experiment.

To address the issue of a kinetically trapped degree of freedom and to enhance the sampling of rotameric states, the extended FEP/H-REMD framework is introduced. As a preliminary test, the boosting potential in FEP/H-REMD was applied exclusively to the χ_1 degree of freedom of Val111, which is the most problematic residue. Eight biasing replicas were used to guarantee a high acceptance ratio (>80%) for exchange attempts between the replicas with adjacent values of the boosting parameter *b*. A binding free energy of −5.1 kcal/mol is achieved using this FEP/H-REMD scheme (Table 2), closer to the experimental value, and also in good agreement with the value of −5.06 kcal/mol obtained by Dill and co-workers using a PMF-based confine-and-release strategy.[4] The largest change occurs for the repulsive free energy contribution ΔG_{rep}, which increases by about 1.5 kcal/mol. The changes for the dispersive and electrostatic free energy contributions are smaller and do not suffer from convergence problems, most likely because they are switched-on after the repulsive core of the ligand has been inserted into the binding cavity. The enhancement of the conformational sampling for the calculation of the repulsive free energy contribution suggests that the dihedral energy barrier is associated with a steric contact with the ligand.

Further insight can be obtained by considering the time evolution of the χ_1 dihedral of Val111, as illustrated in Figure 4. Because all the windows of the FEP/H-REMD simulations are started with Val111 in the rotamer taken from the *holo* state X-ray structure (χ_1 near −60°), the time evolution of the χ_1 dihedral for the *apo* state (i.e., the first window with a completely decoupled ligand) is of particular interest. In the FEP/H-REMD simulations, it is observed that the χ_1 of Val111 rapidly starts to make transitions within ∼10 ps toward 180° (red curve), corresponding to the dominant rotamer for the *apo* state. In contrast, the side chain remains trapped, with χ_1 around −60° in the straight FEP/REMD simulations (black curve). Moreover, the time evolution of χ_1 for the *holo* state (window #40) fluctuates predominantly around 180°, with some excursions to other values. Along the thermodynamic coupling axis (λ), the population of rotamers changes progressively from the appropriate distribution of the *apo* and *holo* states. For the *apo* state, the average populations for *trans*, *gauche*$^+$, and *gauche*$^-$ are 0.99, 0.01, and 0.0, respectively. These results are in good

Table 2. Absolute Binding Free Energies of *p*-Xylene to T4L/L99A (all values in kcal/mol)

	binding site			bulk water	
	REMD	H-REMD (Val111)	H-REMD (7 residues)	PBC	exp
ΔG_{rep}	12.00 ± 0.21	13.55 ± 0.17	13.71 ± 0.10	16.02 ± 0.27	
ΔG_{disp}	−25.08 ± 0.07	−25.25 ± 0.08	−25.25 ± 0.08	−15.44 ± 0.02	
ΔG_{elec}	−0.74 ± 0.02	−0.78 ± 0.01	−0.78 ± 0.01	−1.61 ± 0.02	
ΔG_{rstr}	7.02 ± 0.09	7.02 ± 0.09	7.02 ± 0.09		
total	−7.45 ± 0.23	−6.12 ± 0.19	−5.96 ± 0.17	−1.03 ± 0.24	−0.87
ΔG_b°	**−6.42 ± 0.21**	**−5.09 ± 0.20**	**−4.93 ± 0.18**		**−4.67**

accord with the values estimated from the PMF of Dill and coworkers (0.99, 0.01, and 0.0008).[4] For the *holo* state, the average populations are 0.16, 0.11, and 0.73, again in good accord with the values estimated from the PMF reported by Dill and coworkers (0.23, 0.002, and 0.76). For both the *apo* and *holo* states, the three possible rotamers are ranked correctly, and the probability of the dominant state is reproduced within a few percent. The enhanced sampling provided by FEP/H-REMD is key to producing an accurate binding free energy started from the *holo* configuration.

The ultimate aim of the FEP/H-REMD framework is to enable binding free energy calculation without any prior knowledge of the location of a possible high potential barrier, such as the χ_1 of Val111 in T4L/L99A. In a next round of FEP/H-REMD calculations, we test this concept by applying indiscriminately a biasing boosting potential to seven protein side chains within a distance of 6 Å around the ligand. The residues affected by the boosting potentials are Leu84, Leu91, Leu118, Leu121, Val111, Ile78, and Tyr88. The affected residues around the binding pocket are displayed in color in Figure 1. It should be noted that the selection of residues is done on the basis of the starting (*holo*) configuration and remains unchanged during the entire FEP/H-REMD calculation. In this illustrative test, the stages corresponding to the dispersive and charging contributions were skipped in order to focus mainly on the dominant repulsive contribution. The results given in Table 2 show that a binding free energy of −4.9 kcal/mol is obtained, essentially identical with the calculation where only χ_1 of Val111 was boosted. Figure 5 shows the sampling of rotameric states of four selected residues about the binding pocket. For this FEP/H-REMD calculation, simple boosting potentials extracted from the PMF of small peptides in the gas phase were used. More sophisticated constructs could certainly be designed to further improve the quality of the boost potential. For example, one could also switch off the nonbonding between different side chains to resolve the problem of a kinetic bottleneck caused by steric clashes. Nevertheless, a simple PMF bias seems to be sufficient to enhance the sampling in the present application. For Val111, the distribution of the χ_1 value remains the same as for the former calculations. For Leu and Tyr, no interconversion among rotameric states is observed, which is consistent with the *apo* and *holo* X-ray structures. Interestingly, some infrequent transitions between two rotameric states of Ile78 are also observed. While the present application concerns the binding of *p*-xylene to a relatively small nonpolar cavity, the FEP/H-REMD scheme is expected to scale efficiently with an increased number of freedoms.

It should be emphasized that the enhanced sampling is achieved without any prior knowledge of all the relevant degrees

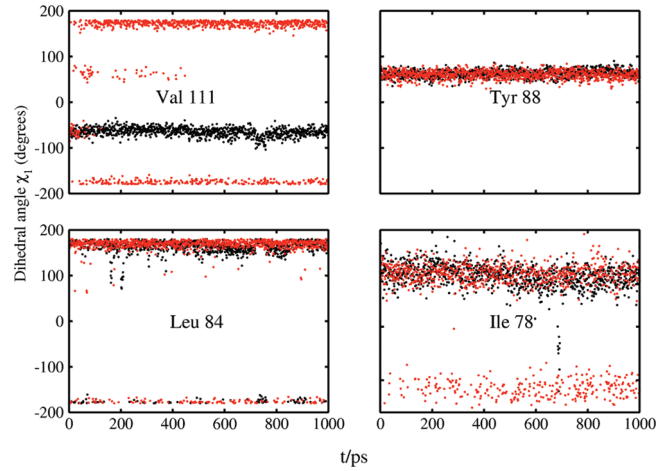

Figure 5. Rotameric states of four selected residues closest to the ligand. The four panels show the χ_1 of residues Val111, Tyr88, Leu84, and Ile78 in the *apo* state. It should be noted that Ile78 obtains a considerable sampling enhancement with FEP/H-REMD.

of freedom that are kinetically trapped. Nevertheless, it is necessary to apply the boosting potential to a finite subset of side chains and treat those via H-REMD. An important advantage is that the FEP/H-REMD scheme is not expected to be strongly size-limited. In principle, the proposed framework is applicable to a region of interest around the ligand, without a considerable loss in efficiency. As the number of side chains grows, longer simulations could be needed to get the same level of sampling. However, the moderate size of typical drug-like molecules ensures that only a finite and relatively small number of side chains should be treated with H-REMD. Effective boost potentials can be precalculated and stored in a library for any type of residues to achieve a universal sampling enhancement of side chain rotamers in the neighborhood of any binding pockets. Extensions to the present framework to include backbone degrees of freedom are in progress.

Conclusion

In summary, a dual FEP simulation scheme suitable for a large supercomputing platform was proposed to enhance the sampling of protein side chains in binding free energy calculations. Extending from our previous work,[15] each system with a given thermodynamic coupling factor λ in the extended ensemble in FEP/REMD is further coupled with a set of replicas evolving on a biased energy surface with boosting potentials accelerating the interconversion among different rotameric states of a set of side chains in the neighborhood of the binding site via a Hamiltonian REMD scheme. An important feature of the FEP/H-REMD scheme is that it can be used to enhance the sampling

of a fairly large number of putative slowly varying degrees of freedom without a considerable loss in efficiency. Sampling of any residue lining the binding pocket can benefit by the boosting H-REMD from a set of precalculated biasing potentials stored in a library. Application of FEP/H-REMD shows that the sampling of rotamers of the side chains surrounding the nonpolar cavity of T4L/L99A is significantly enhanced and that the binding free energy for a large ligand such as *p*-xylene can be calculated accurately by starting from the *holo* protein configuration. Further developments of the present method to include the treatment of backbone reorganization are currently in progress.

Acknowledgment. We are grateful to Dr. Andrew Binkowski for his support. We would like to acknowledge Dr. Milan Hodoscek for collaboration with the programming work for the CHARMM REPDSTR module, and Dr. Albert Lau for valuable discussions about free energy calculations and the replica-exchange scheme. This research is funded by grant MCB-0920261 from the National Science Foundation. Access to the computational resources of ALCF at ANL, supported by the Office of Science of the U.S. Department of Energy (DOE) under contract DE-AC02-06CH11357, was made possible by an INCITE grant from the DOE. The submitted manuscript has been created by UChicago Argonne, LLC, Operator of ANL. ANL, a U.S. DOE Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357.

References

(1) Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1255–1273.

(2) Wang, J.; Deng, Y.; Roux, B. *Biophys. J.* **2006**, *91*, 2798–2814.

(3) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; Mcreynolds, A. C.; Shoichet, B. K.; Dill, K. A. *J. Mol. Biol.* **2007**, *371*, 1118–1134.

(4) Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Theory Comput.* **2007**, *3*, 1231–1235.

(5) Woo, H. J.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825–6830.

(6) Gan, W.; Roux, B. *Proteins* **2009**, *74*, 996–1007.

(7) Hritz, J.; Oostenbrink, C. *J. Phys. Chem. B* **2009**, *113*, 12711–20.

(8) Jorgensen, W. *Acc. Chem. Res.* **2009**, *42*, 724.

(9) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919–11929.

(10) Kannan, S.; Zacharis, M. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 697–706.

(11) Chao, X.; Wang, J.; Liu, H. *J. Chem. Theory Comput.* **2008**, *4*, 1348–1359.

(12) Straatsma, T. P.; McCammon, J., A. *J. Chem. Phys.* **1994**, *101*, 5032–5039.

(13) Fajer, M.; Hamelberg, D.; McCammon, J. A. *J. Chem. Theory Comput.* **2008**, *4*, 1565–1569.

(14) Hritz, J.; Oostenbrink, C. *J. Chem. Phys.* **2008**, *128*, 144121.

(15) Jiang, W.; Hodoscek, M.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5*, 2583–2588.

(16) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B* **2003**, *107*, 13711–13718.

(17) Woodcock, H. L., III; Hodoscek, M.; Sherwood, P.; Lee, Y.; Schaefer, H.; Brooks, B. *Theor. Chem. Acc.* **2003**, *109*, 140–148.

(18) Woodcock, H. L., III; Hodoscek, M.; Gilbert, A. T. B.; Gill, P. M. W.; Schaefer, H. F., III; R., B. B. *J. Comput. Chem.* **2007**, *28*, 1485–1502.

(19) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(20) Deng, Y.; Roux, B. *J. Phys. Chem.* **2004**, *108*, 16567–16576.

(21) Weeks, J. D.; Chandler, D.; Anderson, H. C. *J. Chem. Phys.* **1971**, *54*, 5237–5247.

(22) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Wantanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(23) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(24) Im, W.; Berneche, S.; Roux, B. *J. Chem. Phys.* **2000**, *114*, 2924–2937.

(25) Morton, A.; Matthews, B. W. *Biochemistry* **1995**, *34*, 8576−8588.

CT1001768