Published in final edited form as:

J Chem Theory Comput. 2011 September 13; 7(9): 2721–2727. doi:10.1021/ct200153u.

Computing Alchemical Free Energy Differences with Hamiltonian Replica Exchange Molecular Dynamics (H-REMD) Simulations

Yilin Meng¹, Danial Sabri Dashti², and Adrian E. Roitberg^{3,*}

¹Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois, 60637

²Department of Physics and Quantum Theory Project, University of Florida, Gainesville, Florida, 32611-8435

³Department of Chemistry and Quantum Theory Project, University of Florida, Gainesville, Florida, 32611-8435

Abstract

Alchemical free energy calculations play a very important role in the field of molecular modeling. Efforts have been made to improve the accuracy and precision of those calculations. One of the efforts is to employ a Hamiltonian replica exchange molecular dynamics (H-REMD) method to enhance conformational sampling. In this paper, we demonstrated that HREMD method not only improves convergence in alchemical free energy calculations but also can be used to compute free energy differences directly via the Free Energy Perturbation (FEP) algorithm. We show a direct mapping between the H-REMD and the usual FEP equations, which are then used directly to compute free energies. The H-REMD alchemical free energy calculation (Replica exchange Free Energy Perturbation, REFEP) was tested on predicting the pK_a value of the buried Asp26 in thioredoxin. We compare the results of REFEP with TI and regular FEP simulations. REFEP calculations converged faster than those from TI and regular FEP simulations. The final predicted pK_a value from the H-REMD simulation was also very accurate, only 0.4 pK_a unit above the experimental value. Utilizing the REFEP algorithm significantly improves conformational sampling, and this in turn improves the convergence of alchemical free energy simulations.

Introduction

Free energy, especially the free energy difference between two states, is a crucial quantity in the study of chemical and biological systems.¹ Knowledge of the free energy differences can help us understand the behaviors of such systems. For example, the free energy of binding is one of the criteria used to evaluate the performance of drugs.² Therefore, one important aspect of molecular modeling is to yield accurate free energy difference efficiently. Many free energy calculation methodologies (such as free energy perturbation,³ thermodynamic integration,⁴ umbrella sampling,^{5–7} and Jarzynski's equality⁸) as well as analysis techniques (such as the weighted histogram analysis method⁹ and Bennett acceptance ratio method^{10,11}) have been developed to achieve this goal. In general, free energy calculations could be divided into alchemical free energy and conformational free energy calculations. The alchemical free energy calculations are often employed when studying the free energy differences of processes that involve changes in non-covalent interactions. In an alchemical

*to whom correspondence should be addressed. roitberg@ufl.edu.

free energy simulation, a non-physical reaction coordinate λ is generally adopted in order to connect the initial and final states. This reaction coordinate is usually expressed as an interpolation of the initial and final states. Thus, an alchemical process is achieved through a series of intermediate states having no direct physical meaning. Since the free energy difference between two states is a state function, the actual choice of coordinate cannot, in the limit of infinite sampling, affect the results. Free energy perturbation (FEP) and thermodynamic integration (TI) are two common methodologies that are utilized in alchemical free energy computations.

One important issue in alchemical free energy calculations is the convergence of the free energy difference versus computational cost. The convergence is particularly difficult in systems involving slow structural transition or large environmental reorganization as λ changes.^{12–14} Therefore, conformational sampling is crucial in alchemical free energy calculations. Enhanced sampling methods, such as replica exchange molecular dynamics (REMD),¹⁵ orthogonal space random walk (OSRW),¹⁴ and accelerated molecular dynamics (AMD)¹⁶ have been applied to free energy simulations in order to accelerate conformational sampling, and in turn to yield accurate and converged free energy differences. Among the enhanced sampling methodologies, REMD method is of particular interest because the weight of each state is a priori known (Boltzmann factor). The REMD algorithm was initially introduced by Sugita and Okamoto in 1999. In their REMD algorithm, N non-interacting copies (replicas) of a system are simulated at N different temperatures (one each). Regular MD is performed and periodically an exchange of configurations between two (usually adjacent) temperatures is attempted. Many variants of the original REMD method have been developed. One of them is the so-called “Hamiltonian REMD (H-REMD)”.^{17–20} In the H-REMD algorithms, replicas differ in their potential energies but (usually) have the same temperature. In practice, different ways of assigning potential energy function to replicas have been developed. For example, Fukunishi *et al.*¹⁷ scaled hydrophobic interactions and van der Waals interactions; protein-water as well as water-water interactions are scaled in the replica exchange with solute tempering (REST) algorithm;¹⁸ coarse-grained potential energy functions (low resolution) are combined with all-atom force fields (high resolution) in the resolution REMD algorithm.^{19–21}

Both temperature-based and Hamiltonian-based REMD have been applied to alchemical free energy calculations. Woods *et al.*^{13,22} and Rick²³ have combined the temperature-based REMD with TI calculation. A temperature-based REMD simulation is conducted at each state along the reaction coordinate. Woods *et al.*^{13,22} have also applied H-REMD methodology to FEP and TI calculations. Each replica in the H-REMD simulation represents a state along the reaction coordinate λ and a periodic swap in λ is attempted. Relative solvation free energy of water and methane as well as the relative binding free energies of halides to calis[4]pyrrole have been calculated in this way.²² The Yang group has developed a dual-topology alchemical H-REMD (DTA-HREM) method.²⁴ Their method was tested on the free energy of mutating an asparagine amino acid (with two ends blocked) to leucine. More recently, the Roux group has coupled FEP methodology with the Distributed Replica technique (REPDSTR).^{25,26} An additional acceleration in the sampling of side-chain dihedral angle was also incorporated when Jiang and Roux utilized FEP/H-REMD method to study the absolute binding free energy of *p*-xylene to the T4 lysozyme L99A mutant.²⁶ In all those studies, the conformational sampling and convergence of free energy computations have shown significant improvement when REMD method is applied. The protocol presented here accelerates convergence, but of course does not solve known problems in the field related to enhanced sampling of coordinates orthogonal to lambda space, which would hamper many of the current methods.

In this paper, we will demonstrate that FEP is actually already incorporated in the H-REMD method in an elegant and formal way. The REFEP method is shown to be not only an enhanced sampling method but also a free energy calculation algorithm. We will apply the REFEP method to the pK_a prediction of thioredoxin Asp26. The experimental pK_a value of 7.5 has been shown to be one of the largest shifted from the intrinsic pK_a value,^{27,28} and hence makes it an interesting case to be studied theoretically. TI and FEP (regular molecular dynamics for conformational sampling) alchemical free energy simulations have been conducted in order to compare with REFEP simulations. A very accurate theoretical pK_a value is obtained from REFEP simulations. The convergence of free energy difference and pK_a value is achieved in REFEP simulations much faster than that in the FEP and TI simulations. The advantage and simplicity of using HREMD simulation to compute alchemical free energy difference is clearly shown.

Theory and Method

Free Energy Perturbation (FEP)

The FEP method, which was initially introduced by Zwanzig in 1954,³ is a well established method and is considered as the most frequently employed methodology in alchemical free energy calculations.¹² The details of FEP, as well as TI, methodology and its applications have been extensively reviewed.^{12,29–32} Therefore, only a very brief description of the FEP and TI method will be given here. Consider two states (1 and 2) of a system in the canonical (NVT) ensemble, and their corresponding Helmholtz free energies A_1 and A_2 . The Helmholtz free energy difference between two states can be expressed as:

$$\Delta A_{1 \rightarrow 2} = -k_B T \ln \left\langle e^{-[U_2(q) - U_1(q)]/k_B T} \right\rangle_1 \quad (1)$$

Here, k_B is the Boltzmann constant, T is the temperature, and q is the molecular structure. U_1 and U_2 are the potential energy energies of states 1 and 2, respectively. The bracket with subscript 1 stands for the average calculated over the structural ensemble generated by state 1. In order to compute $\Delta A_{1 \rightarrow 2}$, one simulation of state 1 is performed. Once a configuration q is taken, the potential energy difference at configuration q is computed. The ensemble average of $\left\langle e^{-[U_2(q) - U_1(q)]/k_B T} \right\rangle_1$ can be calculated easily and hence, $\Delta A_{1 \rightarrow 2}$ is obtained. Although the Helmholtz free energies are utilized here, Eq. 1 can be extended to isothermal-isobaric (NPT) ensemble and to the Gibbs free energy in the same manner.

When the fluctuations in ΔU in equation 1 are too large, FEP calculations are notoriously hard to converge. The convergence of FEP calculation will be poor if the overlap in phase space between the two states is small. In order to compute the free energy difference between two states that are very different, intermediate states mixing the two end points are adopted in such a way that the differences between neighbors can be treated as perturbations. A frequently employed method to generate intermediate states is to interpolate potential energy functions linearly, as shown in Eq.2. In Eq.2, U_1 and U_2 are the potential energy function of state 1 and 2, respectively. Free energy differences between neighboring states are then computed. The sum of individual free energy differences will be the targeted free energy difference between states 1 and 2 (Eq.3). There are many ways of executing FEP calculations involving intermediate states. The double-ended, double-wide,^{30,33} and overlap sampling algorithms³⁴ are among the most popular ones. A thorough description of different algorithms and their performance can be found in a recent review by Jorgensen and Thomas.³⁰

$$U(\lambda) = (1 - \lambda) U_1 + \lambda U_2 \quad (2)$$

$$\Delta A_{1 \rightarrow 2} = -k_B T \sum_i \ln \left\langle e^{-(U(\lambda_{i+1}) - U(\lambda_i)) / k_B T} \right\rangle_i \quad (3)$$

In practice, computing $\Delta A_{1 \rightarrow 2}$ (forward free energy difference) is equally easy (or hard) as computing $\Delta A_{2 \rightarrow 1}$ (backward free energy difference) and one is exactly the opposite of the other in principle. Evaluation of forward and backward free energy differences provides an indication of convergence. Furthermore, the potential energy differences generated from both directions can be utilized to reduce statistical error. The Bennett Acceptance Ratio (BAR) method is a frequently employed scheme to improve the precision of a free energy estimator.^{10–12}

Thermodynamic Integration (TI)

Another way of writing the free energy difference between two states 1 and 2 is:

$$\Delta A_{1 \rightarrow 2} = \int_0^1 \left\langle \frac{\partial A}{\partial \lambda} \right\rangle d\lambda = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (4)$$

Here, λ is a reaction coordinate connecting state 1 and 2, U is the potential energy of a state along the reaction coordinate; the bracket represents an ensemble average generated at a value of λ . The integration is often evaluated numerically via trapezoidal rule or Gaussian quadrature. If $U(\lambda)$ is constructed as in Eq.2, the derivative of $U(\lambda)$ with respect to λ is:

$$\frac{\partial U(\lambda)}{\partial \lambda} = U_2 - U_1 \quad (5)$$

And the free energy difference between state 1 and 2 can be expressed as:

$$\Delta A_{1 \rightarrow 2} = \int_0^1 \langle U_2 - U_1 \rangle_\lambda d\lambda \quad (6)$$

Hence, the ensemble average of potential energy gap between state 1 and 2 at each λ value is needed in a TI calculation. In this manuscript we use the term TI to refer to constrained TI, in which the value of lambda is not allowed to change at each window.

Hamiltonian Replica Exchange Molecular Dynamics (H-REMD)

The original REMD method utilizes replicas having different temperatures (T-REMD). Replicas at high temperatures overcome potential energy barriers more easily than those at low temperatures. Another way to overcome potential energy barriers is simply changing the potential energy surface to reduce potential energy barriers. In the H-REMD algorithm, replicas differ in their Hamiltonians but have the same temperature. Regular MD is performed and an exchange of configurations between two neighboring replicas is attempted periodically.

Figure 1 demonstrates the H-REMD algorithm and the free energy computation in an H-REMD simulation. Let us consider two replicas 1 and 2 with corresponding potential energy U_1 and U_2 . By employing the detailed balance condition and Boltzmann weight of each molecular structure, the transition probability can be written as:

$$w(q_1 \rightarrow q_2) = \min \left\{ 1, e^{-[(U_1(q_2) + U_2(q_1) - U_1(q_1) - U_2(q_2))]/k_B T} \right\} \quad (7)$$

where q_1 and q_2 are the molecular structures of replica 1 and 2 before an exchange attempt, respectively. A Monte Carlo-Metropolis criterion³⁵ is used to evaluate if the attempted swap of structures between two replicas should be accepted or not.

Eq. 7 can be regrouped as:

$$w(q_1 \rightarrow q_2) = \min \left\{ 1, e^{-[(U_2(q_1) + U_1(q_1) - U_1(q_2) - U_2(q_2))]/k_B T} \right\} \quad (8)$$

Comparing the exponential terms in Eq. 1 and Eq. 8, it's clear that Eq. 8 incorporates all information necessary for a FEP calculation. $U_2(q_1) - U_1(q_1)$ is the potential energy difference computed on the basis of the structural ensemble generated by U_1 , while $U_1(q_2) - U_2(q_2)$ is the potential energy difference computed on the basis of the structural ensembles generated by U_2 . Every time the transition probability is computed, those potential energy differences can be utilized to compute the ensemble average shown in Eq. 1. Therefore, $\Delta A_{1 \rightarrow 2}$ and $\Delta A_{2 \rightarrow 1}$ can be computed on-the-fly utilizing the double-ended scheme. The ensemble average in Eq. 1 is computed regardless of whether an exchange attempt is accepted or rejected. When employing H-REMD method to improve conformational sampling in the study of alchemical changes, H-REMD simulations are not only able to enhance conformational sampling but also yield free energy difference directly. In fact, a regular FEP calculation can be thought of as an H-REMD calculation where no exchanges are allowed between replicas.

In practice, as shown in Figure 1, there are two free energy difference calculations (ΔA_{up} and ΔA_{down}) continuously associated with each replica. Take replica 1 as an example, $\Delta A_{\text{up}} = \Delta A_{1 \rightarrow 2}$ while $\Delta A_{\text{down}} = \Delta A_{1 \rightarrow 0}$. In principle, when converged, $\Delta A_{1,\text{up}}$ should be equal to the negative of $\Delta A_{2,\text{down}}$:

$$\Delta A_{1,\text{up}} = -k_B T \ln \left\langle e^{-(U_2 - U_1)/k_B T} \right\rangle_1 = -\Delta A_{2,\text{down}} = k_B T \ln \left\langle e^{-(U_1 - U_2)/k_B T} \right\rangle_2 \quad (9)$$

Any difference (except for the sign) between the two is an indication of error or lack of convergence.

Convergence was gauged also by the time dependence of the predicted free energy differences, computing ΔG versus simulation length. This provides an asymptotically unbiased estimator for ΔG and all methods presented here must eventually reach the same final value (within error bars). REFEP is presented in this article as showing faster convergence towards the final value.

Simulation Details

Accurately determining the pK_a values of ionizable residues, especially those with large shifts from intrinsic pK_a values, is of great interest both experimentally and

computationally.^{27,28,36} In this paper, the pK_a calculation of Asp26 in thioredoxin has been selected as a test case in order to compare the performance of alchemical free energy simulations. Asp26 has been found deeply buried in thioredoxin and possesses one of the largest pK_a shifts among protein carboxylic groups.^{27,28} Following the protocol employed in the paper of Simonson *et al.*,³⁶ the thermodynamic cycle utilized to compute the pK_a value of an ionizable residue is given in Figure 2. As can be seen in Figure 2, the use of a model compound as an auxiliary leg in the thermodynamic cycle makes ΔG_3 (proton to proton) be equal to zero. Not that Essentially, the pK_a shift relative to the intrinsic value ($pK_{a,model}$) is computed as:

$$pK_a(\text{protein}) = pK_a(\text{model}) + \frac{1}{2.303k_B T} [\Delta G(\text{proteinAH} \rightarrow \text{proteinA}^-) - \Delta G(\text{AH} \rightarrow \text{A}^-)] \quad (10)$$

where $\Delta G(\text{proteinAH} \rightarrow \text{proteinA}^-)$ and $\Delta G(\text{AH} \rightarrow \text{A}^-)$ are the free energy differences between protonated and deprotonated aspartic acid in the protein environment and in aqueous solution, respectively. Alchemical free energy simulations were performed in order to yield those two terms. In Eq.10, the Gibbs free energy differences are used because experiments determining pK_a values are generally conducted under an isobaric-isothermal condition.

Aspartic acid dipeptide in implicit water solvent was taken as the model compound with a pK_a value taken as 4.0.³⁷ The oxidized form of thioredoxin (PDB code 2TRX)³⁸ in implicit water was used in our simulation. Changes in ionization was represented by changes in the partial charges of the aspartic acid side chain (ASH→ASP in the AMBER terminology). Since the van der Waals radius of the proton in aspartic acid is zero for both protonated and deprotonated species the free energy difference only contains the electrostatic interactions.

Three types of free energy simulations have been performed for both model compound and the protein: TI (forward and backward), H-REMD-FEP (REFEP), and regular FEP simulations. Our regular FEP simulations were carried out via H-REMD simulations but with all exchange attempts rejected. Comparing the pK_a prediction and free energy convergence from FEP and REFEP simulations will directly indicate the effect of the enhanced conformational sampling due to the exchanges. Linear interpolation of point charges were carried out in order to assign side chain charges for intermediate states. A 7-point Gaussian quadrature has been selected to compute total free energy difference for TI calculations. Therefore, 8 λ values (one end point is needed in either direction) were utilized in the TI simulation. Due to the implementation of the TI algorithm in AMBER, 16 replicas were utilized to ensure the same amount of simulation time for all free energy simulations. A simulation time of 5 ns was used for each λ value and for each replica in the study of the model compound, while for thioredoxin we used 4 ns runs. Structural swaps between neighboring replicas were attempted every 2 ps (1000 MD steps). No particular attempt was made in this work to optimize the number or location of the replicas, nor the exchange attempt frequency. Work in this area is in progress.

All simulations were done using the AMBER 10 molecular simulation suite³⁹, locally modified to add H-REMD/REFEP capabilities. The AMBER ff99SB force field⁴⁰ was utilized in all the simulations. The SHAKE algorithm⁴¹ was used to constrain the bonds connecting hydrogen atoms with heavy atoms in all the simulations which allowed a use of a 2 fs time step. The OBC (Onufriev, Bashford and Case) Generalized Born implicit solvent model (*igb*=5 in the AMBER terminology)⁴² was used to model water environment in all

our calculations. The cutoff for non-bonded interaction and the Born radii was set to 99 Å. This value is larger than the dimension of both systems. Langevin dynamics was employed in order to maintain the temperature at 300 K, using a friction coefficient of 3.0 ps⁻¹.

Results and Discussions

Acceptance Ratio of H-REMD Simulations

The accuracy of FEP depends on the overlaps between phase spaces which can be measured as overlaps between potential energy difference distributions.¹² The acceptance ratio in an H-REMD simulation is an indication of the overlap between two potential energy difference distributions.²⁴ Therefore, it could be utilized to monitor the convergence of free energy calculation qualitatively. In our study, large acceptance ratios were observed in both model compound and protein H-REMD simulation. The acceptance ratio between two neighbors ranged from 0.7 to 0.9 in all H-REMD simulations. Those large acceptance ratios indicate that the overlap in phase space is large.

Aspartic Acid Model Compound Study

The free energy differences on the right-hand side of Eq. 10 were calculated as described in the “Theory and Method” section. The cumulative average free energy difference as a function of time is reported here. Figure 3A shows the $\Delta G(\text{AH} \rightarrow \text{A}^-)$ from TI, H-REMD, and FEP simulations (as mentioned before, a FEP simulation has been performed by rejecting all exchange attempts in an H-REMD simulation). The differences between forward and backward $\Delta G(\text{AH} \rightarrow \text{A}^-)$ are shown in Figure 3B. A converged alchemical free energy simulation should generate the same forward and backward free energy numerically (except for an opposite sign). Any non-zero value is an indication of free energy not converged.

For a simple system as aspartic acid in implicit water, 5 ns of simulation time was long enough for $\Delta G(\text{AH} \rightarrow \text{A}^-)$ to stabilize in all three alchemical free energy simulations as shown in Figure 3A. The forward and backward $\Delta G(\text{AH} \rightarrow \text{A}^-)$ at the end of each free energy calculation and the corresponding error bars are listed in Table 1. The forward and backward free energy differences are the same (within error bars) for both REFEP and FEP simulations. However, the TI simulations failed to do that, although the difference was very small (the difference between forward and backward $\Delta G(\text{AH} \rightarrow \text{A}^-)$ was only 0.13 kcal/mol). The average of forward and backward $\Delta G(\text{AH} \rightarrow \text{A}^-)$ was taken as the final value of $\Delta G(\text{AH} \rightarrow \text{A}^-)$ for the model compound and is also reported in Table 1. Clearly, as shown in Figure 3B, the REFEP simulations have converged much faster than the FEP calculations did.

Study on Asp26 in Thioredoxin

The free energy difference between protonated and deprotonated Asp26 is shown in Figure 3C and 3D. By analogy with the model compound plots, the cumulative average as a function of time is reported. The cumulative average was clearly not converged during the TI simulation and neither was the difference between forward and backward $\Delta G(\text{proteinAH} \rightarrow \text{proteinA}^-)$. According to Table 1, after 4 ns of TI simulation, the difference between forward and backward free energy was 1.4 kcal/mol, while the uncertainty of forward and backward free energy difference was 0.61 and 0.39 kcal/mol, respectively. Data not presented here shows that TI requires roughly 40 ns of dynamics before converging to results comparable with FEP/REFEP. Is it worth noting that this comparison is slightly unfair to TI and deserves further explanation. First, we used 8 intermediate states for TI versus 16 for FEP/REFEP. This setup, when executed within Amber, uses the same CPU time since the TI implementation is done with dual-topology methods. In fact, re-using the ensemble

generated with the FEP Hamiltonians and computing TI values on that ensemble produces very fast-converging results.

For regular FEP free energy calculations, the cumulative averages stabilized after roughly 2.2 ns of simulation, while the cumulative averages for the REFEP simulation stabilized much faster (shown in Figure 3C). Furthermore, Figure 3D illustrates that the difference between forward and backward $\Delta G(\text{proteinAH} \rightarrow \text{proteinA}^-)$ in the REFEP reached a value very close to zero (~ 0.05 kcal/mol) very quickly. As described previously, the final value of $\Delta G(\text{proteinAH} \rightarrow \text{proteinA}^-)$ was calculated as the average of forward and backward free energy differences. Although the final free energy differences computed from 4 ns of simulation were the same for REFEP and regular FEP, the calculations converged much faster in REFEP than in FEP simulation. Since the H-REMD and FEP calculations only differed in whether structures were allowed to be exchanged or not, the improvement in alchemical free energy convergence resulted from employing enhanced conformational sampling technique is significant. Data not presented here shows that the histograms of $P_1(\Delta U)\exp(-\beta(\Delta U))$ for the calculation of the free energy difference between replicas 1 and 2 for different sampling times are slightly different for FEP and REFEP. The REFEP distributions converge faster with time and sample the left side of the distribution better. This helps rationalize the faster convergence of our technique.

pK_a Prediction for Asp26 in Thioredoxin

The pK_a value of Asp26 in thioredoxin can be computed by Eq. 10. The final value of $\Delta G(\text{proteinAH} \rightarrow \text{proteinA}^-)$ from the REFEP simulation was -54.3 kcal/mol. with a predicted pK_a value of 7.9, which is only 0.4 pK_a units above the experimental value. The predicted pK_a value with respect to time from REFEP simulations was plotted in Figure 4 in order to demonstrate the convergence of the pK_a prediction. Figure 4 shows that REFEP simulations not only yielded accurate predicted pK_a value but also achieved convergence very fast. The regular FEP simulation predicted a pK_a value of 8.2, which is 0.7 pK_a units above the experimental value. The convergence in the regular FEP simulation was also worse than that in the REFEP simulation.

Conclusions

Conformational sampling is crucial in free energy calculations. In the case of alchemical free energy calculations, H-REMD is a useful and popular method to enhance the accuracy and convergence of free energy simulations. In this paper, we have demonstrated that REFEP not only improves conformational sampling in free energy calculations but also yields a free energy difference directly via the FEP algorithm. The implementation of REFEP is trivial, once a HREMD code is in place. The REFEP alchemical free energy calculation was tested on predicting the pK_a value of Asp26 in thioredoxin and compared with TI and regular FEP simulations. Free energy differences from the REFEP simulation converged faster than those from TI and regular FEP simulations. The final predicted pK_a value from the REFEP simulation was very accurate, only 0.4 pK_a unit above the experimental value. Utilizing the REFEP algorithm significantly improves conformational sampling, and this in turn improves the convergence of alchemical free energy simulations.

Acknowledgments

This work is supported from National Institute of Health under contract 1R01AI073674. Computer resources and support were provided by the Large Allocations Resource Committee through grant TG-MCA05S010 and the University of Florida High-Performance Computing Center. We thank the reviewers for extremely useful comments that have made this manuscript better.

References

- (1). Chipot, C.; Pohorille, A., editors. Free energy calculations. Theory and applications in chemistry and biology. Springer Verlag; Berlin, New York: 2007.
- (2). Bash PA, Singh UC, Brown FK, Langridge R, Kollman PA. Science. 1987; 235:574–576. [PubMed: 3810157]
- (3). Zwanzig RW. J. Chem. Phys. 1954; 22:1420–1426.
- (4). Kirkwood JG. J. Chem. Phys. 1935; 3:300–313.
- (5). Mezei M. J. Comput. Phys. 1987; 68:237–248.
- (6). Roux B. Comput. Phys. Commun. 1995; 91:275–282.
- (7). Torrie GM, Valleau JP. J. Comput. Phys. 1977; 23:187–199.
- (8). Jarzynski C. Phys. Rev. Lett. 1997; 78:2690–2693.
- (9). Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. J. Comput. Chem. 1992; 13:1011–1021.
- (10). Bennett CH. J. Comput. Phys. 1976; 22:245–268.
- (11). Shirts MR, Chodera JD. J. Chem. Phys. 2008; 129 124105.
- (12). Pohorille A, Jarzynski C, Chipot C. J. Phys. Chem. B. 2010; 114:10235–10253. [PubMed: 20701361]
- (13). Woods CJ, Essex JW, King MA. J. Phys. Chem. B. 2003; 107:13703–13710.
- (14). Zheng LQ, Chen MG, Yang W. Proc. Natl. Acad. Sci. U. S. A. 2008; 105:20227–20232. [PubMed: 19075242]
- (15). Sugita Y, Okamoto Y. Chem. Phys. Lett. 1999; 314:141–151.
- (16). Hamelberg D, Mongan J, McCammon JA. J. Chem. Phys. 2004; 120:11919–11929. [PubMed: 15268227]
- (17). Fukunishi H, Watanabe O, Takada S. J. Chem. Phys. 2002; 116:9058–9067.
- (18). Liu P, Kim B, Friesner RA, Berne BJ. Proc. Natl. Acad. Sci. U. S. A. 2005; 102:13749–13754. [PubMed: 16172406]
- (19). Liu P, Voth GA. J. Chem. Phys. 2007; 126 045106.
- (20). Lyman E, Ytreberg FM, Zuckerman DM. Phys. Rev. Lett. 2006; 96 028105.
- (21). Lwin TZ, Luo R. J. Chem. Phys. 2005; 123 194904.
- (22). Woods CJ, Essex JW, King MA. J. Phys. Chem. B. 2003; 107:13711–13718.
- (23). Rick SW. J. Chem. Theory Comput. 2006; 2:939–946.
- (24). Min DH, Li HZ, Li GH, Bitetti-Putzer R, Yang W. J. Chem. Phys. 2007; 126 144109.
- (25). Jiang W, Hodoscek M, Roux B. J. Chem. Theory Comput. 2009; 5:2583–2588. [PubMed: 21857812]
- (26). Jiang W, Roux B. J. Chem. Theory Comput. 2010; 6:2559–2565. [PubMed: 21857813]
- (27). Dyson HJ, Tennant LL, Holmgren A. Biochemistry. 1991; 30:4262–4268. [PubMed: 1850621]
- (28). Langsetmo K, Fuchs JA, Woodward C. Biochemistry. 1991; 30:7603–7609. [PubMed: 1854757]
- (29). Christ CD, Mark AE, van Gunsteren WF. J. Comput. Chem. 2010; 31:1569–1582. [PubMed: 20033914]
- (30). Jorgensen WL, Thomas LL. J. Chem. Theory Comput. 2008; 4:869–876. [PubMed: 19936324]
- (31). Kollman P. Chemical Reviews. 1993; 93:2395–2417.
- (32). Straatsma TP, Mccammon JA. Annu. Rev. Phys. Chem. 1992; 43:407–435.
- (33). Jorgensen WL, Ravimohan C. J. Chem. Phys. 1985; 83:3050–3054.
- (34). Lu ND, Kofke DA, Woolf TB. J. Comput. Chem. 2004; 25:28–39. [PubMed: 14634991]
- (35). Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. J. Chem. Phys. 1953; 21:1087–1092.
- (36). Simonson T, Carlsson J, Case DA. J. Am. Chem. Soc. 2004; 126:4167–4180. [PubMed: 15053606]
- (37). Bashford D, Case DA, Dalvit C, Tennant L, Wright PE. Biochemistry. 1993; 32:8045–8056. [PubMed: 8347606]

- (38). Katti SK, Lemaster DM, Eklund H. J. Mol. Biol. 1990; 212:167–184. [PubMed: 2181145]
- (39). Case, DA.; Darden, TA.; Cheatham, TE.; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Crowley, M.; Walker, RC.; Zhang, W.; Merz, KM.; B.; W.; Hayik, S.; Roitberg, AE.; Seabra, G.; Kolossváry, I.; Wong, KF.; Paesani, F.; Vanicek, J.; X.; W.; Brozell, SR.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, DH.; Seetin, MG.; Sagui, C.; Babin, V.; Kollman, PA. University of California: San Francisco; 2008.
- (40). Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Proteins: Struct., Funct., Bioinf. 2006; 65:712–725.
- (41). Ryckaert JP, Ciccotti G, Berendsen HJC. J. Comput. Phys. 1977; 23:327–341.
- (42). Onufriev A, Case DA, Bashford D. J. Comput. Chem. 2002; 23:1297–1304. [PubMed: 12214312]

Figure 1.

Diagrams displaying the H-REMD exchange algorithm and free energy calculation. (A): Exchange attempt orders. Replicas connected by a curve are neighbors and molecular configurations (q) are attempted to be exchanged. (B): Free energy calculations in the H-REMD method. Each replica has two free energy differences: ΔA_{up} and the ΔA_{down} from its attempting neighbor form a pair and are computed simultaneously, while ΔA_{down} and the ΔA_{up} from its attempting neighbor form the other pair. At exchange attempts (regardless if the attempts are accepted or rejected), two pairs of free energy differences are computed in an alternating fashion utilizing Eq. 1.

Figure 2.

Thermodynamic cycle used to compute the pK_a shift. Both acid dissociation reactions occur in aqueous solution. The “protein-AH” represents the ionizable residue in a protein environment. The “AH” represents the model compound which is usually the same ionizable residue with capped termini. In practice, a proton does not disappear but instead becomes a dummy atom. The proton still has its position and velocity. The bonded interactions involving the proton are still effective. However, there are no non-bonded interactions for that proton. The change in ionization state is reflected by changes of partial charges in the ionizable residue.

Figure 3.

(A) Cumulative average free energy differences between protonated and deprotonated aspartic acid in the model compound ($\Delta G(AH \rightarrow A^-)$). (B) The differences between forward and backward $\Delta G(AH \rightarrow A^-)$. (C) Cumulative average free energy differences between protonated and deprotonated Asp26 in thioredoxin ($\Delta G(\text{protein}AH \rightarrow \text{protein}A^-)$). (D) The differences between forward and backward $\Delta G(\text{protein}AH \rightarrow \text{protein}A^-)$.

Figure 4.

Predicted pK_a value of Asp26 in thioredoxin as a function of time. The $(\Delta G(AH \rightarrow A^-))$ values utilized in Eq.10 were -59.68 and -59.78 kcal/mol for REFEP and FEP, respectively. The experimental value is 7.5.

Table 1

Free energy difference between protonated and deprotonated aspartic acids obtained from TI, REFEP, and FEP alchemical free energy simulations. Free energy differences were calculated by utilizing all data points from a simulation (5 ns for the model compound and 4 ns for Asp26). The ΔG difference is given by $\Delta G(\text{proteinAH} \rightarrow \text{proteinA}^-) - \Delta G(\text{AH} \rightarrow \text{A}^-)$. All backward free energy differences have positive signs and hence multiplied by minus one. Then, the average values of forward and backward free energy differences were computed and reported here. All free energies have units of kcal/mol. The numbers in parenthesis are error bars. The error bars for forward and backward free energy differences of "model compound" and "Asp26 in thioredoxin" were calculated via block averages (a simulation was truncated into 5 blocks). The rest were obtained by error propagations.

		TI	REFEP	FEP
ASP Model Compound	Forward	-59.43 (0.06)	-59.69 (0.05)	-59.84 (0.06)
	Backward	-59.56 (0.06)	-59.66 (0.05)	-59.72 (0.06)
	Average	-59.50 (0.08)	-59.68 (0.08)	-59.78 (0.08)
Asp26 in thioredoxin	Forward	-54.35 (0.61)	-54.29 (0.17)	-54.23 (0.56)
	Backward	-55.82 (0.39)	-54.24 (0.14)	-53.84 (0.56)
	Average	-55.09 (0.72)	-54.27 (0.22)	-54.04 (0.79)
ΔG Difference	Forward	5.08 (0.61)	5.40 (0.18)	5.61 (0.56)
	Backward	3.74 (0.39)	5.42 (0.15)	5.88 (0.56)
	Average	4.41 (0.72)	5.41 (0.23)	5.74 (0.79)
Predicted $pK_{a, \text{protein}}$	Forward	7.7 (0.4)	7.9 (0.1)	8.1 (0.4)
	Backward	6.7 (0.3)	7.9 (0.1)	8.3 (0.4)
	Average	7.2 (0.5)	7.9 (0.2)	8.2 (0.6)