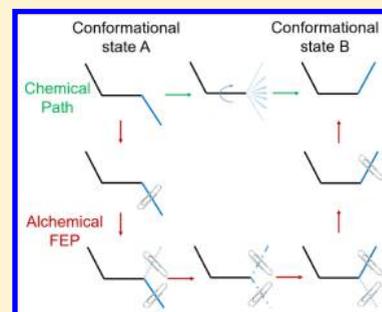


Conformational Free Energy Changes via an Alchemical Path without Reaction Coordinates

Peng He,[†] Bin W. Zhang,^{*,†} Shima Arasteh,[†] Lingle Wang,[‡] Robert Abel,[‡] and Ronald M. Levy^{*,†}[†]Center for Biophysics and Computational Biology, Department of Chemistry and Institute for Computational Molecular Science, Temple University, Philadelphia, Pennsylvania 19122, United States[‡]Schrödinger, Inc., 120 West 45th Street, New York, New York 10036, United States

Supporting Information

ABSTRACT: We introduce a novel method called restrain–free energy perturbation–release (R-FEP-R) to estimate conformational free energy changes via an alchemical path, which for some conformational landscapes like those associated with cellular signaling proteins in the kinase family is more direct and readily converged than the corresponding free energy changes along the physical path. The R-FEP-R method was developed from the dual topology free energy perturbation method that is widely applied to estimate the binding free energy difference between two ligands. In R-FEP-R, the free energy change between two conformational basins is calculated by free energy perturbations that remove those atoms involved in the conformational change from their initial conformational basin while simultaneously growing them back according to the final conformational basin. Both the initial and final dual topology states are unphysical, but they are designed in a way such that the unphysical contributions to the initial and final partition functions cancel. Compared with other advanced sampling algorithms such as umbrella sampling and metadynamics, the R-FEP-R method does not require predetermined transition pathways or reaction coordinates that connect the two conformational states. As a first illustration, the R-FEP-R method was applied to calculate the free energy change between conformational basins for alanine dipeptide in solution and for a side chain in the binding pocket of T4 lysozyme. The results obtained by R-FEP-R agree with the benchmarks very well.



Conformational free energy differences play an essential role in many areas of biophysics, such as enzyme activation, inhibitor specific binding, and allosteric effects.^{1–5} The straightforward way to estimate the conformational free energy differences is to run a long molecular dynamics (MD) simulation and collect the population of each conformational state. However, if the average first passage time or the relaxation time of the system is long compared with the times that are accessible given the available computational power, it is not possible to obtain converged results by using brute-force MD simulations.^{6,7} Because free energy is a state function, the conformational free energy difference can be obtained by following the free energy change along a transition pathway connecting the initial and final states with an umbrella sampling algorithm.^{8–11} However, sometimes finding a pathway is challenging itself although lots of transition pathway algorithms have been developed in the last two decades.^{12–24} Furthermore, the conformational changes along a transition pathway can be much larger than the conformational difference between the end points. For such transition pathways, it usually requires many umbrella sampling windows and long simulation time to obtain converged results. Wang–Landau sampling and metadynamics can also be applied to estimate the conformational free energy differences.^{25,26} These kind of algorithms choose a reduced reaction coordinate connecting the two target states of the complex system beforehand. Then

biases are added to the Hamiltonian function during the sampling until the potential of mean force along the reaction coordinate is constant. The success of Wang–Landau sampling and metadynamics algorithms strongly depends on the choice of the reduced reaction coordinate, which remains an open question in this field of research for complicated transitions.^{27,28} Many of the problems at the forefront of computational biophysics involve simulations of large allosteric conformational changes of proteins, like those associated with cellular signaling, where intermediate states along the physical transition pathways can differ much more in structure from the initial state than the final target state does. One such example is the transition of epidermal growth factor receptor kinase from the active conformational state to one of the many inactive conformational states; the root-mean-square deviation (RMSD) between the active and inactive states is only ~ 3.7 Å, whereas intermediate states along the physical path differ by as much as 9–12 Å.^{3,29} These processes pose a great challenge for the tools currently available to map free energy landscapes. To address problems of this kind, we introduce in this Letter a novel method called restrain–free energy perturbation–release (R-FEP-R) that estimates the conformational free energy

Received: June 13, 2018

Accepted: July 19, 2018

Published: July 19, 2018

difference between target states without choosing collective variables or knowing transition pathways beforehand. The R-FEP-R method was developed based on the idea of the dual topology free energy perturbation (FEP) transformation in relative binding free energy calculations.⁴

The dual topology free energy perturbation method is widely applied to estimate the binding free energy difference between two ligands that are similar in structure.^{4,30,31} The topology file used for the dual topology FEP simulations contains the topological information on the receptor and the two ligands of interest. The ligands' topology can be decomposed into three sets: the shared set and two dual topology sets. The shared set includes the topology of the same atoms that are shared by both ligands, and the two dual topology sets include the topology of the unique atoms of the two ligands, respectively. Suppose that the topology of the first ligand consists of the shared set and the first dual topology set and the topology of the second ligand consists of the shared set and the second dual topology set. At the beginning of the FEP simulations, the atoms described by the shared set and the first dual topology set have full interactions with the environment, namely the receptor and the solvent, while the atoms described by the second dual topology set have no interactions with the environment except a linkage with the shared part of the ligand. Then one decreases the interactions between the atoms described by the first dual topology set and the environment from full effect to zero by using a series of intermediate states. At the end of this step, the atoms described by the first dual topology set have no interactions with the environment except a linkage with the shared part of the ligand. Next (or simultaneously) one increases the interactions between the atoms described by the second dual topology set and the environment from zero to full effect by using another series of intermediate states. Suppose the free energy change during this

whole procedure is ΔG_{σ} , where the subscript denotes that the FEP simulations are performed when the ligands are a part of the binding complex. One needs to repeat the same procedure and obtain the free energy change of switching from the first ligand to the second in pure solvent ΔG_{σ} . Eventually, the binding free energy difference between these two ligands is estimated by

$$\Delta G_b = \Delta G_c - \Delta G_s \quad (1)$$

As can be seen, the atoms described by the first dual topology set change from real to virtual during the FEP simulations. Therefore, they will be termed the dual-RV set in this Letter. The atoms described by the second dual topology set will be termed the dual-VR set.

The basic idea of the dual topology free energy perturbation method can be used to estimate conformational free energy change. More specifically, the conformational free energy change can be calculated by FEP simulations that remove those atoms involved in the conformational change from their initial conformational basin and grow them back according to the final conformational basin.

The atoms of the molecule of interest can be divided into the shared set and the dual set by comparing two conformations representative of the initial and final conformational states. For simple cases, this can be done by inspection. For more complicated cases, one approach is to align the two conformations and calculate the root-mean-square deviation (RMSD) of atomic positions. The dual set can be chosen to contain all the atoms that have contributions to the RMSD significantly larger than the thermal fluctuations, and the shared set contains all the other atoms of the molecule. In the canonical ensemble, the conformational free energy change is

$$\begin{aligned} \Delta G &= -k_B T \ln \frac{Z_1}{Z_0} \\ &= -k_B T \ln \frac{\int \{dx^{(w)} dy^{(w)} dz^{(w)}\} \int \{dx^{(s)} dy^{(s)} dz^{(s)}\} \int_{S_1} \{dx^{(d)} dy^{(d)} dz^{(d)}\} \exp\{-\beta U\}}{\int \{dx^{(w)} dy^{(w)} dz^{(w)}\} \int \{dx^{(s)} dy^{(s)} dz^{(s)}\} \int_{S_0} \{dx^{(d)} dy^{(d)} dz^{(d)}\} \exp\{-\beta U\}} \end{aligned} \quad (2)$$

where $\beta = 1/(k_B T)$ is the inverse temperature and U is the total potential energy of the system. $\{x^{(w)}, y^{(w)}, z^{(w)}\}$, $\{x^{(s)}, y^{(s)}, z^{(s)}\}$, and $\{x^{(d)}, y^{(d)}, z^{(d)}\}$ are the Cartesian coordinates of the solvent (water), the shared set, and the dual set, respectively. The limits of integration S_0 and S_1 denote the definition of the initial and the final conformational states (free energy basins).

Before applying dual topology FEP to estimate the conformational free energy change, we introduce restraints to the dual set to prevent it from leaving the initial or the final conformational state and also to accelerate convergence of the FEP simulations along the λ coordinate. For example, one configuration is chosen from the initial conformational state S_0 . Then we apply a harmonic restraint to each dihedral angle of the dual set by using the dihedral angles of a chosen structure in the initial target state as the reference values. The partition function for the molecule in conformational state S_0 with restraints is

$$\begin{aligned} Z'_0 &= \int \{dx^{(w)} dy^{(w)} dz^{(w)}\} \int \{dx^{(s)} dy^{(s)} dz^{(s)}\} \\ &\int \{dx^{(d)} dy^{(d)} dz^{(d)}\} \exp\{-\beta(U + U_r(S_0))\} \end{aligned} \quad (3)$$

where $U_r(S_0)$ is the total restraint potential. Note the limits of integration for the dual set, S_0 , is omitted because we assume that the harmonic restraints are so strong that the molecule does not leave the conformational state S_0 in a finite simulation. The free energy change of adding these restraints, $\Delta G_0 = -k_B T \ln(Z'_0/Z_0)$ can be obtained by FEP simulations. Similarly, we apply a harmonic restraint to each dihedral angle of the dual set according to a chosen structure in the final state S_1 and estimate the free energy change $\Delta G_1 = -k_B T \ln(Z'_1/Z_1)$ where

$$\begin{aligned} Z'_1 &= \int \{dx^{(w)} dy^{(w)} dz^{(w)}\} \int \{dx^{(s)} dy^{(s)} dz^{(s)}\} \\ &\int \{dx^{(d)} dy^{(d)} dz^{(d)}\} \exp\{-\beta(U + U_r(S_1))\} \end{aligned} \quad (4)$$

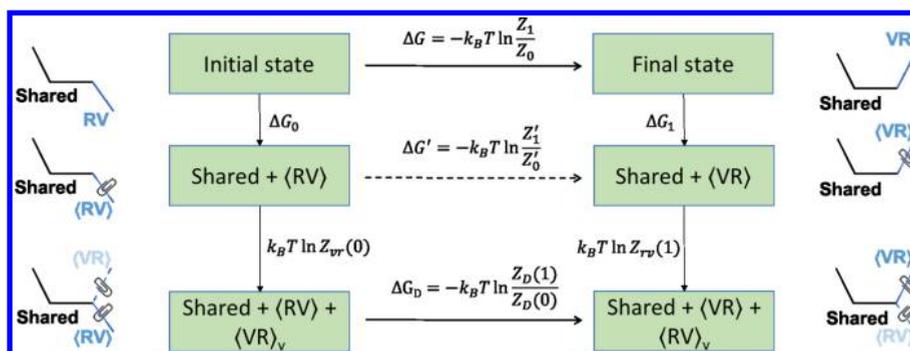


Figure 1. Thermodynamic cycles of the restrain-free energy perturbation-release method. The R-FEP-R method can be explained by two thermodynamic cycles. The top horizontal leg ΔG is the conformational free energy change, where the initial conformational state consists of the shared set and the dual-RV set, and the final conformational state consists of the shared and the dual-VR set. To estimate ΔG , harmonic restraints are applied to the dual-RV and the dual-VR set first. We use paperclips and angle brackets to present restraints in this figure. The free energy changes of applying harmonic restraints, ΔG_0 and ΔG_1 , are obtained by two FEP simulations. Based on the thermodynamic cycle on the top, the conformational free energy change ΔG can be estimated by the sum of ΔG_0 , $-\Delta G_1$, and $\Delta G'$, where $\Delta G'$ is the conformational free energy change with restraints. The horizontal leg in the middle, $\Delta G'$, is estimated by the bottom thermodynamic cycle. The two vertical legs in the bottom thermodynamic cycle represent that the addition of the restrained virtual fragments $\langle VR \rangle_v$ and $\langle RV \rangle_v$ to the restrained initial and final conformational states, respectively. We showed that the free energy changes of adding these two fragments, $k_B T \ln Z_{rv}(0)$ and $k_B T \ln Z_{rv}(1)$, are equal. Therefore, the middle horizontal leg equals the bottom horizontal leg, which is estimated by a dual topology free energy perturbation calculation. Finally, the conformational free energy change ΔG is estimated by the sum of ΔG_0 , $-\Delta G_1$, and ΔG_D .

As shown in Figure 1, the conformational free energy change can be estimated by summing three other legs in the top thermodynamic cycle

$$\begin{aligned} \Delta G &= -k_B T \ln \frac{Z_1}{Z_0} \\ &= -k_B T \ln \frac{Z_1}{Z'_1} - k_B T \ln \frac{Z'_1}{Z'_0} - k_B T \ln \frac{Z'_0}{Z_0} \\ &= \Delta G_0 - \Delta G_1 + \Delta G' \end{aligned} \quad (5)$$

where

$$\Delta G' = -k_B T \ln \frac{Z'_1}{Z'_0} \quad (6)$$

is the middle horizontal leg shown in Figure 1, which can be estimated by the dual topology free energy perturbation method as described below.

Similarly, the atoms of the molecule of interest in the dual topology FEP simulations can be divided into three sets. The shared set is the same as defined previously. The dual-RV set and the dual-VR set are copies of the dual set with harmonic restraints $U_r(S_0)$ and $U_r(S_1)$, respectively. We run parallel simulations at multiple λ -states, and the partition function of the system at the i th λ -state of the dual topology FEP is

$$\begin{aligned} Z_D(\lambda_i) &= \int \{dx dy dz\} \exp\{-\beta U(\lambda_i)\} \\ &= \int \{dx^{(w)} dy^{(w)} dz^{(w)}\} \int \{dr d\theta d\phi\} J(\{x, y, z\}, \{r, \theta, \phi\}) \\ &\quad \times \exp\{-\beta(U^{(s+w)} + U^{(rv)}(\lambda_i) + U^{(vr)}(\lambda_i))\} \end{aligned} \quad (7)$$

where $U(\lambda_i)$ is the total potential energy of the system at the i th λ -state. In the second line of eq 7, we change the variables of the integration of the solute from Cartesian coordinates to internal coordinates, where $J(\{x, y, z\}, \{r, \theta, \phi\})$ is the Jacobian determinant. Suppose there are n atoms in the shared set and m atoms in the dual-RV or the dual-VR set. The variables of the integration can be written as

$$\begin{aligned} \int \{dr d\theta d\phi\} &\sim \int dr_1^{(s)} d\theta_1^{(s)} d\phi_1^{(s)} \dots dr_n^{(s)} d\theta_n^{(s)} d\phi_n^{(s)} \\ &\quad \times \int dr_1^{(rv)} d\theta_1^{(rv)} d\phi_1^{(rv)} \dots dr_m^{(rv)} d\theta_m^{(rv)} d\phi_m^{(rv)} \\ &\quad \times \int dr_1^{(vr)} d\theta_1^{(vr)} d\phi_1^{(vr)} \dots dr_m^{(vr)} d\theta_m^{(vr)} d\phi_m^{(vr)} \\ &\sim \int \{dr^{(s)} d\theta^{(s)} d\phi^{(s)}\} \times \int \{dr^{(rv)} d\theta^{(rv)} d\phi^{(rv)}\} \\ &\quad \times \int \{dr^{(vr)} d\theta^{(vr)} d\phi^{(vr)}\} \end{aligned} \quad (8)$$

The Jacobian determinant of this change of variables is^{32,33}

$$\begin{aligned} J(\{x, y, z\}, \{r, \theta, \phi\}) &= (r_1^{(s)} r_2^{(s)} \dots r_n^{(s)})^2 \sin \theta_1^{(s)} \dots \sin \theta_n^{(s)} \\ &\quad \times (r_1^{(rv)} r_2^{(rv)} \dots r_m^{(rv)})^2 \sin \theta_1^{(rv)} \dots \sin \theta_m^{(rv)} \\ &\quad \times (r_1^{(vr)} r_2^{(vr)} \dots r_m^{(vr)})^2 \sin \theta_1^{(vr)} \dots \sin \theta_m^{(vr)} \\ &= J(\{r^{(s)}, \theta^{(s)}\}) \times J(\{r^{(rv)}, \theta^{(rv)}\}) \times J(\{r^{(vr)}, \theta^{(vr)}\}) \end{aligned} \quad (9)$$

The total potential energy $U(\lambda_i)$ is rewritten as a sum of three components in the second line of eq 7. $U^{(rv)}(\lambda_i)$ includes the intragroup interaction energies of the dual-RV set, its restraint potential energy, and the interaction energies between the dual-RV set and all the other atoms—the shared set and solvent. On the basis of the type of interactions that the potential energy functions describe, $U^{(rv)}(\lambda_i)$ can be written as

$$\begin{aligned} U^{(rv)}(\lambda_i) &= (U_{\text{bond}}^{(rv)} + U_{\text{angle}}^{(rv)} + U_{\text{improper}}^{(rv)} + U_r(S_0)) \\ &\quad + (1 - \lambda_i)(U_{\text{proper}}^{(rv)} + U_{\text{vdW}}^{(rv)} + U_{\text{elec}}^{(rv)}) \end{aligned} \quad (10)$$

where $U_{\text{bond}}^{(rv)} = U_{\text{bond}}^{(rv)}(\{r^{(rv)}\})$ is the bond length potential; $U_{\text{angle}}^{(rv)} = U_{\text{angle}}^{(rv)}(\{\theta^{(rv)}\})$ is the bond angle potential; $U_{\text{proper}}^{(rv)} = U_{\text{proper}}^{(rv)}(\{\phi^{(rv)}\})$ is the proper torsional potential; $U_{\text{improper}}^{(rv)} = U_{\text{improper}}^{(rv)}(\{r^{(rv)}, \theta^{(rv)}, \phi^{(rv)}\})$ is the improper torsional potential; $U_{\text{vdW}}^{(rv)} = U_{\text{vdW}}^{(rv)}(\{x^{(w)}, y^{(w)}, z^{(w)}\}, \{r^{(s)}, \theta^{(s)}, \phi^{(s)}\}, \{r^{(rv)}, \theta^{(rv)}, \phi^{(rv)}\})$ is the van der Waals potential; and $U_{\text{elec}}^{(rv)} = U_{\text{elec}}^{(rv)}(\{x^{(w)}, y^{(w)}, z^{(w)}\}, \{r^{(s)}, \theta^{(s)}, \phi^{(s)}\}, \{r^{(rv)}, \theta^{(rv)}, \phi^{(rv)}\})$ is the Coulomb potential. Formally, it is possible to integrate over the solvent

coordinates, in which case the effective potential energy of each solute conformation is the solvent averaged potential of mean force within which the solute moves. Note that the last two terms in eq 10 include both the intragroup interaction energy of the dual-RV set and the interaction energy between the dual-RV set and the other atoms in the shared set and the solvent. As can be seen, when λ changes from 0 to 1, the following nonquadratic energy terms change from the full effect to 0: (a) the proper dihedral potential energy, the intragroup van der Waals, and the Coulomb potential energies of the dual-RV set and (b) the van der Waals and the Coulomb potential energies between the dual-RV set and the other atoms (including the shared set and solvent). Similarly, $U^{(vr)}(\lambda_i)$ includes the intragroup interaction energies of the dual-VR set itself, its restraint potential, and the interaction energies between the dual-VR set and the other atoms. $U^{(vr)}(\lambda_i)$ can be written as

$$U^{(vr)}(\lambda_i) = (U_{\text{bond}}^{(vr)} + U_{\text{angle}}^{(vr)} + U_{\text{improper}}^{(vr)} + U_r(S_1)) + \lambda_i(U_{\text{proper}}^{(vr)} + U_{\text{vdW}}^{(vr)} + U_{\text{elec}}^{(vr)}) \quad (11)$$

where the nonquadratic energy terms listed previously increase from 0 to the full effect when λ changes from 0 to 1. Here, we use one parameter, λ , to control both $U^{(rv)}$ and $U^{(vr)}$ simultaneously. However, the interaction between the dual-RV set and the dual-VR set is always 0 in all the FEP simulations. The third component $U^{(s+w)} = U^{(s+w)}(\{x^{(w)}, y^{(w)}, z^{(w)}\}, \{r^{(s)}, \theta^{(s)}, \phi^{(s)}\})$ in eq 7 includes all the other potential energy terms of the system.

Next we examine the partition functions of the end point states. The partition function of the $\lambda = 0$ state is

$$Z_D(0) = \int \{dx^{(w)}dy^{(w)}dz^{(w)}\} \int \{drd\theta d\phi\} J(\{x, y, z\}, \{r, \theta, \phi\}) \times \exp\{-\beta(U^{(s+w)} + U^{(rv)}(0) + U^{(vr)}(0))\} = \iiint \exp\{-\beta(U^{(s+w)} + U^{(rv)}(0))\} J(\{r^{(s)}, \theta^{(s)}\}) \{dr^{(s)}d\theta^{(s)}d\phi^{(s)}\} \times J(\{r^{(rv)}, \theta^{(rv)}\}) \{dr^{(rv)}d\theta^{(rv)}d\phi^{(rv)}\} \{dx^{(w)}dy^{(w)}dz^{(w)}\} \times \int \exp\{-\beta U^{(vr)}(0)\} J(\{r^{(vr)}, \theta^{(vr)}\}) \{dr^{(vr)}d\theta^{(vr)}d\phi^{(vr)}\} = Z_{s+rv}(0) \times Z_{vr}(0) \quad (12)$$

where

$$Z_{s+rv}(0) = \iiint \exp\{-\beta(U^{(s+w)} + U^{(rv)}(0))\} J(\{r^{(s)}, \theta^{(s)}\}) \{dr^{(s)}d\theta^{(s)}d\phi^{(s)}\} \times J(\{r^{(rv)}, \theta^{(rv)}\}) \{dr^{(rv)}d\theta^{(rv)}d\phi^{(rv)}\} \{dx^{(w)}dy^{(w)}dz^{(w)}\} \quad (13)$$

and

$$Z_{vr}(0) = \int \exp\{-\beta U^{(vr)}(0)\} J(\{r^{(vr)}, \theta^{(vr)}\}) \{dr^{(vr)}d\theta^{(vr)}d\phi^{(vr)}\} = \int \exp\{-\beta(U_{\text{bond}}^{(vr)} + U_{\text{angle}}^{(vr)} + U_{\text{improper}}^{(vr)} + U_r(S_1))\} \times J(\{r^{(vr)}, \theta^{(vr)}\}) \{dr^{(vr)}d\theta^{(vr)}d\phi^{(vr)}\} \quad (14)$$

Notice that the partition function $Z_D(0)$ can be written as a product of $Z_{s+rv}(0)$ and $Z_{vr}(0)$ because these two partition functions are not correlated, which becomes obvious after we changed the variables of integration of the solute to internal coordinates. Furthermore, a careful examination reveals that $Z_{s+rv}(0)$ is the partition function for the molecule in conformational state S_0 with restraints, namely, equivalent to

Z'_0 defined by eq 3. Therefore, the partition function of the $\lambda = 0$ state can be written as

$$Z_D(0) = Z'_0 \times Z_{vr}(0) \quad (15)$$

Similarly, the partition function of the $\lambda = 1$ state can be written as

$$Z_D(1) = Z'_1 \times Z_{vr}(1) \quad (16)$$

where

$$Z'_1 = \iiint \exp\{-\beta(U^{(s+w)} + U^{(rv)}(1))\} J(\{r^{(s)}, \theta^{(s)}\}) \{dr^{(s)}d\theta^{(s)}d\phi^{(s)}\} \times J(\{r^{(rv)}, \theta^{(rv)}\}) \{dr^{(rv)}d\theta^{(rv)}d\phi^{(rv)}\} \{dx^{(w)}dy^{(w)}dz^{(w)}\} \quad (17)$$

and

$$Z_{vr}(1) = \int \exp\{-\beta U^{(rv)}(1)\} J(\{r^{(rv)}, \theta^{(rv)}\}) \{dr^{(rv)}d\theta^{(rv)}d\phi^{(rv)}\} = \int \exp\{-\beta(U_{\text{bond}}^{(rv)} + U_{\text{angle}}^{(rv)} + U_{\text{improper}}^{(rv)} + U_r(S_0))\} \times J(\{r^{(rv)}, \theta^{(rv)}\}) \{dr^{(rv)}d\theta^{(rv)}d\phi^{(rv)}\} \quad (18)$$

A comparison of eqs 14 and 18 reveals that the only differences between $Z_{rv}(1)$ and $Z_{vr}(0)$ are the terms of the harmonic restraint applied to the proper dihedral angles of the dual set, namely, $U_r(S_0)$ and $U_r(S_1)$. The two parameters that define a harmonic restraint are the reference value and the force constant. As the dual-RV and the dual-VR sets are both copies of the dual set of the molecule, we can choose the same force constant for the same dihedral angle in these two sets. The reference values of the same dihedral angle in $U_r(S_0)$ and $U_r(S_1)$ are usually different because one is determined by a configuration chosen from the conformational state S_0 and the other is determined by a configuration chosen from state S_1 . However, if the force constant is strong, the integration of a harmonic restraint term in $Z_{rv}(1)$ or $Z_{vr}(0)$ does not depend on the reference value because

$$\int_{-\pi}^{\pi} \exp\left\{-\frac{1}{2}\beta k_d(\phi - \phi_0)^2\right\} d\phi \approx \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\beta k_d(\phi - \phi_0)^2\right\} d\phi = \left(\frac{2\pi k_B T}{k_d}\right)^{1/2} \quad (19)$$

where k_d is the force constant and ϕ_0 is the reference value of the harmonic restraint. Therefore, the two partition functions of the virtual part at the end point states, $Z_{rv}(1)$ and $Z_{vr}(0)$, indeed are equal. Finally, the free energy difference between the end point states of the dual topology FEP is

$$\Delta G_D = -k_B T \ln \frac{Z_D(1)}{Z_D(0)} = -k_B T \ln \frac{Z'_1 \times Z_{vr}(1)}{Z'_0 \times Z_{vr}(0)} = -k_B T \ln \frac{Z'_1}{Z'_0} = \Delta G' \quad (20)$$

Combining eqs 5 and 20 shows that the conformational free energy change can be estimated by these three legs of the thermodynamic cycles shown in Figure 1

$$\Delta G = \Delta G_0 - \Delta G_1 + \Delta G_D \quad (21)$$

First we applied the R-FEP-R method to estimate the conformational free energy change of an alanine dipeptide molecule (AlaD) in water.³⁴ Figure 2 shows the Ramachandran

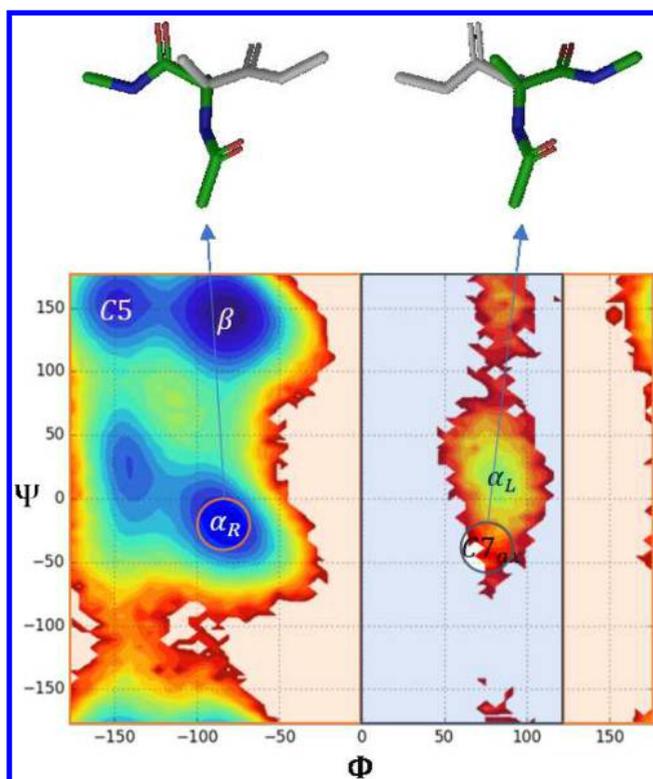


Figure 2. Ramachandran plot of AlaD. The region with the ivory background is the initial conformational state, and the region with the sky blue background is the final conformational state. Before running the dual topology FEP, we choose one configuration each from the α_R and the $C7_{ax}$ basins for the references and apply harmonic restraints to the dihedral angles. The AlaD molecule on the top left side represents a configuration sampled at the $\lambda = 0$ state. The dual-RV set (colored) is real, and the dual-VR set (monochrome) is virtual. The AlaD molecule on the top right side represents a configuration sampled at the $\lambda = 1$ state. The dual-RV set (monochrome) is virtual, and the dual-VR set (colored) is real.

plot of AlaD. We define the region ($-180 < \phi < 0$ and $120 < \psi < 180$) as the initial conformational states and the region ($0 < \phi < 120$) as the final conformational state. As can be seen, the initial conformational state contains the α_R , C5, and β free energy basins. The final conformational state contains the α_L and $C7_{ax}$ basins. Based on the temperature RE simulation results, the free energy difference between the initial and final conformational states is 2.91 ± 0.06 kcal/mol, which serves as the benchmark. The free energy difference of 2.91 ± 0.06 kcal/mol corresponds to a 131-fold enhancement of the population in the $\alpha_R + C5 + \beta$ free energy basins relative to the $\alpha_L + C7_{ax}$ free energy basin.

To apply R-FEP-R, the atoms comprising the N-terminal peptide plane and the C_α atom of AlaD are grouped together as the shared set and the atoms comprising the C-terminal peptide plane, the side chain (methyl group), and the α -hydrogen atom are grouped together as the dual set. Two copies of the dual set in different conformational states, dual-RV and dual-VR, are included in the system (Figure 2). We then chose one configuration each from the α_R and the $C7_{ax}$ basins for the references and apply harmonic restraints to the dihedral angles before running the dual topology FEP simulations. The free energy changes of restraint, FEP, and release are listed in Table 1 as simulation 1. The conformational free energy difference between the initial and final states

Table 1. Free Energy Changes (kcal/mol) and Standard Errors of Each Step in the R-FEP-R Calculations for AlaD

no.	ΔG_0	ΔG_1	ΔG_D	ΔG
1	3.12 ± 0.02	-4.23 ± 0.01	4.07 ± 0.06	2.96 ± 0.07
2	3.48 ± 0.02	-4.23 ± 0.01	3.60 ± 0.02	2.85 ± 0.03

estimated by the R-FEP-R method is 2.96 ± 0.07 kcal/mol, which agrees with the benchmark (2.91 ± 0.06 kcal/mol) within the statistical error, which is small. To clarify that the free energy difference estimated by the R-FEP-R method does not depend on the configurations that we choose for the reference to apply harmonic restraints within the (initial or final) conformational basin, we redid the R-FEP-R simulations, but the reference configuration of the restraint for the initial state was chosen from the C5 basin instead of the α_R basin. The results are listed in Table 1 as simulation 2. Compared with simulation 1, the difference in the restraint procedure (ΔG_0) is canceled by the difference in the dual topology FEP (ΔG_D) so that the final results of the second R-FEP-R simulation also match the benchmark. The total computation times of running temperature RE (the benchmark) and R-FEP-R (simulation 1) are 4900 and 930 ns, respectively. Assuming the magnitude of uncertainty is proportional to $1/\sqrt{n}$, where n is the total number of data, the R-FEP-R method is more efficient (~ 5 times) than the temperature RE simulation in the calculation of conformational free energy changes for this example.

Next we applied the R-FEP-R method to estimate the conformational free energy change of a side chain in the active site of T4 lysozyme L99A. T4 lysozyme is a well-known model system for the study of ligand-induced conformational transitions. As shown in Figure 3a, the side-chain dihedral angle χ of residue Val111 rotates from 180° (trans) to -60° (gauche+) upon ligand binding. Mobley et al. proposed a method called “confine-and-release” to obtain the correct binding free energy for the binding complex with ligand-induced conformational changes.³⁵ The first step of the “Confine-and-Release” method is to estimate the free energy difference between the trans and gauche+ states of T4 lysozyme without the ligand. The free energy profile constructed by umbrella sampling simulations is shown in Figure 3b, where the reaction coordinate is the side-chain dihedral angle χ of residue Val111. We define the region $-180^\circ < \chi < -150^\circ$ and $150^\circ < \chi < 180^\circ$ as the trans conformational state, and the region $-90^\circ < \chi < -30^\circ$ as the gauche+ conformational state. The free energy difference between these two states is 1.63 ± 0.07 kcal/mol, which serves as the benchmark. In the calculation of R-FEP-R, the side chain of the residue Val111 is defined as the dual set, and all the other atoms of T4 lysozyme are the shared set. We chose one configuration each from the trans and the gauche+ basins for the references and apply harmonic restraints to the dihedral angles of the dual-RV and dual-VR sets, respectively. The dual topology FEP was run to remove the dual-RV set and simultaneously grow the dual-VR set by using 15 λ -states. In this R-FEP-R calculation, the restraint, FEP, and release procedures were coupled by the Hamiltonian replica exchange algorithm to accelerate the convergence. The R-FEP-R estimate for the conformational free energy change of T4 lysozyme is 1.64 ± 0.03 kcal/mol, which agrees with the benchmark (1.63 ± 0.07 kcal/mol) very well. On the basis of the comparisons of the total simulation times and the

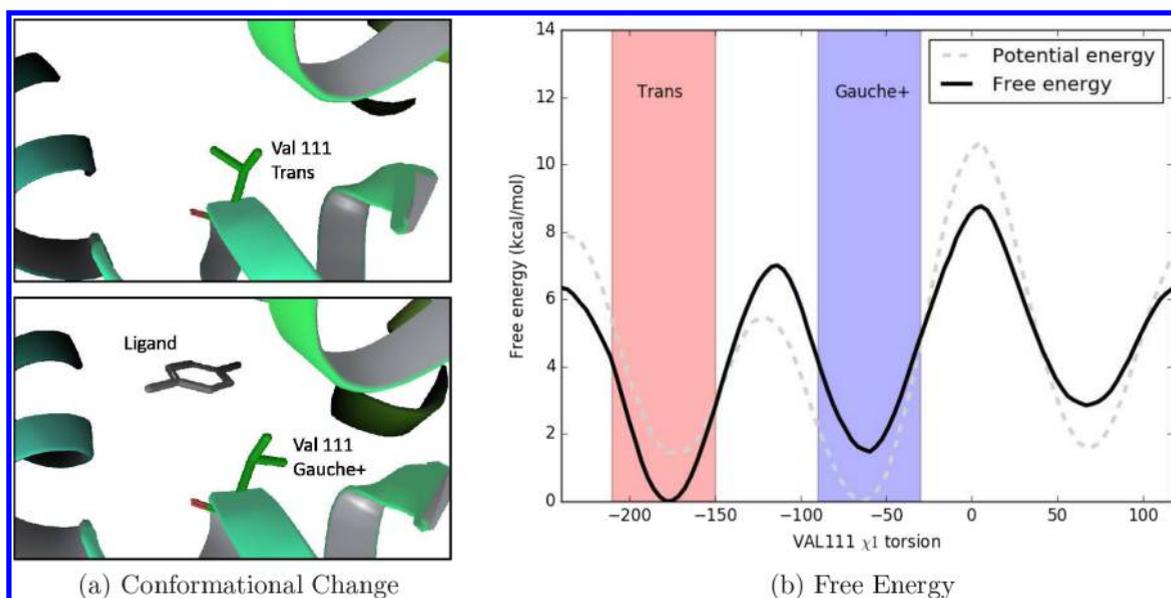


Figure 3. Conformational change of T4 lysozyme L99A. (a) Upon ligand binding, the side-chain dihedral angle χ of residue Val111 rotates from the trans state to the gauche+ state to avoid clash. (b) Dependence of the free energy and potential energy on the side-chain dihedral angle χ for the apo enzyme. The region with the red background is the trans state, and the region with the blue background is the gauche+ state.

uncertainties, we found that the R-FEP-R method is ~ 4 times more efficient than the umbrella sampling method in the calculation of conformational free energy changes for this example.

In this Letter, we introduce a method called R-FEP-R (restrain-free energy perturbation-release) to estimate conformational free energy differences via an alchemical path. The R-FEP-R method was developed based on the dual topology free energy perturbation method that is widely used to estimate the relative binding free energies of two ligands. To be more precise, the conformational free energy change is calculated by free energy perturbations that remove those atoms involved in the conformational change from their initial conformational state and simultaneously grow them back according to the final conformational state. After a brief review of the original dual free energy perturbation method, we showed theoretically that the R-FEP-R method can be applied to calculate conformational free energy differences.

The major difference between the dual topology free energy perturbation method for estimating the relative binding affinity of two ligands (DT-FEP-RB) and the R-FEP-R method for estimating the conformational free energy difference between two conformational basins is how the unphysical contributions from the dual topology dummy atoms to the initial and final partition functions is canceled, so that the free energy difference between the initial and final states contains only the physical contributions. The DT-FEP-RB calculation of binding free energy differences between two different ligands consists of two independent FEP simulations, one in the pure solvent box (usually water) and one in the receptor in water; in each simulation, the dual topology atoms represent the atoms which are unique to the initial or final ligand. In each of the two independent simulations (one in water, the other in the receptor), the dual atoms are transformed from dummy atoms to real atoms and vice versa. When the free energy difference between the two independent simulations is taken, the contribution of the dual dummy atom ensembles to the free

energy difference between the initial and final states cancels if the simulations are converged.

In the R-FEP-R method for estimating conformational free energy differences, however, there is only one simulation performed, which corresponds to an alchemical transformation from the initial conformational free energy basin to the final free energy basin. The two sets of dual topology atoms contain the same atom types with the same harmonic bond stretching and bending force constants; the torsional potential function of the dual topology atoms is also the same. In the R-FEP-R transformation, the nonharmonic torsional ensembles of the initial and final conformational states are both converted to Gaussian torsional ensemble distributions (with different centers). The free energy cost of these two torsional nonharmonic transformations (restraining and releasing the virtual dual atoms) are included in the estimate of the conformational free energy difference; the contribution of the virtual dual atom Gaussian ensembles in the initial and final states cancel so that the R-FEP-R transformation also includes only physical contributions. The two thermodynamic cycles also differ in that the potential function parameters of the real dual atoms are different at the physical end points for the DT-FEP-RB cycle, whereas they are the same for the R-FEP-R cycle.

Then the R-FEP-R method was tested for conformational changes for two model systems: alanine dipeptide and a side chain in the active site of T4 lysozyme. For both cases, the conformational free energy differences estimated by the R-FEP-R method agree with the benchmark very well. For these two examples, the R-FEP-R method for estimating conformational free energy differences is ~ 4 – 5 times more efficient as compared with replica exchange or umbrella sampling. Compared with other popular biased sampling algorithms such as umbrella sampling and metadynamics, the advantages of the R-FEP-R method are that the transition pathways or predetermined reaction coordinates that connect the two conformational states are not required and for some problems R-FEP-R may introduce a smaller perturbation to the system.

Although temperature RE simulations can also be used to obtain the conformational free energy differences without transition pathways or reaction coordinates, the AlaD example suggests that the R-FEP-R method may be more efficient than temperature RE in the calculation of conformational free energy changes at least in some cases. It will be challenging to use R-FEP-R to calculate conformational free energy changes when the dual atom set specified by the dual topology is large and the dual-RV and the dual-VR atom sets do not overlap in space, because it is difficult to slowly grow or annihilate a large collection of atoms in solution. However, when the dual set contains a large number of atoms, if the conformational difference between the initial and final ensembles containing the dual atoms is smaller than the conformational changes that are associated with the physical transition paths, we expect that calculating conformational free energy differences along alchemical paths designed to take advantage of the dual topology framework will be more efficient than using a physical path to determine the corresponding free energy changes. The use of alchemical paths to calculate conformational free energy changes, although different from our approach based on the dual topology framework described here, has been proposed previously.³⁶ The ability to calculate free energy differences for large allosteric conformational changes of proteins like those associated with cellular signaling of various kinds remains an unsolved problem. There is considerable room left to explore different ways to construct alchemical paths for larger conformational free energy changes with the objective of accelerating their convergence.

■ SIMULATION DETAILS

All simulations were performed using the GROMACS (v2016.3) simulation package.³⁷ Before the production MD simulations, each system was prepared at 300 K by running (1) 5000 steps of energy minimization, (2) 50 000 steps of equilibration under a *NVT* ensemble, and (3) 500 000 steps of equilibration under a *NPT* ensemble. The step size of MD simulations was 2 fs with the SHAKE constraint algorithm. Long-range electrostatic interactions were computed by the PME method. To calculate the conformational free energy change of AlaD, a single AlaD molecule was solvated in 880 TIP3P water molecules in an octahedron box. The clearance on each side of the solute was 1.2 nm. OPLS-AA force field³⁸ was used, and the production MD simulations were run under a *NVT* ensemble using the leapfrog stochastic dynamics integrator. The time constant for the temperature coupling (τ) is 1.0 ps. The benchmark was obtained by running temperature replica exchange simulations with seven temperatures ranging from 300 to 350 K (300, 307, 315, 322, 330, 339, and 350 K). Before the dual topology FEP was run, the force constants of the harmonic restraints applied to the ϕ and ψ dihedral angles of AlaD were increased from 0 to 239 kcal/mol by using eight λ -states. The chosen λ values were 0.0, 0.004, 0.05, 0.2, 0.4, 0.6, 0.8, and 1.0. There were 15 λ -states in the dual topology FEP simulations. The chosen λ values were 0.0, 0.02, 0.05, 0.1, 0.15, 0.24, 0.35, 0.5, 0.65, 0.76, 0.85, 0.9, 0.95, 0.98, and 1.0. To calculate the conformational change of T4 lysozyme, the T4 lysozyme molecule was solvated in 13 329 TIP3P water molecules in an octahedron box. The clearance on each side of the solute was 1.2 nm. AMBER99SB-ildn force field³⁹ were used, and the production MD simulations were run under a *NPT* ensemble using the leapfrog stochastic dynamics integrator and the Parrinello–Rahman pressure

coupling. The time constant for the temperature coupling (τ) is 1.0 ps, and the time constant for the pressure coupling (τ) is 2.0 ps. The benchmark was obtained by running umbrella sampling with 26 windows covering the Val111 χ space. The US simulations were coupled by replica exchange to accelerate the convergence. In the R-FEP-R calculations, the force constants applied to the χ dihedral angles of Val111 increase from 0 to 239 kcal/mol by using eight λ -states. The chosen λ values were 0.0, 0.004, 0.05, 0.2, 0.4, 0.6, 0.8, and 1.0. There were 15 λ -states in the dual topology FEP. The chosen λ values were 0.0, 0.02, 0.05, 0.1, 0.15, 0.24, 0.35, 0.5, 0.65, 0.76, 0.85, 0.9, 0.95, 0.98, and 1.0. All three steps—restraint, FEP, and release—were coupled by replica exchange to accelerate the convergence. The FEP simulation data were analyzed by UWAM.⁴⁰

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.8b01851.

Uncertainty estimate and convergence plot for two examples (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: bin.w.zhang@temple.edu.

*E-mail: ronlevy@temple.edu.

ORCID

Bin W. Zhang: 0000-0003-3007-4900

Lingle Wang: 0000-0002-8170-6798

Ronald M. Levy: 0000-0001-8696-5177

Author Contributions

L.W. and R.A. co-initiated the idea of using a dual topology framework for calculating conformational free energies, and L.W. suggested an alchemical path for which the virtual dual atom contributions to the initial and final partition functions cancel. P.H., B.W.Z., and R.M.L. proposed the model systems and the benchmark calculations. P.H. performed the calculations with assistance from S.A. P.H., B.W.Z., and R.M.L. wrote the manuscript.

Notes

The authors declare the following competing financial interest(s): R.M.L. is consultant to Schrodinger, Inc. and is on its Scientific Advisory Board.

■ ACKNOWLEDGMENTS

This work has been supported by National Institutes of Health grants RO1-GM030580 and computer grant S10-OD020095. The authors acknowledge invaluable discussions with Nanjie Deng at Pace University and Di Cui, Junchao Xia, and Yoshitake Sakae at Temple University.

■ REFERENCES

- (1) Gallicchio, E.; Levy, R. M. Recent theoretical and computational advances for modeling protein–ligand binding affinities. *Adv. Protein Chem. Struct. Biol.* **2011**, *85*, 27–80.
- (2) Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R. Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative Binding Affinity Predictions of CDK2 Inhibitors. *J. Chem. Theory Comput.* **2013**, *9*, 1282–1293.

- (3) Shan, Y.; Arkipov, A.; Kim, E. T.; Pan, A. C.; Shaw, D. E. Transitions to catalytically inactive conformations in EGFR kinase. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 7270–7275.
- (4) Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, *10*, 2632–2647.
- (5) Sultan, M. M.; Denny, R. A.; Unwalla, R.; Lovering, F.; Pande, V. S. Millisecond dynamics of BTK reveal kinome-wide conformational plasticity within the apo kinase domain. *Sci. Rep.* **2017**, *7*, 15604.
- (6) Zwier, M. C.; Chong, L. T. Reaching Biological Timescales with All-Atom Molecular Dynamics Simulations. *Curr. Opin. Pharmacol.* **2010**, *10*, 745–752.
- (7) Makarov, D. E. *Single Molecule Science: Physical Principles and Models*; CRC Press: Boca Raton, FL, 2015; Chapter 5, p 59.
- (8) Ravindranathan, K. P.; Gallicchio, E.; Levy, R. M. Conformational Equilibria and Free Energy Profiles for the Allosteric Transition of the Ribose-binding Protein. *J. Mol. Biol.* **2005**, *353*, 196–210.
- (9) Dickson, A.; Warmflash, A.; Dinner, A. R. Nonequilibrium umbrella sampling in spaces of many order parameters. *J. Chem. Phys.* **2009**, *130*, 074104.
- (10) Law, S.; Feig, M. Base-Flipping Mechanism in Postmismatch Recognition by MutS. *Biophys. J.* **2011**, *101*, 2223–2231.
- (11) Lin, Y.-L.; Meng, Y.; Jiang, W.; Roux, B. Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 1664–1669.
- (12) Pratt, L. R. A Statistical Method for Identifying Transition States in High Dimensional Problems. *J. Chem. Phys.* **1986**, *85*, 5045–5048.
- (13) Dellago, C.; Bolhuis, P. G.; Chandler, D. Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements. *J. Chem. Phys.* **1998**, *108*, 9236–9245.
- (14) Zuckerman, D. M.; Woolf, T. B. Dynamic reaction paths and rates through importance-sampled stochastic dynamics. *J. Chem. Phys.* **1999**, *111*, 9475–9484.
- (15) E, W.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2002**, *66*, 052301.
- (16) Faradjian, A. K.; Elber, R. Computing time scales from reaction coordinates by milestoneing. *J. Chem. Phys.* **2004**, *120*, 10880–10889.
- (17) Aristoff, D.; Bello-Rivas, J. M.; Elber, R. A Mathematical Framework for Exact Milestoning. *Multiscale Model. Simul.* **2016**, *14*, 301–322.
- (18) Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18043–18048.
- (19) Zwier, M. C.; Adelman, J. L.; Kaus, J. W.; Pratt, A. J.; Wong, K. F.; Rego, N. B.; Suárez, E.; Lettieri, S.; Wang, D. W.; Grabe, M.; et al. WESTPA: An Interoperable, Highly Scalable Software Package for Weighted Ensemble Simulation and Analysis. *J. Chem. Theory Comput.* **2015**, *11*, 800–809.
- (20) Stock, G.; Ghosh, K.; Dill, K. A. Maximum Caliber: A variational approach applied to two-state dynamics. *J. Chem. Phys.* **2008**, *128*, 194102.
- (21) Molloy, K.; Shehu, A. Elucidating the ensemble of functionally-relevant transitions in protein systems with a robotics-inspired method. *BMC Struct. Biol.* **2013**, *13*, S8.
- (22) Martini, L.; Kells, A.; Covino, R.; Hummer, G.; Buchete, N.-V.; Rosta, E. Variational Identification of Markovian Transition States. *Phys. Rev. X* **2017**, *7*, 031060.
- (23) Maximova, T.; Moffatt, R.; Ma, B.; Nussinov, R.; Shehu, A. Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics. *PLoS Comput. Biol.* **2016**, *12*, e1004619.
- (24) Chong, L. T.; Saglam, A. S.; Zuckerman, D. M. Path-sampling strategies for simulating rare events in biomolecular systems. *Curr. Opin. Struct. Biol.* **2017**, *43*, 88–94.
- (25) Wang, F.; Landau, D. P. Efficient, Multiple-Range Random Walk algorithm to Calculate the Density of States. *Phys. Rev. Lett.* **2001**, *86*, 2050–2053.
- (26) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.
- (27) Tiwary, P.; Berne, B. J. Spectral gap optimization of order parameters for sampling complex molecular systems. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2839–2844.
- (28) Mendels, D.; Piccini, G.; Parrinello, M. Collective Variables from Local Fluctuations. *J. Phys. Chem. Lett.* **2018**, *9*, 2776–2781.
- (29) Vijayan, R. S. K.; He, P.; Modi, V.; Duong-Ly, K. C.; Ma, H.; Peterson, J. R.; Dunbrack, R. L.; Levy, R. M. Conformational Analysis of the DFG-Out Kinase Motif and Biochemical Profiling of Structurally Validated Type II Inhibitors. *J. Med. Chem.* **2015**, *58*, 466–479.
- (30) Pearlman, D. A. A Comparison of Alternative Approaches to Free Energy Calculations. *J. Phys. Chem.* **1994**, *98*, 1487–1493.
- (31) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (32) Herschbach, D. R.; Johnston, H. S.; Rapp, D. Molecular Partition Functions in Terms of Local Properties. *J. Chem. Phys.* **1959**, *31*, 1652–1661.
- (33) Gō, N.; Scheraga, H. A. On the Use of Classical Statistical Mechanics in the Treatment of Polymer Chain Conformation. *Macromolecules* **1976**, *9*, 535–542.
- (34) Deng, N.; Zhang, B. W.; Levy, R. M. Connecting Free Energy Surfaces in Implicit and Explicit Solvent: An Efficient Method To Compute Conformational and Solvation Free Energies. *J. Chem. Theory Comput.* **2015**, *11*, 2868–2878.
- (35) Mobley, D. L.; Chodera, 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.
- (36) Park, S.; Lau, A. Y.; Roux, B. Computing conformational free energy by deactivated morphing. *J. Chem. Phys.* **2008**, *129*, 134102.
- (37) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (38) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (39) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950–1958.
- (40) Tan, Z.; Gallicchio, E.; Lapelosa, M.; Levy, R. M. Theory of Binless Multi-State Free Energy Estimation with Applications to Protein-Ligand Binding. *J. Chem. Phys.* **2012**, *136*, 144102.