

Practically Efficient and Robust Free Energy Calculations: Double-Integration Orthogonal Space Tempering

Lianqing Zheng[†] and Wei Yang*,^{†,‡}

[†]Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306, United States

[‡]Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32306, United States

ABSTRACT: The orthogonal space random walk (OSRW) method, which enables synchronous acceleration of the motions of a focused region and its coupled environment, was recently introduced to enhance sampling for free energy simulations. In the present work, the OSRW algorithm is generalized to be the orthogonal space tempering (OST) method via the introduction of the orthogonal space sampling temperature. Moreover, a double-integration recursion method is developed to enable practically efficient and robust OST free energy calculations, and the algorithm is augmented by a novel θ -dynamics approach to realize both the uniform sampling of order parameter spaces and rigorous end point constraints. In the present work, the double-integration OST method is employed to perform alchemical free energy simulations, specifically to calculate the free energy difference between benzyl phosphonate and difluorobenzyl phosphonate in aqueous solution, to estimate the solvation free energy of the octanol molecule, and to predict the nontrivial Barnase–Barstar binding affinity change induced by the Barnase NS8A mutation. As demonstrated in these model studies, the DI-OST method can robustly enable practically efficient free energy predictions, particularly when strongly coupled slow environmental transitions are involved.

I. INTRODUCTION

Developing free energy calculation methods has been a focal area in the quantitative aspect of molecular simulation.^{1–5} A major goal is to achieve accurate estimation of target free energy changes within as short as possible sampling length. Facing the bottleneck sampling challenge, various methods have been proposed; among many ingenious efforts, generalized ensemble (GE) based algorithms^{5–8} have attracted tremendous attention. The essential idea of GE free energy simulation methods is to employ a modified ensemble, which permits quick escaping of local energy wells, to efficiently produce accurate distributions for free energy estimations.^{5–8} In classical GE (or the first-order GE) free energy simulations, the design of a modified ensemble is focused on a prechosen order parameter λ , as reflected by the biasing energy term $f_m(\lambda)$ in the following target potential:

$$U_m = U_o(\lambda) + f_m(\lambda) \quad (1)$$

When λ is a spatial order parameter, $U_o(\lambda)$ represents the target energy function; when λ is an alchemical order parameter, $U_o(\lambda)$ stands for a hybrid energy function that is constructed on the basis of the constraints of $U_o(0) = U^A$ and $U_o(1) = U^B$ (then, two end states A and B are respectively represented by $\lambda = 0$ and $\lambda = 1$).^{9–18} In the first-order GE regime, the biasing term $f_m(\lambda)$ is adaptively updated to approach $-G_o(\lambda)$, which is the negative of the λ -dependent free energy profile corresponding to the canonical ensemble with $U_o(\lambda)$ as the potential energy function; thereby, an order parameter space random walk can be achieved to uniformly sample all the states in a target range. To adaptively estimate $G_o(\lambda)$, three major recursion approaches have been developed; they include the adaptive umbrella sampling method,¹⁹ in which free energy estimations are based on order parameter probability distributions, the adaptive biasing force (ABF) method²⁰ (in alchemical free energy simulations; it is called the generalized ensemble thermodynamic

integration method²¹ in the molecular dynamics scheme, or the adaptive integration method²² in the Monte Carlo or hybrid Monte Carlo scheme), in which free energy estimations are based on the thermodynamic integration (TI) formula,^{23,24} and the multiplicative approaches (including the metadynamics method for molecular dynamics simulations²⁵ and the Wang–Landau method for Monte Carlo or hybrid Monte Carlo simulations²⁶), which are realized through a dynamic force-balancing relationship. It is noted that various hybrid recursion methods based on the above three major approaches have been explored as well.^{27–35}

Although in first-order GE simulations, free energy surfaces along prechosen order parameters are flattened, “hidden” free energy barriers usually exist in the space perpendicular to the order parameter directions. Notably these “hidden” free energy barriers can impose great sampling challenges, e.g., slow environmental relaxations. This situation can be illustrated via a simple two-dimension model (Figure 1a), where the x direction represents the order parameter direction, and the y direction represents the “hidden” direction of the environmental response. The red dashed line in Figure 1b shows that the explicit free energy barrier (along the order parameter x direction with a value of about 12.8 kcal/mol) is lower than the real energy barrier [at the point of ($x = 0.0, y = 0.0$) with a value of about 15.0 kcal/mol as shown in Figure 1a] by about 2.2 kcal/mol; the existence of such a hidden energy barrier (e.g., the difference between the real energy barrier and the explicit free energy barrier) is due to the participation of the hidden dimension y in the energy increase near the transition state region (at $x = 0$). As discussed in our earlier works,^{36,37} the generalized force F_λ can serve as a collective variable to describe the

Received: October 14, 2011

Published: January 25, 2012



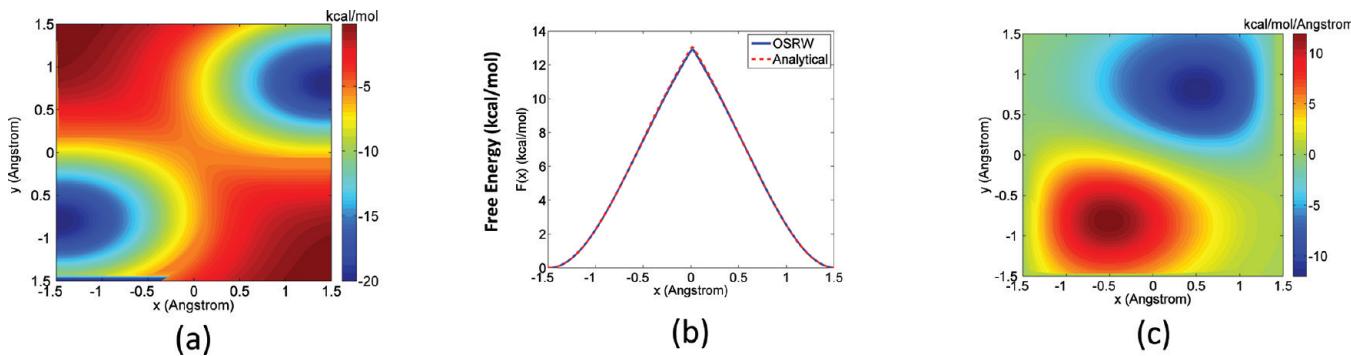


Figure 1. The model of the strongly coupled hidden free energy barrier issue. (a) The model energy surface where x represents the “known” order parameter direction and y represents the “hidden” direction of the environment response. (b) The free energy profile along the order parameter x direction. The red dashed line represents the analytical solution; the blue solid line represents the OSRW simulation estimation. (c) The spatial-dependent $(\partial U_o)/(\partial x)$ function.

progress of the hidden processes that strongly couple with the order parameter move. It is noted that F_λ is defined as $\partial U_o/\partial \lambda - RT (\partial \ln |\mathbf{J}| / \partial \lambda)$, where $|\mathbf{J}|$ is the Jacobian term corresponding to the transformation from the Cartesian system to a new system with λ as a coordinate direction, and it is equal to $\partial U_o/\partial x$ in this model case because of the fact that here an original Cartesian coordinate x is employed as the order parameter. The above insight was originally derived from the Marcus theory;³⁸ in our development,³⁷ we generalized the vertical energy gap,^{38,39} which was to describe electron transfer processes, to be the generalized force for the description of transitions between neighboring order parameter states; it can be clearly revealed by the spatial-dependent $\partial U_o/\partial x$ function (Figure 1c), which shows that near the state transition region [$x \in (-0.5, 0.5)$], $\partial U_o/\partial x$ decreases monotonically with the increase of y . Accordingly, the second-order GE simulation scheme, originally the orthogonal space random walk (OSRW) algorithm,^{36,37} was formulated as shown in the following modified energy function:

$$U_m = U_o(\lambda) + f_m(\lambda) + g_m(\lambda, F_\lambda) \quad (2)$$

where $f_m(\lambda)$ is targeted toward $-G_o(\lambda)$, and $g_m(\lambda, F_\lambda)$ is targeted toward $-G_o(\lambda, F_\lambda)$, the negative of the free energy profile along (λ, F_λ) in the ensemble corresponding to the energy function $U_o(\lambda) - G_o(\lambda)$. It is noted that, different from the first-order GE methods, OSRW requires two recursion components to respectively update $g_m(\lambda, F_\lambda)$ and $f_m(\lambda)$. The recursion component responsible for the update of $g_m(\lambda, F_\lambda)$ is called the “recursion kernel”, and the recursion component responsible for the update of $f_m(\lambda)$ is called the “recursion slave” because of the fact that the target of $f_m(\lambda)$, $-G_o(\lambda)$, depends on the target of $g_m(\lambda, F_\lambda)$: $-G_o(\lambda, F_\lambda)$.⁴⁰ In the original development, the recursion slave was based on the TI formula, and the metadynamics method was employed as the recursion kernel. Notably, in practice, the recursion kernel can be based on any of the three recursion methods as previously mentioned; one of the focuses in the present work is to design a robust recursion kernel.

Since its birth, the OSRW method has shown very encouraging sampling power; however, the originally implemented method suffers from the lack of robustness, especially in the aspect of long-time scale convergence. As elaborated in the Results and Discussion, two inter-related aspects contribute to this robustness issue: (1) because of the fact that free energy surfaces along generalized force directions are completely

flattened (e.g., the effective sampling temperature in the orthogonal space is infinity), there is no boundary to confine the orthogonal space sampling exploration; (2) the metadynamics-based recursion kernel can be replaced by a more robust recursion strategy. Regarding the first aspect, here, we are proposing to generalize the OSRW method to the orthogonal space tempering (OST) technique, which can be described through the following modified energy function:

$$U_m = U_o(\lambda) + f_m(\lambda) + \frac{T_{ES} - T_o}{T_{ES}} g_m(\lambda, F_\lambda) \quad (3)$$

where $g_m(\lambda, F_\lambda)$ is still targeted toward $-G_o(\lambda, F_\lambda)$; its contribution to the overall potential is scaled by a parameter of $(T_{ES} - T_o)/T_{ES}$; here T_o is the system reservoir temperature, and a preset parameter T_{ES} can be called the orthogonal space sampling temperature because of the fact that for any given λ' state, probability distributions in the target ensemble follow $\exp[-G_o(\lambda', F_\lambda)/kT_{ES}]$, where k is the Boltzmann constant. Thereby, the sampling boundary in the orthogonal space is naturally defined. In regard to the second aspect, we are proposing an alternative double-integration recursion scheme, where the recursion kernel is based on our customized ABF approach, the “dynamic reference restraining” (DRR) method. Notably, the long-time convergence of the ABF recursion strategy has been mathematically proven;⁴¹ therefore, we will employ this recursion approach as a key component of our recursion kernel design to ensure overall free energy recursion robustness.

In the present paper, the double-integration OST (DI-OST) method is described in the context of alchemical free energy simulation (or called the “free energy perturbation” calculation);^{9–18} for the purpose of GE sampling, the dynamics of the scaling parameter λ are introduced via a specially designed extended Hamiltonian scheme, which will be elaborated in the Algorithm Designs section. The present double-integration OST (DI-OST) method is demonstrated on alchemical free energy simulations, specifically to calculate the free energy difference between benzyl phosphonate and difluorobenzyl phosphonate in aqueous solution, to estimate the solvation free energy of the octanol molecule, and to predict the nontrivial Barnase–Barstar binding affinity change induced by the Barnase NS8A mutation. As shown in these model studies, the DI-OST method is a practically efficient and robust free energy calculation method, particularly when strongly coupled slow environmental transitions are involved.

II. ALGORITHM DESIGNS

II.A. A Novel θ -Dynamics Approach for Generalized Ensemble Alchemical Free Energy Simulations. The present work is only focused on alchemical free energy simulations, by which protein–ligand binding affinity changes, protein–protein binding affinity changes, solvation energies, pK_a values, and other chemical state related thermodynamic properties can be predicted.^{9–18} Certainly, the proposed DI-OST algorithm is also applicable to the geometry-based potential of mean force calculations; a related work will be published in the near future.

To carry out alchemical free energy calculations, as described in eq 1, a scaling parameter λ needs to be introduced to connect two target chemical states. A simplest hybrid energy function is the linear form:

$$U_o(\lambda) = (1 - \lambda)U_s^A + \lambda U_s^B + U_e \quad (4)$$

where U_s^A and U_s^B are the energy terms unique in the two end chemical states; U_e represents the common environmental energy terms shared by the two end states. When dummy atoms are employed in one of the end states, soft-cores potentials^{42,43} are commonly applied to treat the van der Waals terms or/and the electrostatic terms in U_s^A and U_s^B to avoid the end point singularity issue.

In GE alchemical free energy simulations, λ needs to be dynamically coupled with the motion of the rest of the system. Such extended dynamics can be realized either via the hybrid Monte Carlo method,^{44–46} where the scaling parameter jumps along a prearranged discrete λ ladder are enabled through the metropolis acceptance/rejection procedure, or via the λ -dynamics method,^{47,48} where λ moves in the continuous region between 0 and 1 are enabled through an extended Hamiltonian approach. The extended dynamics of the scaling parameter in OSRW are implemented on the basis of the λ -dynamics method. In the original λ -dynamics free energy calculation method, the scaling parameter λ is treated as a one-dimensional fictitious particle. In our implementation, especially to rigorously constrain λ between 0 and 1, a novel θ -dynamics approach is proposed. In this θ -dynamics, λ is set as the function $\lambda(\theta)$; the variable θ is treated as a one-dimensional fictitious particle, which travels periodically between $-\pi$ and π . It is noted that a similar approach was employed for multistate simulations,^{49,50} where $\lambda(\theta)$ is set as for instance $\sin^2 \theta$, and due to its Jacobian contribution in the related coordinate transformation, high densities of visits around two ends states can be achieved. In contrast, in OSRW simulations, uniform distributions are targeted. Here, the usage of the θ -dynamics approach is mainly for the purpose of constraining the λ range; actually, it is preferable to have uniform sampling in the λ space. For the above purpose, our functional form of $\lambda(\theta)$ is designed as follows:

$$\lambda(\theta) = \begin{cases} r \sin^2 \frac{\theta}{2}, & |\theta| \leq \theta_o \\ a\theta + b, & \theta_o < \theta < \pi - \theta_o \\ -a\theta + b, & \theta_o - \pi < \theta < -\theta_o \\ r \sin^2 \frac{\theta}{2} + c, & \pi - \theta_o \leq |\theta| \leq \pi \end{cases} \quad (5)$$

in which $r = 1/(1 - \cos \theta_o + 1/2(\pi - 2\theta_o)\sin \theta_o)$, $a = r/2 \sin \theta_o$, $b = r/2(1 - \cos \theta_o - \theta_o \sin \theta_o)$, and $c = r/2$

$(\pi - \theta_o)\sin \theta_o + r/2(1 - \cos \theta_o - \theta_o \sin \theta_o) - r \sin^2((\pi - \theta_o)/2)$. In eq 5, θ_o is the parameter utilized to separate the linear region and the end-state ($\lambda = 0, 1$) transition region. In OSRW and OST simulations, θ_o should be set as a tiny value so that A is almost 1 and B is almost zero; thus the Jacobian contribution from the $\lambda(\theta)$ function can be negligible. The propagation and the thermolysis of the θ particle are based on the Langevin equation, the same as how the λ particle is treated in the original λ -dynamics method.

II.B. A Brief Introduction to the Implementation of the Original OSRW Method. The OSRW method is based on the modified potential energy function as described in eq 2. The OSRW algorithm has two recursion components: the recursion kernel to adaptively update $g_m(\lambda, F_\lambda)$ toward its target function $-G_o(\lambda, F_\lambda)$ and the recursion slave to adaptively update $f_m(\lambda)$ toward its target function $-G_o(\lambda)$ based on the concurrent $g_m(\lambda, F_\lambda)$ function. In the original implementation, the metadynamics strategy²⁶ is employed as the recursion kernel. Specifically, the free energy biased potential $g_m(\lambda, F_\lambda)$ can be obtained by repetitively adding a relatively small Gaussian-shaped repulsive potential:

$$h_o \exp\left(-\frac{|\lambda - \lambda(t_i)|^2}{2w_1^2}\right) \exp\left(-\frac{|F_\lambda - F_\lambda(t_i)|^2}{2w_2^2}\right) \quad (6)$$

which is centered around $[\lambda(t_i), F_\lambda(t_i)]$ at the scheduled update time t_i , and thereby discourages the system from often visited configurations. With this procedure repeated, the overall biasing potential

$$g_m(\lambda, F_\lambda) = \sum_{t_i} h_o \exp\left(-\frac{|\lambda - \lambda(t_i)|^2}{2w_1^2}\right) \times \exp\left(-\frac{|F_\lambda - F_\lambda(t_i)|^2}{2w_2^2}\right) \quad (7)$$

will build up and eventually flatten the underlying curvature of the free energy surface in the (λ, F_λ) space. Then, the free energy profile along the reaction coordinate (λ, F_λ) , which should eventually converge to $-G_o(\lambda, F_\lambda)$, can be estimated as $-g_m(\lambda, F_\lambda)$.

Since for a state λ' , the free energy profile along its generalized force direction can be estimated as $-g_m[\lambda', F_\lambda(\lambda')]$, the generalized force distribution should be proportional to $\exp\{\beta_o g_m[\lambda', F_\lambda(\lambda')]\}$, in which β_o represents $1/(kT_o)$. On the basis of the above discussion, free energy derivatives at each state can be obtained as follows:

$$\frac{dG_o}{d\lambda} \Big|_{\lambda'} = \langle F_\lambda \rangle_{\lambda'} = \frac{\int_{F_\lambda} F_\lambda \exp\{\beta_o [g_m(\lambda, F_\lambda)]\} \delta(\lambda - \lambda')}{\int_{F_\lambda} \exp\{\beta_o [g_m(\lambda, F_\lambda)]\} \delta(\lambda - \lambda')} \quad (8)$$

Following the TI formula,^{23,24} the free energy change between the initial state with λ_i , which is the lower bound of the collective variable range, and any target state with the order parameter λ can unfold as a function of λ :

$$G_o(\lambda) = \int_{\lambda_i}^{\lambda} \frac{dG_o}{d\lambda} \Big|_{\lambda'} d\lambda' \quad (9)$$

In the original OSRW implementation, the metadynamics strategy, as described in eq 7, serves as the recursion kernel; the

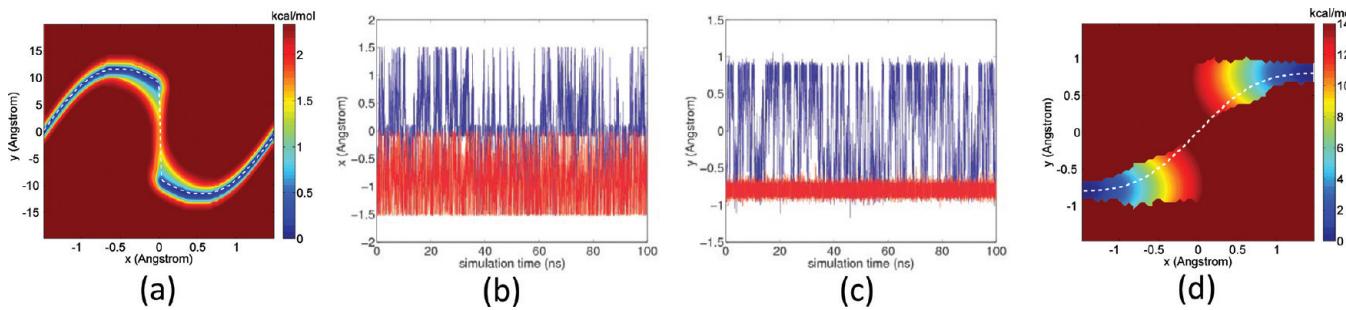


Figure 2. The model illustration of the OSRW method. (a) The contour plot is the converged profile $-g_m(x, (\partial U_o)/(\partial x))$ obtained from the OSRW simulation; the white dashed line is the free energy derivative curve. (b) The time dependent x changes in the equilibrium generalized ensemble simulations based on the biasing potentials, $f_m(x)$ (red) and $f_m(x) + g_m(x, (\partial U_o)/(\partial x))$ (blue). (c) The time dependent y changes in the equilibrium generalized ensemble simulations based on the biasing potentials, $f_m(x)$ (red) and $f_m(x) + g_m(x, (\partial U_o)/(\partial x))$ (blue). (d) The free energy surface obtained on the basis of the samples in the equilibrium generalized ensemble simulation based on the biasing potential of $f_m(x) + g_m(x, (\partial U_o)/(\partial x))$; the minimum energy path connecting the two energy wells is presented by the white dashed line.

TI based formula (eqs 8 and 9) serves as the recursion slave with $f_m(\lambda)$ recursively set as instantaneously estimated $-G_o(\lambda)$.

II.B. A Model Illustration and Analysis. On the basis of the above OSRW procedure, we carried out a free energy simulation study on the model system (as illustrated in Figure 1). The model simulation was performed on the basis of two-dimensional Langevin dynamics, where the temperature was set as 50 K and the particle mass was set as 100 g/mol. The OSRW simulation led to a converged free energy profile $G_o(x)$ [targeted as $-f_m(x)$], as shown by the blue line in Figure 1b and a converged $-g_m(x, \partial U_o/\partial x)$ (in the model case, $\partial U_o/\partial x$ is the generalized force) as shown in Figure 2a, where two energy minima are smoothly connected along $\partial U_o/\partial x$ at the transition state region. When converged, this represents the residual free energy surface after the free energy surface flattening treatment $-g_m(x, \partial U_o/\partial x)$ along the order parameter. Figure 2a [$-g_m(x, \partial U_o/\partial x)$] reveals the fact that the residual free energy barrier exists around the transition state region. It can be traced along $\partial U_o/\partial x$ near the transition state, and more importantly, the residual barrier height (about 2.2 kcal/mol) is similar to that of the hidden energy barrier. Obviously, in this model system, the generalized force can reveal the direction of the order-parameter-coupled hidden process; this is certainly a prerequisite for efficient and accurate calculations of the target free energy profile $G_o(x)$. In Figure 2a, the x -dependent free energy derivative dG_o/dx curve is shown as a white dashed line; as discussed in the context of eq 8, this curve was generated on the basis of the concurrent recursion kernel result $g_m(x, \partial U_o/\partial x)$.

To further understand the role of $\partial U_o/\partial x$ and the difference between the OSRW sampling [e.g., based on $U_o + f_m(x) + g_m(x, \partial U_o/\partial x)$ as in eq 2] and the classical generalized ensemble sampling [e.g., based on $U_o + f_m(x)$ as in eq 1], we respectively employed the biasing energy functions $f_m(x)$ and $f_m(x) + g_m(x, \partial U_o/\partial x)$, which were obtained in the recursion step, to perform two corresponding equilibrium generalized ensemble simulations. As shown in Figure 2b and c, the OSRW sampling (the blue lines) allows the system to travel repetitively between two energy minima; as a comparison in the classical generalized ensemble simulation (the red lines), the system is trapped in the original energy minimum state due to the lack of sampling acceleration along the hidden dimension. Furthermore, according to the umbrella sampling reweighting relationship, the samples collected from the OSRW simulation can be employed to recover the free energy surface (Figure 2d) along

x and y , the well-sampled region of which is the same as the target energy surface (Figure 1a). As shown from this recovered free energy surface, the samples are more concentrated along the minimum energy path (shown by the dash white line) that connects two energy wells.

II.C. The Orthogonal Space Tempering (OST) Method.

In an OSRW simulation, the sampling volume in the orthogonal space increases with the elongation of the simulation length; additionally, the diffusion sampling overhead around the states, where no hidden barrier exists, continuously increases. As discussed in the Introduction section, the OSRW method can be generalized to the orthogonal space tempering (OST) algorithm. The target energy function of the OST scheme is described in eq 3. In the OST scheme, free energy surfaces along the generalized force direction are not completely flattened. Then, the orthogonal space effective sampling temperature T_{ES} can impose an effective sampling boundary to ensure the long-time scale convergence. A larger T_{ES} allows more efficient crossing of hidden free energy barriers but introduces more diffusion sampling overhead. Interestingly, when T_{ES} approaches the infinity limit, the OST method becomes the original OSRW algorithm; when T_{ES} approaches the system reservoir temperature T_o , the second-order GE sampling turns to the first-order GE sampling as described in eq 1.

II.D. The Novel Double-Integration Recursion Method to Enable Robust Orthogonal Space Tempering (OST) Simulations. The metadynamics method achieves adaptive recursions based on a dynamic force-balancing relationship. As it is known, its performance strongly depends on energy surface ruggedness and preset parameters.⁵¹ To improve the convergence behavior of OST, in the present work, we designed an alternative method to gain robust recursions.

Among various recursion methods, the adaptive biasing force (ABF) algorithm has a similar efficiency to that of the metadynamics algorithm. In contrast to the metadynamics technique, the ABF method has been mathematically proven;⁴¹ thus the usage of the ABF method as the recursion kernel, specifically via the calculation of the F_λ -dependent free energy profile $G_o(\lambda', F_\lambda)$ at each λ' state, can ensure free energy convergence robustness. A challenging issue remains: how to numerically calculate the generalized force of F_λ to estimate target F_λ -dependent free energy profiles. As a matter of fact, calculating generalized forces of complex order parameters has been known to be a difficult issue in the ABF algorithm implementation.⁵² To circumvent this issue, in our OST

implementation, we propose a “dynamic reference restraining” (DRR) recursion strategy. Specifically, the target OST potential in eq 3 is rewritten as follows:

$$\begin{aligned} U_m = & U_o(\lambda) + \frac{1}{2}k_\phi(F_\lambda - \phi)^2 + f_m(\lambda) \\ & + \frac{T_{ES} - T_o}{T_{ES}}g_m(\lambda, \phi) \end{aligned} \quad (10)$$

in which the generalized force fluctuation is restrained to the move of another dynamic particle ϕ . In eq 10, $f_m(\lambda)$ is still targeted toward $-G_o(\lambda)$, and $g_m(\lambda, \phi)$ is targeted toward $-G_o(\lambda, \phi)$, the negative of the free energy surface along (λ, ϕ) in the canonical ensemble with the energy function $U_o(\lambda) + 1/2k_\phi(F_\lambda - \phi)^2 - G_\phi(\lambda)$, where $G_\phi(\lambda)$ is the λ -dependent free energy surface in the canonical ensemble with $U_o(\lambda) + 1/2k_\phi(F_\lambda - \phi)^2$ as the energy function. On the basis of eq 10, motions along F_λ are indirectly activated via the restraining treatment to the dynamic reference: ϕ . Here, the dynamics of the ϕ particle are also realized through the same extended Hamiltonian method as in λ -dynamics or θ -dynamics, which is discussed in section II.A.

According to the OST target function in eq 10, we need to design a recursion kernel to estimate $G_o(\lambda, \phi)$ in order to adaptively update $g_m(\lambda, \phi)$. To obtain the two-dimensional function $G_o(\lambda, \phi)$, first, the ABF method is directly employed to calculate the ϕ -dependent free energy profile at each λ' state, specifically on the basis of the following TI relationship:

$$G_o'(\lambda', \phi) = \int_\phi \left\langle \frac{\partial U_o(\lambda, \phi)}{\partial \phi} \delta(\lambda - \lambda') \right\rangle_{\phi'} d\phi' \quad (11)$$

Here, $U_o(\lambda, \phi)$ represents $U_o(\lambda) + 1/2k_\phi(F_\lambda - \phi)^2$; then $\partial U_o(\lambda, \phi)/\partial \phi$ can be simply evaluated as $-k_\phi(F_\lambda - \phi)$. It is noted that the numerical boundary of $G_o(\lambda', \phi)$, i.e., the integration boundary in eq 11, changes as the recursion proceeds. Following the general ABF strategy, $\langle \partial U_o(\lambda, \phi) / \partial \phi \delta(\lambda - \lambda') \rangle_{\phi'}$ can be adaptively estimated as

$$\frac{\sum_i -k_\phi[F_\lambda(t_i) - \phi(t_i)] \delta[\lambda(t_i) - \lambda'] \delta[\phi(t_i) - \phi']}{\sum_i \delta[\lambda(t_i) - \lambda'] \delta[\phi(t_i) - \phi']} \quad (12)$$

where t_i is the i th scheduled sample-collecting time. Equations 11 and 12 only allow the obtaining of the one-dimension function $G_o(\lambda', \phi)$ at each λ' state. The height of the $G_o(\lambda', \phi)$ function can be recalibrated as

$$\begin{aligned} G_o''(\lambda', \phi) = & G_o'(\lambda', \phi) - G_{o', \min}(\lambda', \phi) \\ & - RT \ln \int_\phi \exp\left(-\frac{G_{o'}(\lambda', \phi) - G_{o', \min}(\lambda', \phi)}{kT_o}\right) \end{aligned} \quad (13)$$

where $G_{o', \min}(\lambda', \phi)$ is the lowest value in the free energy curve $G_o(\lambda', \phi)$; $G_{o'}(\lambda', \phi)$ represents the postcalibration function of $G_o(\lambda', \phi)$. All of the calibrated one-dimension $G_o(\lambda', \phi)$ functions can be assembled to be the target two-dimension $G_o(\lambda, \phi)$ function. Then, $g_m(\lambda, \phi)$ can be adaptively updated as instantaneously estimated $-G_o(\lambda, \phi)$. This calibration procedure is based on the $g_m(\lambda, \phi)$ function definition in eq 10, specifically to fulfill the condition that the target energy function for $g_m(\lambda, \phi)$ free energy flattening treatment has already been

flattened along the λ direction. In the proposed DI-OST method, eqs 11–13 constitute the recursion kernel.

Regarding the recursion slave, the TI formula in eq 9 is still used to estimate $G_o(\lambda)$; then, $(dG_o/d\lambda)|_{\lambda'}$ at each λ' state needs to be evaluated. Different from the recursion in the original OSRW algorithm, where the target function of the recursion kernel is $-G_o(\lambda, F_\lambda)$, here, the target function of the recursion kernel $-G_o(\lambda, \phi)$ does not provide direct information on generalized force F_λ distributions. For the fact that F_λ is restrained to ϕ , a simple but an approximate way of estimating $(dG_o/d\lambda)|_{\lambda'}$ can be made on the basis of the assumption of $\langle \phi \rangle_{\lambda'} = \langle F_\lambda \rangle_{\lambda'}$. Thus, $(dG_o/d\lambda)|_{\lambda'}$ can be approximated via the following equation:

$$\frac{dG_o}{d\lambda} \Big|_{\lambda'} = \langle F_\lambda \rangle_{\lambda'} \approx \langle \phi \rangle_{\lambda'} = \frac{\int_\phi \phi \exp\{\beta[g_m(\lambda, \phi)]\} \delta(\lambda - \lambda')}{\int_\phi \exp\{\beta[g_m(\lambda, \phi)]\} \delta(\lambda - \lambda')} \quad (14)$$

To more rigorously estimate $(dG_o/d\lambda)|_{\lambda'}$, $G_o(\lambda', F_\lambda)$ needs to be calculated for each λ' state; here, the $G_o(\lambda', F_\lambda)$ definition is the same as that in section II.B. Notably, the samples collected at the state λ' with $F_\lambda = F_{\lambda'}$ can be considered as being obtained from multiple independent ensembles, each of which corresponds to a unique restraining reference value ϕ' . According to the umbrella integration relationship,⁵³ based on the samples from each (λ', ϕ') restraining ensemble, $(dG_o(\lambda', F_\lambda)/dF_\lambda)|_{F_{\lambda'}, \lambda'}$ can be estimated as $1/(\beta_o)(F_{\lambda'} - \overline{F_{\lambda'}^{\lambda', \phi'}})/(\sigma_{\lambda'}^{\lambda', \phi'})^2 - k_\phi(F_{\lambda'} - \phi')$, where $\overline{F_{\lambda'}^{\lambda', \phi'}}$ stands for the average of the F_λ values of all of the samples in the (λ', ϕ') restraining ensemble and $\sigma_{\lambda'}^{\lambda', \phi'}$ represents the variance of these samples. Using the multihistogram approach to combine the estimations from all of the restraining ensembles that are visited at the λ' state, $(dG_o(\lambda', F_\lambda)/dF_\lambda)|_{F_{\lambda'}, \lambda'}$ can be calculated as

$$\frac{dG_o(\lambda', F_\lambda)}{dF_\lambda} \Big|_{F_{\lambda'}, \lambda'} = \frac{\int_{\phi'} \rho(\phi', \lambda', F_{\lambda'}) \left[\frac{1}{\beta_o} \frac{F_{\lambda'} - \overline{F_{\lambda'}^{\lambda', \phi'}}}{(\sigma_{\lambda'}^{\lambda', \phi'})^2} - k_\phi(F_{\lambda'} - \phi') \right]}{\int_{\phi'} \rho(\phi', \lambda', F_{\lambda'})} \quad (15)$$

where $\rho(\phi', \lambda', F_{\lambda'})$ denotes the total number of the $(\lambda', F_{\lambda'})$ samples that are collected from the ϕ' restraining ensemble.

Then, based on the TI relationship, $G_o(\lambda', F_\lambda)$ can be calculated:

$$G_o'(\lambda', F_\lambda) = \int_{F_{\lambda'}} \frac{dG_o(\lambda', F_\lambda)}{dF_\lambda} \Big|_{F_{\lambda'}, \lambda'} dF_{\lambda'} \quad (16)$$

Again, like in eq 11, the numerical boundary of $G_o(\lambda', F_\lambda)$, i.e., the integration boundary in eq 16, changes as the recursion proceeds. Following the corresponding derivation in the original OSRW method, we can obtain $(dG_o/d\lambda)|_{\lambda'}$ at the state λ'

$$\frac{dG_o}{d\lambda} \Big|_{\lambda'} = \langle F_\lambda \rangle_{\lambda'} = \frac{\int_{F_\lambda} F_\lambda \exp\{-\beta_o[G_o'(\lambda, F_\lambda)]\} \delta(\lambda - \lambda')}{\int_{F_\lambda} \exp\{-\beta_o[G_o'(\lambda, F_\lambda)]\} \delta(\lambda - \lambda')} \quad (17)$$

On the basis of the corresponding TI formula in eq 9, $f_m(\lambda)$, which is targeted as $-G_o(\lambda)$, can then be adaptively updated. In the proposed DI-OST method, eqs 15–17 and 9 constitute the recursion slave. Notably, $f_m(\lambda)$ does not have to be equal to $-G_o(\lambda)$ in a strict manner; here, it is highly recommended to employ the approximate approach based on eqs 11–14 and 9 to update $f_m(\lambda)$, and the more rigorous approach based on eqs 15–17 and 9 to estimate $G_o(\lambda)$, because of the fact that $\langle \phi \rangle_\lambda$ in eq 14, is directly estimated from ϕ -space ABF calculations (eqs 11 and 12) and should converge faster. In the DI-OST method, both the recursion kernel and the recursion slave are based on the integration schemes. Therefore, it is named the double-integration recursion method.

III. COMPUTATIONAL DETAILS

The double-integration recursion based OST method is implemented in the “orthogonal space sampling module”, which is currently coupled with our customized CHARMM program.^{54,55} In the present study, the following van der Waals soft-core potential form is employed to treat the atoms which are annihilated:

$$U_{\text{softcore}}^{\text{vdW}} = (1 - \lambda) \left[\frac{A}{(\alpha_{\text{vdW}} \lambda^2 + r^6)^2} - \frac{B}{\alpha_{\text{vdW}} \lambda^2 + r^6} \right] \quad (18)$$

where α_{vdW} is the van der Waals soft-core shifting parameter. It is noted that eq 18 is different from the one in the currently released CHARMM program.⁵⁵ The electrostatic soft-core potential is based on a common equation:

$$U_{\text{softcore}}^{\text{elec}} = \frac{(1 - \lambda) Q_A Q_B}{\sqrt{\alpha_{\text{elec}} \lambda + r^2}} \quad (19)$$

where α_{elec} is the electrostatic soft-core shifting parameter. In eqs 18 and 19, the annihilation is assumed to occur at the state of $\lambda = 1$; to be consistent, in this study, all of the dummy atoms are set at the state of $\lambda = 1$.

In the present study, the DI-OST method is demonstrated in the context of alchemical free energy simulation, specifically to calculate the free energy difference between benzyl phosphonate and difluorobenzyl phosphonate in aqueous solution, to estimate the solvation free energy of the octanol molecule, and to predict the Barnase–Barstar nontrivial binding affinity change induced by the Barnase N58A mutation.

III.A. The Free Energy Difference between Benzyl Phosphonate and Difluorobenzyl Phosphonate in Aqueous Solution. The molecules of benzyl phosphonate (BP) and difluorobenzyl phosphonate (F2BP) are the side chain analogues of prototypical phosphotyrosine mimetics, which are common targets in drug discovery.⁵⁶ The free energy difference between these two molecules in aqueous solution, $\Delta G_{\text{BP} \rightarrow \text{F2BP}}^{\text{aqueous}}$, has been used as a test-bed to analyze free energy simulation methods.⁵⁷ In practical studies, if combined with the free energy difference in gas phase $\Delta G_{\text{BP} \rightarrow \text{F2BP}}^{\text{gas}}$, $\Delta G_{\text{BP} \rightarrow \text{F2BP}}^{\text{aqueous}} - \Delta G_{\text{BP} \rightarrow \text{F2BP}}^{\text{gas}}$ gives rise to the value of the solvation energy difference; if combined with the free energy difference in a protein binding site $\Delta G_{\text{BP} \rightarrow \text{F2BP}}^{\text{protein}}$, $\Delta G_{\text{BP} \rightarrow \text{F2BP}}^{\text{protein}} - \Delta G_{\text{BP} \rightarrow \text{F2BP}}^{\text{aqueous}}$ gives rise to the value of the binding free energy difference. Here, the test calculations on $\Delta G_{\text{BP} \rightarrow \text{F2BP}}^{\text{aqueous}}$ are used to comparatively

evaluate the original OSRW method and the currently proposed DI-OST method in the aspects of algorithm robustness and long-time convergence. On the basis of each of the two methods, five sets of independent simulations were carried out. The MD simulation setup was the same as the one in the earlier studies, where the BP and F2BP molecules are described with the CHARMM22 parameter.⁵⁶ In total, 294 water molecules are included in the truncated octahedral box; the water molecules are treated with the TIP3P model.⁵⁸ Figure 3 shows the setup of the



Figure 3. The setup of the alchemical transition from benzyl phosphonate (BP) and difluorobenzyl phosphonate (F2BP).

alchemical transition from BP to F2BP. For the fact that there is no vanishing atom in either of the end states, the linear hybrid energy function (as described by eq 4) is used in this model study.

In the five OSRW simulations, $g(\lambda, F_\lambda)$ (in eq 7) was updated every 10 time steps; the height of the Gaussian function h was set as 0.01 kcal/mol; the widths of the Gaussian function, ω_1 and ω_2 , were set as 0.01 and 4 kcal/mol respectively; and $f_m(\lambda)$ was updated (based on eqs 8 and 9) once per 1000 time steps. In the five DI-OST simulations, the samples were collected every time step; $g(\lambda, \phi)$ was updated (based on eqs 11–13) once per 1000 time steps; $f_m(\lambda)$ was updated (based on eqs 17–19 and 9) once per 1000 time steps; and T_{ES} was set as 600 K (the system reservoir temperature is 300 K). The length of each simulation is 20 nanoseconds (ns).

III.B. The Solvation Free Energy of the Octanol Molecule. The model calculation on the octanol solvation free energy is to understand the role of the orthogonal space sampling temperature T_{ES} . The octanol molecule (Figure 4),

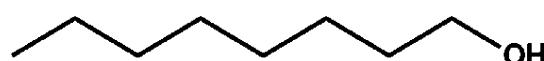


Figure 4. The octanol molecule.

which is described by the CHARMM general force field (CGFF),⁵⁹ is embedded in a truncated octahedral water box with a total of 713 TIP3P water molecules. In the alchemical free energy simulation setup, the solvated octanol molecule ($\lambda = 0$) is changed to a gas phase molecule ($\lambda = 1$), which does not have any interaction with the solvent molecules. Accordingly, all of the van der Waals and the electrostatic energy terms describing the solute–solvent interactions are subject to the soft-core treatment, in which α_{vdW} is set as 0.5 and α_{elec} is set as 5.0. Then, the solvation free energy of octanol $G_{\text{octanol}}^{\text{solvaton}}$ can be estimated as the negative of the free energy difference $-\Delta G_{\lambda=0 \rightarrow \lambda=1}$ between the two end states.

To understand the influence of T_{ES} on sampling efficiency, two sets of independent DI-OST simulations were run, each of which includes eight simulations with T_{ES} respectively set as 750 and 375 K (the system reservoir temperature is 300 K). The samples were collected every time step. $g_m(\lambda, \phi)$ was updated (based on eqs 11–13) once per 1000 time steps. $f_m(\lambda)$ was also updated (based on eqs 17–19 and 9) once per 1000 time steps. The length of each simulation is 17 ns.

III.C. The Barnase–Barstar Binding Affinity Change upon the Barnase N58A Mutation. The model study on the binding between barnase, an extracellular RNase of *Bacillus amyloliquefaciens*, and barstar, the intracellular polypeptide inhibitor of barnase (Figure 5b), is to demonstrate the DI-OST

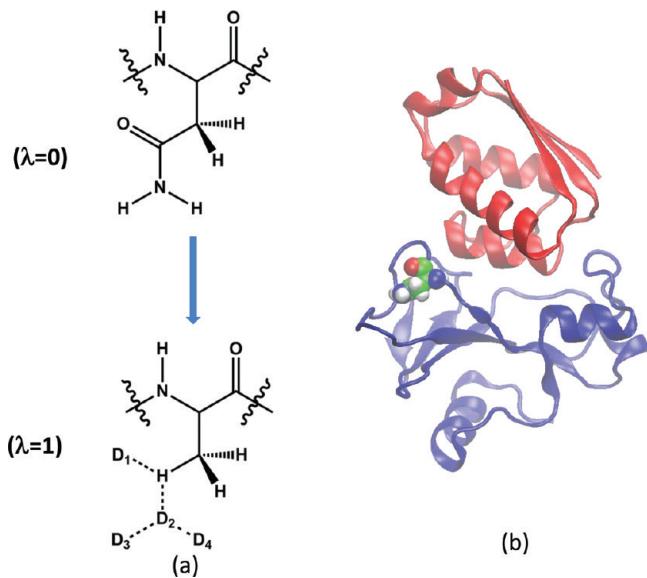


Figure 5. (a) The setup of the Barnase NS9A alchemical mutation. (b) The structure of the Barnase–Barstar complex. Barnase is colored in blue, and Barstar is colored in red. The Asn59 residue is colored by atoms.

method in predicting mutation induced protein–protein binding affinity changes. The barnase NS8A mutation is located at the second layer of the binding interface; this noncharging mutation causes about 3.1 kcal/mol of the binding affinity loss.⁶⁰ The DI-OST simulations were performed to calculate the alchemical free energy changes in two environments: $\Delta G_{\text{Asn} \rightarrow \text{Ala}}^{\text{complex}}$ in the barnase–barstar complex and $\Delta G_{\text{Asn} \rightarrow \text{Ala}}^{\text{barnase}}$ in the unbound barnase. The binding affinity change $\Delta\Delta G_{\text{Asn} \rightarrow \text{Ala}}$ can be calculated as $\Delta G_{\text{Asn} \rightarrow \text{Ala}}^{\text{complex}} - \Delta G_{\text{Asn} \rightarrow \text{Ala}}^{\text{barnase}}$. All of the systems are treated with the CHARMM27/CMAP model.⁶¹ In the model for the $\Delta G_{\text{Asn} \rightarrow \text{Ala}}^{\text{complex}}$ calculation, the barnase–barstar complex (with the PDB code of 1BRS) is embedded in the octahedral box with 18 902 water molecules; in the model for the $\Delta G_{\text{Asn} \rightarrow \text{Ala}}^{\text{barnase}}$ calculation, the unbound barnase (also based on the PDB code of 1BRS) is embedded in the octahedral box with 11 291 water molecules.

In the alchemical free energy simulation setup (Figure 5a), the vanishing atoms in Asn58 ($\lambda = 0$) are switched to the corresponding dummy atoms at $\lambda = 1$. The bond, angle, and dihedral terms associated with the dummy atoms are set identical to the corresponding ones of the original asparagine residue. All of the van der Waals terms and the electrostatic energy terms associated with the dummy atoms are subject to the soft-core treatment, in which α_{vdW} was set as 0.5 and α_{elec} was set as 5.0. The three DI-OST simulations were performed with T_{ES} set as 1500 K (the system reservoir temperature is 300 K); the samples were collected every time step. $g(\lambda, \phi)$ was updated (based on eqs 11–13) once per 1000 time steps. $f_m(\lambda)$ was also updated (based on eqs 17–19 and 9) once per 1000 time steps.

III.D. The General Molecular Dynamics Simulation Setup.

The CGFF parameters were generated through the

CHARMM-GUI server.⁶² The particle mesh ewald (PME) method⁶³ was applied to take care of the long-range columbic interactions while the short-range interactions were totally switched off at 12 Å. The Nose–Hoover method⁶⁴ was employed to maintain a constant reservoir temperature at 300 K, and the Langevin piston algorithm⁶⁵ was used to maintain the constant pressure at 1 atm. The time step was set as 1 fs.

IV. RESULTS AND DISCUSSION

IV.A. The Free Energy Difference between Benzyl Phosphonate and Difluorobenzyl Phosphonate in Aqueous Solution: Robustness and Long-Time Convergence. The results from one of the five DI-OST simulations are summarized in Figure 6. In about 800 ps, the scaling parameter λ completed the first one-way trip, which started at $\lambda = 0$ (Figure 6a). It is noted that free energy estimations are only possible when the sampling covers the entire λ space. At 820 ps, the initial estimation of $\Delta G_{\text{BP} \rightarrow \text{F2BP}}^{\text{aqueous}}$ gives 299.91 kcal/mol, which is very close to the finally converged result 299.77 kcal/mol. In the DI-OST scheme, the $((T_{\text{ES}} - T_o)/T_{\text{ES}})g_m(\lambda, \phi)$ biasing term enables the accelerating of ϕ moves, which through the restraint term $1/2k_\phi(F_\lambda - \phi)^2$ induces simultaneous fluctuation enlargement of the generalized force F_λ . In these simulations, the restraint force constant k_ϕ was set as 0.1 (kcal/mol)⁻¹; F_λ and ϕ are robustly synchronized as revealed in Figure 6b. As shown by the red line in Figure 6c, the recursive orthogonal space tempering treatment allows F_λ fluctuations to be continuously enlarged until around 8 ns; then the ϕ space sampling boundary imposed by T_{ES} was reached. Subsequent recursion kernel and recursion slave updates enable continuous refinement of the $g_m(\lambda, \phi)$ and $f_m(\lambda)$ terms. Figure 6d illustrates the $g_m(\lambda, \phi)$ function that was generated at the end of the 20 ns simulation; as shown, the orthogonal space sampling temperature 600 K allows the fluctuations of ϕ and F_λ to overcome $\sim 9KT$ strongly coupled free energy barriers that are hidden in the orthogonal space.

The BP and D2BP molecules differ only in their local polarity. One would expect moderate environment changes to be associated with the target alchemical transition; simulating the BP-D2BP transition may not fully demonstrate the sampling power of the DI-OST method. However, for its simplicity, this is an ideal system to test the robustness and the long-time convergence behavior of a free energy simulation method. As shown in Figure 7a and b, the estimated free energies from the five DI-OST simulations converge to the average value of 299.77 kcal/mol, which quantitatively agrees with the results obtained from the classical free energy simulation studies.^{56,57} Notably, as mentioned in the Computational Details, in this model study, we only targeted our calculations on the estimation of the alchemical free energy difference $\Delta G_{\text{BP} \rightarrow \text{F2BP}}^{\text{aqueous}}$, the value of which alone does not have any physical meaning. With 20 ns of the simulation lengths, the variance of the five independently estimated values is as low as 0.01 kcal/mol. Within only 940 picoseconds (ps), all five DI-OST simulations had completed their first one-way trips. Then, the average of the estimated values is 299.82 kcal/mol, and the variance of the calculation results is 0.12 kcal/mol. In 2 ns, the average of the estimated values converges to 299.79 kcal/mol, and the variance of the calculation results is 0.04 kcal/mol. In DI-OST simulations, $G_o(\lambda)$ [the negative of $f_m(\lambda)$] should converge faster than $G_o(\lambda, \phi)$ [the negative of $g_m(\lambda, \phi)$] because of the fact that the free energy derivative $dG_o(\lambda)/d\lambda$ is largely determined by the lower region of the free energy surface

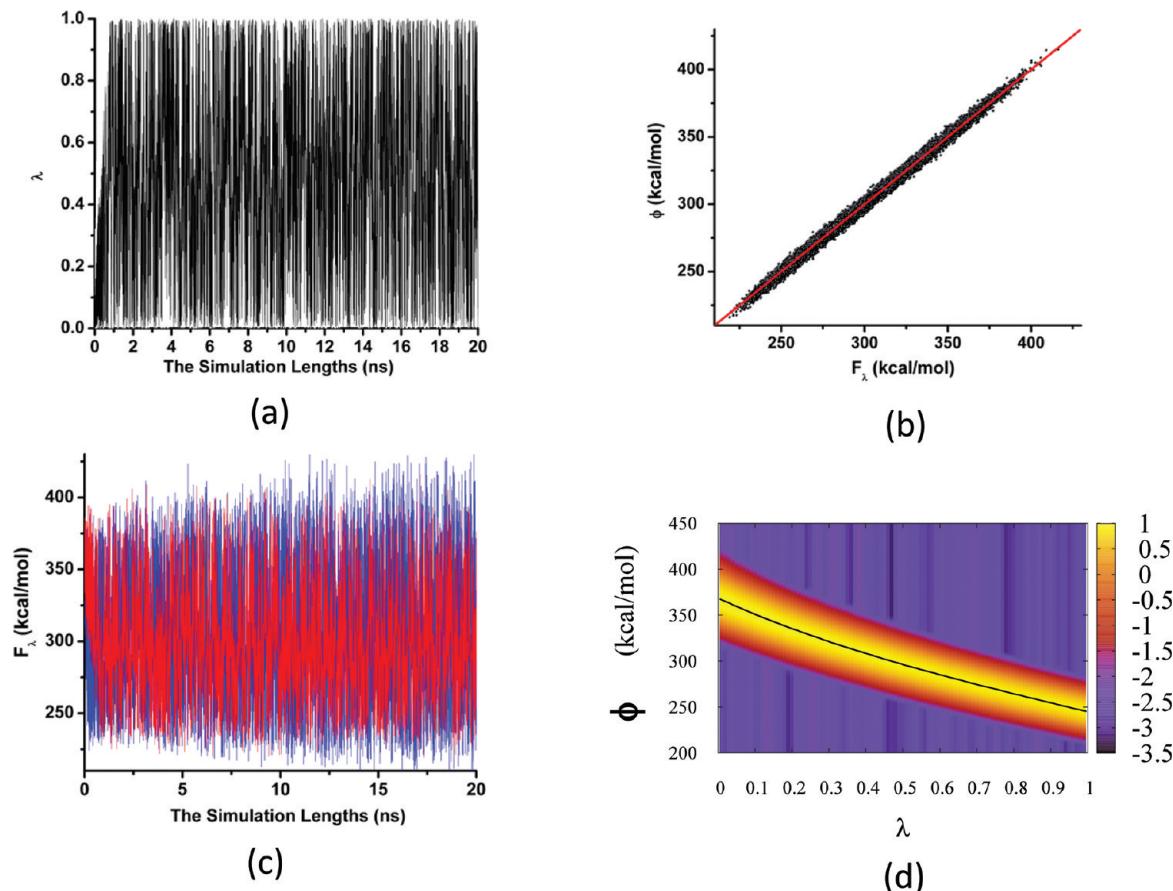


Figure 6. The results of one of the five DI-OST simulations on the BP-F2BP transition. (a) The time-dependent λ changes. (b) The F_λ and ϕ synchronization; $F_\lambda = \phi$ is represented by the red line. (c) The time-dependent F_λ changes. The result from the DI-OST simulation is shown in red, and the result from a comparison OSRW simulation is shown in blue. (d) The $g_m(\lambda, \phi)$ function generated from 20 ns of the DI-OST simulation.

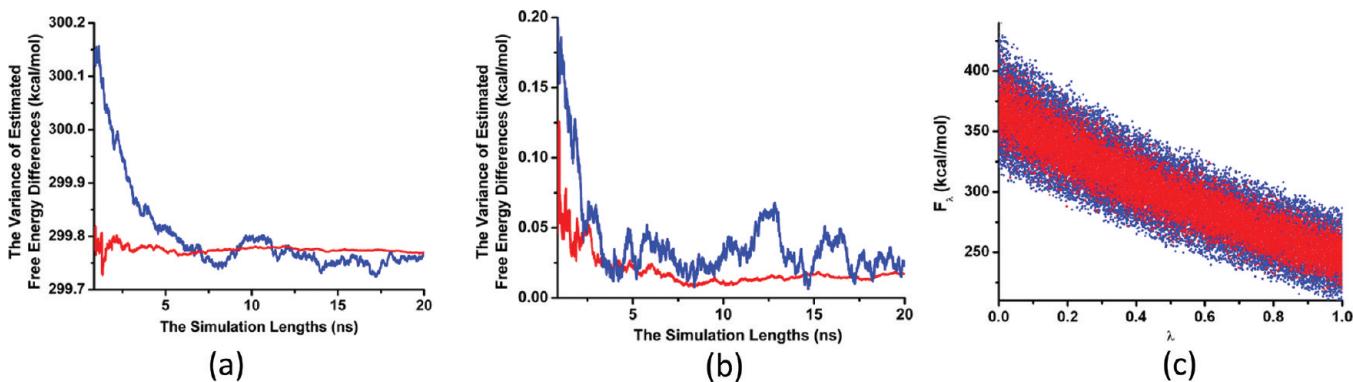


Figure 7. The comparison of the DI-OST and OSRW simulation results on the BP-F2BP transition. (a) The time-dependent free energy averages from the five independent DI-OST (in red) and OSRW (in blue) simulations. (b) The time-dependent variances among the free energy values estimated from the five independent DI-OST (in red) and OSRW (in blue) simulations. (c) The representation of the samples in the (λ, F_λ) space. The samples from a representative DI-OST simulation are colored in red, and the samples from a representative OSRW simulation are colored in blue.

along (λ, F_λ) . Besides the sampling efficiency, the free energy estimation robustness and the long-time convergence rigorousness of the DI-OST method can be clearly revealed by the blue curves in Figure 7a and b.

As discussed earlier, the original OSRW method is limited in two aspects. First, the orthogonal space sampling temperature T_{ES} is effectively infinity; thus, there is no boundary to restrict the magnitude of F_λ fluctuation enlargement. As shown by the

blue curve in Figure 6c, the orthogonal space free energy surface flattening treatment enlarges F_λ fluctuations boundlessly. In comparison with the DI-OST simulations, which have their sampling boundaries imposed by the finite T_{ES} value (600 K), the OSRW simulations have ever-increasing sampling coverage (Figure 7c). Consequently, both the average and the variance of the free energy results show time-dependent oscillatory behaviors (the blue curves in Figure 7b and c).

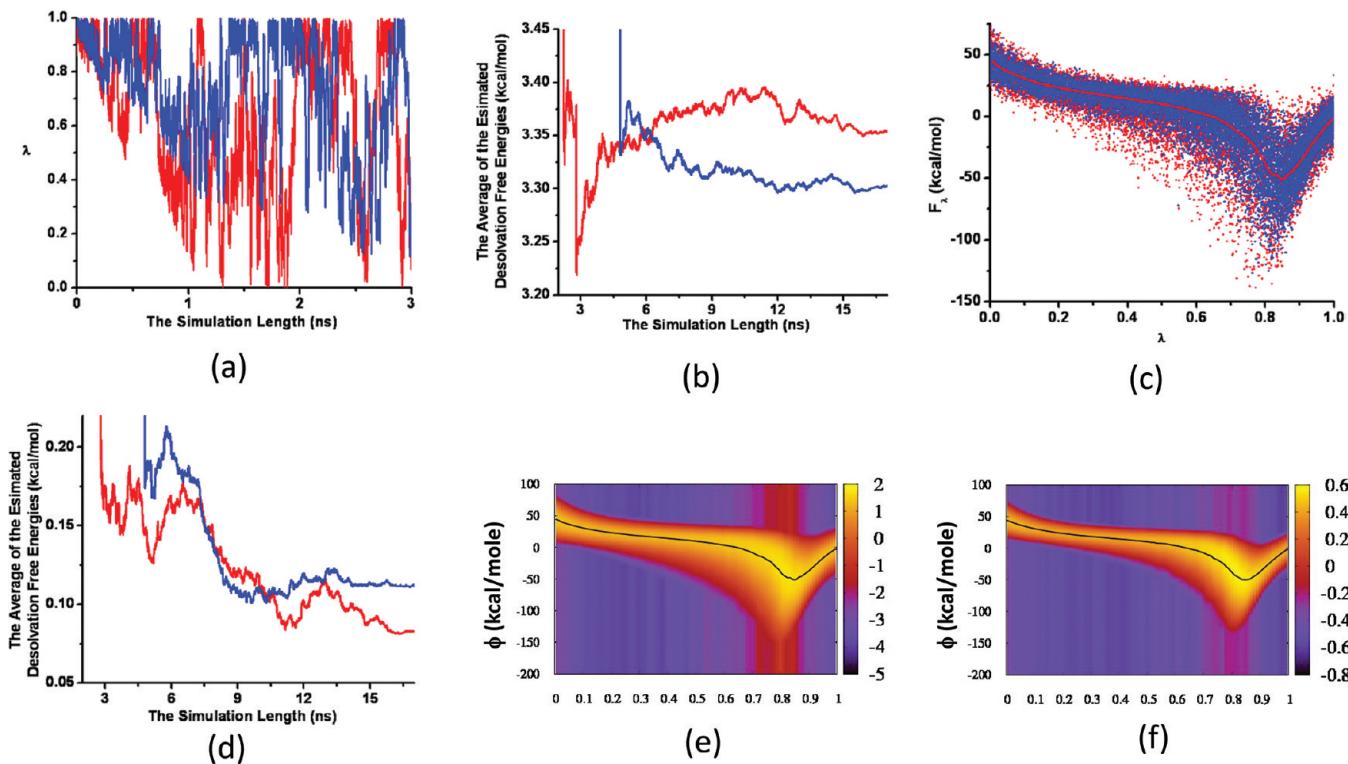


Figure 8. The results of the two sets of DI-OST calculations on the octanol desolvation. (a) The time-dependent λ changes in one of the $T_{ES} = 375$ K DI-OST simulations and one of the $T_{ES} = 750$ K DI-OST simulations. (b) The time-dependent free energy averages from the eight independent $T_{ES} = 750$ K DI-OST simulations (in red) and the eight independent $T_{ES} = 375$ K DI-OST simulations (in blue). (c) The representation of the samples in the (λ, F_λ) space. The samples from a representative $T_{ES} = 750$ K DI-OST simulation are colored in red, The samples from a representative $T_{ES} = 375$ K DI-OST simulation are colored in blue, and the red line is the free energy derivative curve estimated based on the $T_{ES} = 750$ K DI-OST samples. (d) The time-dependent variances among the free energy values estimated from the eight independent $T_{ES} = 750$ K DI-OST simulations (in red) and the eight independent $T_{ES} = 375$ K DI-OST simulations (in blue). (e) The $g_m(\lambda, \phi)$ function generated from one of the $T_{ES} = 750$ K DI-OST simulations. (f) The $g_m(\lambda, \phi)$ function generated from one of the $T_{ES} = 375$ K DI-OST simulations.

Second, the original OSRW method is based on the metadynamics recursion kernel. The metadynamics kernel provides extra dynamic boosts on λ moves. Then, the first one-way trips can be quickly completed (around 350 ps in average). Although the free energy estimations could be started earlier, both of the short-time and long-time convergence behaviors of the OSRW simulations are much worse than those of the DI-OST simulations (Figure 7b and c). For instance, at 2 ns, the average of the free energy values from the OSRW simulations converges to 299.97 kcal/mol, and the variance of these results is about 0.10 kcal/mol. The metadynamics sampling in the OSRW simulations is by nature in the nonequilibrium regime; in comparison, the sampling in the DI-OST simulations starts in the near-equilibrium regime and rigorously approaches the equilibrium regime with the converging of the two recursion target functions. Certainly, the robustness and the convergence behavior of OSRW simulations can be improved with the decreasing of the employed Gaussian height; however, it is expected that then the orthogonal space recursion (the recursion kernel) efficiency will be lower and the $g_m(\lambda, F_\lambda)$ convergence will be slower.

Obviously, the DI-OST algorithm allows the orthogonal space sampling strategy to be more robustly realized for free energy simulations. It should be noted that although in the above comparison, better robustness and long-time convergence behavior of the DI-OST simulations have been demonstrated; indeed, within the simulated time scale, the absolute performance of the OSRW simulations is still quite encouraging. As shown in Figure 7a, even the average of the initially estimated (after

the first one-way trips) results from OSRW can be as good as 300.1 kcal/mol, and the final estimated results converge to the same values as the ones predicted on the basis of the DI-OST method.

IV.B. The Solvation Free Energy of the Octanol Molecule: the Orthogonal Space Sampling Temperature T_{ES} .

Among various alchemical free energy simulation applications, solvation free energy calculations are unique because of the fact that they may require extensive sampling but the results are still quantitatively verifiable by classical free energy simulations. In this study, we carried out solvation energy calculations on the octanol molecule to understand the role of the orthogonal space sampling temperature T_{ES} in the DI-OST method.

Figure 8a shows the time-dependent λ changes in the first 3 ns of two representative DI-OST simulations, in which T_{ES} is respectively set as 750 K (the blue curve) and 375 K (the red curve). As discussed earlier, the sampling length required to achieve the first one-way trip is a key factor in sampling efficiency measurement. The average of the first one-way trip sampling lengths in the eight $T_{ES} = 750$ K DI-OST simulations is 1.6 ns; the variance of these sampling lengths is 0.53 ns. In comparison, the average of the first one-way trip sampling lengths in the eight $T_{ES} = 375$ K DI-OST simulations is 3.57 ns, and the variance of the first one-way trip sampling lengths is 0.63 ns. As shown in Figure 8a, the sampling bottleneck is located in the region of $\lambda \in (0.7, 0.8)$; infrequent crossing of this region slows down overall λ round-trip diffusivity.

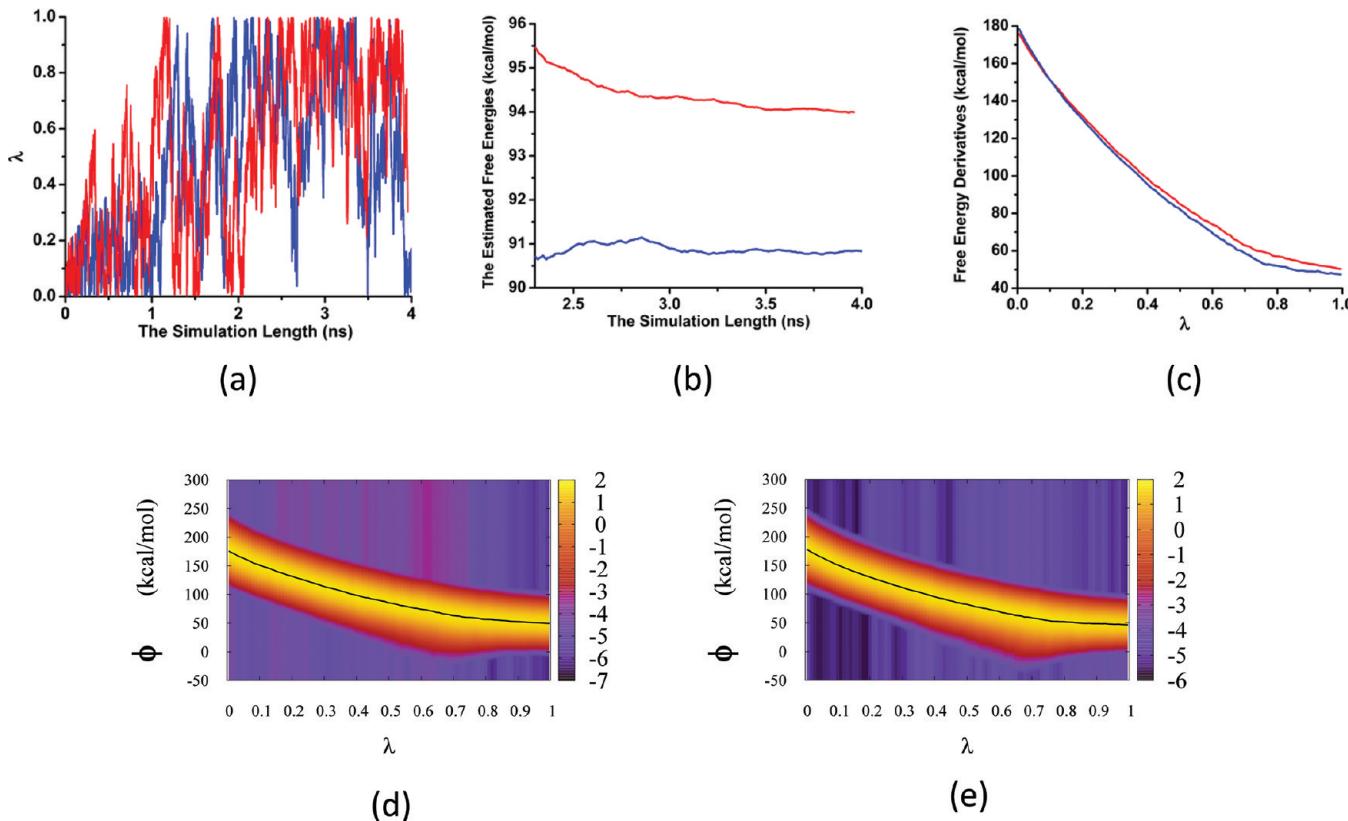


Figure 9. The results on the Barnase–Barstar binding affinity change induced by the Barnase N58A mutation. (a) The time-dependent λ changes in the bound state (in red) and unbound state (in blue) DI-OST simulations. (b) The time-dependent estimated free energy values from the bound state (in red) and unbound state (in blue) DI-OST simulations. (c) The free energy derivative curves from values from the bound state (in red) and unbound state (in blue) DI-OST simulations. (d) The $g_m(\lambda, \phi)$ function generated from the bound state DI-OST simulation. (e) The $g_m(\lambda, \phi)$ function generated from the unbound state DI-OST simulation.

As shown in Figure 8c, the solute appearance/annihilation transition is the major event in this sampling bottleneck region. It is noted that due to the employment of the soft-core potential, the solute appearance/annihilation transition is shifted from $\lambda = 1$, the expected region when the linear hybrid alchemical potential is applied, to this new region. As is generally known, solvent molecule reorganizations are the “hidden” events that are associated with solute insertions/annihilations. When the orthogonal space sampling temperature T_{ES} is higher (for instance 750 K), the magnitude of the F_λ fluctuation is expected to be larger (as shown in Figure 8c) and hidden free energy barriers associated with solvent reorganizations can be more quickly crossed; thereby, the sampling of the bottleneck region can be more efficient.

The red curve in Figure 8b shows the time-dependent averages of the estimated desolvation free energies from the eight $T_{ES} = 750$ K DI-OST simulations, and the red curve in Figure 8d shows the time-dependent variances of the estimated desolvation free energies from the eight $T_{ES} = 750$ K DI-OST simulations. At around 2 ns, the average of the estimated values is 3.45 kcal/mol and the variance of these values is about 0.23 kcal/mol (Figure 8b and d). At around 6 ns, the average of the estimated values drops to around 3.35 kcal/mol, while their variance decreases to 0.17 kcal/mol. At around 13.5 ns, the free energy estimations reach very nice convergence with the average value of 3.36 kcal/mol, and the estimation variance drops below 0.1 kcal/mol. By the inclusion of the long-range Lennard-Jones correction⁶⁶ (0.79 kcal/mol), the predicted solvation energy, -4.15 ± 0.1 kcal/mol, is in excellent

agreement with the experimental value -4.09 kcal/mol.⁶⁷ At 17 ns, a nicely converged $g_m(\lambda, \phi)$ function was obtained with the variance further reduced to 0.08 kcal/mol; as shown in Figure 8e, the orthogonal space sampling temperature 750 K allows the fluctuations of ϕ and F_λ to quickly escape ~ 5 kT strongly coupled free energy barriers. In comparison, the eight $T_{ES} = 350$ K DI-OST simulations have smaller sampling coverage in the orthogonal space (Figure 8c). The lack of sampling in the orthogonal space not only leads to the longer sampling length requirement for the first one-way trips as discussed above but also leads to the slower convergence (Figure 8b and d). Figure 8d shows that at 17 ns, some of the $T_{ES} = 350$ K DI-OST simulations have not yet converged well because of the fact that the variance among them is still larger than 0.1 kcal/mol. As a result, the average of these values is about 0.05 kcal/mol away from the average of the values estimated from the $T_{ES} = 750$ K simulations (Figure 8b). Figure 8f illustrates the $g_m(\lambda, \phi)$ function obtained at 17 ns of a representative $T_{ES} = 350$ K DI-OST simulation; with $T_{ES} = 350$ K, the orthogonal space sampling treatment temperature 350 K only allows the fluctuations of ϕ and F_λ to escape less than $2 kT$ strongly coupled hidden free energy barriers.

As shown in the above analysis, the orthogonal space tempering treatment allows the sampling bottleneck regions, where hidden free energy barriers exist, to be more efficiently explored. If there is no hidden free energy barrier in the orthogonal space, a higher orthogonal space sampling temperature T_{ES} may introduce more diffusion sampling overhead, which might lower free energy estimation precision. To further

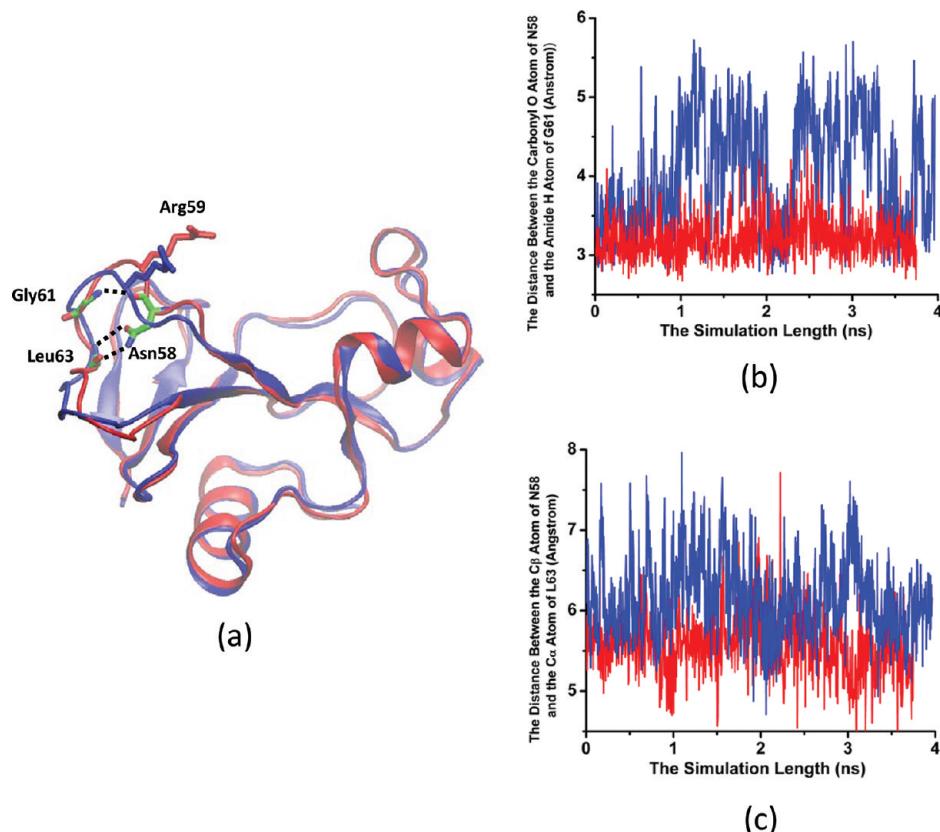


Figure 10. The structural changes associated with the N59A mutation. (a) The comparison of the unbound structures of the wild and the mutant Barnase proteins. The wild type structure is colored in red; for clarity, Asn58 and the backbones of Gly61 and Leu63 are colored by atom. The mutant structure (obtained from the unbound DI-OST simulation at 3 ns when $\lambda = 1$) is colored in blue. (b) The time-dependent changes of the distance between the carbonyl O atom of residue 58 and the amide H atom of residue 61 of Barnase. (d) The time-dependent changes of the distance between the C_β atom of residue 58 and the C_α atom of residue 63 of Barnase.

improve the DI-OST method, currently, we are actively working on a strategy to optimize $T_{ES}(\lambda)$ as the function of λ . In practical biomolecular simulation studies, there usually exist large hidden free energy barriers, and then, obtaining accurate free energy estimation should be a higher priority than improving estimation precision, as long as the estimation precision is in a reasonable range. On the basis of our experience, when a new system is explored, we would like to recommend setting T_{ES} in a range between 750 and 1500 K.

IV.C. The Barnase–Barstar Binding Affinity Change Induced by the Barnase N58A Mutation. It has been known that charge–charge interactions are directly responsible for the strong binding between Barnase and Barstar.⁶⁸ The Barnase Asn58 residue is located at the second layer of the binding interface (Figure 5b). As measured experimentally,⁶⁰ the noncharging N58A mutation causes 3.1 kcal/mol of the binding affinity loss. This unusual electrostatic response suggests that nontrivial conformational changes are likely to be coupled with the N58A mutation. To quantitatively understand the N59A induced binding affinity change, a specialized technique like the DI-OST method should be applied to ensure adequate sampling of the coupled structural transitions. To confidently sample such conformational changes, in the DI-OST simulations, T_{ES} is set as 1500 K.

Two DI-OST simulations, which are respectively based on the Barnase–Barstar (bound) complex structure and the Barnase (unbound) structure, were performed. In 4 ns, multiple λ round-trips were realized in both of the DI-OST simulations.

As shown in Figure 9a, it took the bound-state simulation only 1.1 ns to complete the first one-way trip, while it took the unbound-state sampling about 1.8 ns to cover the entire order parameter range. The dynamics of the scaling parameter λ in the unbound-state simulation reveals that the region of $\lambda = 0.4$ is the sampling bottleneck area, where slow gating events need to occur for λ continuing travels. In 4 ns, decent convergence was realized in both of the free energy simulations (Figure 9b). Through the DI-OST recursion treatment, the λ -dependent free energy derivatives $dG_o/d\lambda$ were calculated; Figure 9c reveals that the binding affinity change $\Delta\Delta G_{Asn \rightarrow Ala}$ is largely responsible for the difference that occurs near the alanine state ($\lambda = 1$), where the two free energy derivative curves are distinct from each other. As discussed later, the conformational change of the mutated (N58A) Barnase induced by the binding/unbinding of Barstar is mainly responsible for $\Delta\Delta G_{Asn \rightarrow Ala}$. On the basis of the TI formula (eq 9), $\Delta G_{Asn \rightarrow Ala}^{complex}$ is estimated to be 94.0 kcal/mol and $\Delta G_{Asn \rightarrow Ala}^{Barnase}$ is estimated to be 91.1 kcal/mol; thus $\Delta\Delta G_{Asn \rightarrow Ala}$ can be predicted to be 2.9 kcal/mol, which is in excellent agreement with the experimental value of 3.1 kcal/mol. Figure 9d and e illustrate the $g_m(\lambda, \phi)$ functions that are generated respectively from the bound state and the unbound state simulations; as shown, the orthogonal space tempering treatment allows the fluctuations of ϕ and F_λ to overcome $\sim 12\text{--}14 kT$ of the strongly coupled hidden free energy barriers.

The comparison of the crystal structures (1BRS and 1BNR) suggests that the Barnase protein has the identical conformation at the bound and the unbound states. As shown in

Figure 10 (the part with the protein backbone colored in red and the side chains colored by atom), the Barnase Asn58 is located on a Barstar-binding loop, but at the opposite side from the binding interface residues, for instance, Arg59. In these structures, the binding interface region on the Arg59-containing loop is zipped by the hydrogen bond between the amide group of Gly61 and the carbonyl group of Asn58; thereby Arg59 can be accurately positioned into the binding site. This zipped structure is further locked by two additional hydrogen bonds between the Asn58 side chain and the backbone amide/carbonyl groups. In the bound-state DI-OST simulation, with residue 58 repeatedly interconverted between the two end chemical states: asparagine and alanine, the structure of the Arg59-containing loop stayed unchanged, even when λ approached the alanine state ($\lambda = 1$). As shown by the red line in Figure 10b, the hydrogen bond between the amide group of Gly61 and the carbonyl group of Asn58 was not broken during the entire simulation. The fluctuation of the distance between residues 58 and 63 was modest (the red line in Figure 10c). In contrast, in the unbound-state simulation, synchronously with the λ move, the Arg59-containing loop varied back and forth between the original zipped conformation (at the asparagine state when $\lambda = 0$) and a newly formed unzipped conformation (at the alanine state when $\lambda = 1$). As shown by the blue line in Figure 10c, when residue 58 turned to alanine, the distance between residues 58 and 63 increased, and when λ traveled back to the asparagine state, the canonical hydrogen bonds between these two residues were formed again. Correspondingly, the zipping hydrogen bond repetitively broke and reformed; these concerted changes can be revealed by the blue line in Figure 10b. Figure 10a shows the comparison between the unbound structures of the wild type (in red) and the mutant (in blue) proteins. On the unzipped loop of the unbound NS9A mutant, Arg59 flips away from its wild-type gesture that is originally preorganized to bind Barstar.

The above analysis suggests that there is strong coupling between the Barnase–Barstar binding and the Arg59-containing loop zipping, and Asn58 plays a pivotal role in stabilizing the zipped conformation of the Arg59-containing loop when Barnase is in the unbound state. Therefore, the Barnase–Barstar binding can be enhanced. When Asn58 is mutated to alanine, the Arg59-containing loop in the unbound Barnase is unzipped due to the loss of both the locking hydrogen bonds by Asn58 and the binding of the Barstar. When the NS8A mutant binds Barstar, some free energy penalty needs to be paid in order to form the bound conformation, which, as revealed by the bound state DI-OST simulation, stays zipped in the Barstar-bound state regardless of the existence of Asn58. As shown in Figure 9c, the two simulations share the similar free energy derivative curves near the asparagine ($\lambda = 0$) state; this indicates that there is only modest contribution from the direct electrostatic interaction difference to the binding affinity change. In essence, the binding affinity change induced by the NS8A mutation is largely responsible for the mutation-induced conformational change at the unbound state. The DI-OST method allows the corresponding conformational change to be synchronously sampled with the λ moves; therefore, the binding affinity change can be efficiently predicted.

V. CONCLUDING REMARKS

As is generally known, achieving adequate sampling for free energy calculations can be extremely challenging. The bottle-

neck problem lies in the fact that order parameter transitions might be gated by slow structural changes. The orthogonal space random walk (OSRW) method, which enables synchronous acceleration of the motions of a focused region and its strongly coupled environment, was recently introduced to enhance sampling for free energy simulations. Although the OSRW method has shown very encouraging sampling power, the originally implemented method suffers from a lack of robustness, especially in the aspect of the long-time scale convergence. To resolve the robustness issue, in the present work, the OSRW algorithm is generalized to be the orthogonal space tempering (OST) technique via the introduction of the orthogonal space sampling temperature. Here, the orthogonal space sampling temperature can be employed to control orthogonal space sampling boundaries and diffusion sampling overheads. While the OST method can also be implemented with the well-tempered metadynamics method⁶⁹ as the recursion kernel technique, to enable practically robust OST free energy updates, a double-integration recursion strategy is developed; specifically, a “dynamic reference restraining” method is employed as the recursion kernel. Furthermore, the algorithm is augmented by a novel θ -dynamics approach to realize both uniform sampling of the order parameter space and rigorous end point constraints.

In the present work, the double-integration OST method is employed to perform alchemical free energy simulations, specifically to calculate the free energy difference between benzyl phosphonate and difluorobenzyl phosphonate in aqueous solution, to estimate the solvation free energy of the octanol molecule, and to predict the nontrivial Barnase–Barstar binding affinity change induced by the Barnase N58A mutation. As demonstrated in these model studies, the DI-OST method can robustly enable practically efficient free energy predictions.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: yyang2@fsu.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The funding support (MCB 0919983 to W.Y.) from National Science Foundation is acknowledged. We would like to thank the Florida State University high performance computing (HPC) center, the NSF Teragrid facility, and the institute of molecular biophysics computing facility (Dr. Michael Zawrotny) for the computing support. We would like to thank Dr. Stefan Boresch for providing the benzyl phosphonate to difluorobenzyl phosphonate alchemical free energy simulation setup.

■ REFERENCES

- (1) Jorgensen, W. L. Free-energy calculations – A breakthrough for modeling organic-chemistry in solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.
- (2) Beveridge, D. L.; Dicapua, F. M. Free-energy via molecular simulation – Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492.
- (3) Kollman, P. Free-energy calculations – Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (4) Adcock, S. A.; McCammon, J. A. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem. Rev.* **2006**, *106*, 1589–1615.

- (5) Chipot, C.; Shell, M. S.; Pohorille, A. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Pohorille, A., Chipot, C., Eds.; Springer: Heidelberg, Germany, 2007; pp 1–31.
- (6) Yang, W.; Cui, Q.; Min, D.; Li, H. Z. QM/MM alchemical free energy simulations: Challenges and recent developments. *Ann. Rep. Comput. Chem.* **2010**, *6*, 51–62.
- (7) Okamoto, Y. Generalized-ensemble algorithms: Enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J. Mol. Graphics Modell.* **2004**, *22*, 425–439.
- (8) Yang, W.; Nymeyer, H.; Zhou, H. X.; Berg, B. A.; Brüschweiler, R. Quantitative computer simulations of biomolecules: A snapshot. *J. Comput. Chem.* **2008**, *29*, 668–672.
- (9) Warshel, A. Simulating the energetics and dynamics of enzymatic reactions. *Pont. Acad. Sci. Scr. Var.* **1983**, *55*, 59–81.
- (10) Tembe, B. L.; McCammon, J. A. Ligand receptor interactions. *Comp. Chem.* **1984**, *8*, 281–283.
- (11) Jorgensen, W. L.; Ravimohan, C. Monte-Carlo simulation of differences in free-energies of hydration. *J. Am. Chem. Soc.* **1985**, *83*, 3050–3054.
- (12) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. Free-energy calculations by computer-simulation. *Science* **1987**, *236*, 564–568.
- (13) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. Hidden thermodynamics of mutant proteins – A molecular-dynamics analysis. *Science* **1989**, *244*, 1069–1072.
- (14) Straatsman, T. P.; McCammon, J. A. Computational alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.
- (15) Kollman, P. Free-energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (16) Simonson, T.; Archontis, G.; Karplus, M. Free energy simulations come of age: Protein-ligand recognition. *Acc. Chem. Res.* **2002**, *35*, 430–437.
- (17) Gilson, M. K.; Zhou, H. X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.
- (18) Jorgensen, W. L.; Thomas, L. L. Perspective on free-energy perturbation calculations for chemical equilibria. *J. Chem. Theory Comput.* **2008**, *4*, 869–876.
- (19) Bartels, C.; Karplus, M. Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide. *J. Comput. Chem.* **1997**, *18*, 1450–1462.
- (20) Darve, E.; Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.* **2001**, *115*, 9169–9183.
- (21) Bitetti-Putzer, R.; Yang, W.; Karplus, M. Generalized ensembles serve to improve the convergence of free energy simulations. *Chem. Phys. Lett.* **2003**, *377*, 633–641.
- (22) Fasnacht, M.; Swendsen, R. H.; Rosenberg, J. M. Adaptive integration method for Monte Carlo simulations. *Phys. Rev. E* **2004**, *69*, 056704.
- (23) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.
- (24) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.* **1989**, *156*, 472–477.
- (25) Wang, F. G.; 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) Marsili, S.; Barducci, A.; Chelli, R.; Procacci, P.; Schettino, V. Self-healing umbrella sampling: A non-equilibrium approach for quantitative free energy calculations. *J. Phys. Chem. B* **2006**, *110*, 14011–14013.
- (28) Min, D. H.; Liu, Y. S.; Carbone, I.; Yang, W. On the convergence improvement in the metadynamics simulations: A Wang-Landau recursion approach. *J. Chem. Phys.* **2007**, *126*, 194104.
- (29) Zheng, L. Q.; Yang, W. Essential energy space random walks to accelerate molecular dynamics simulations: Convergence improvement via an adaptive-length self-healing strategy. *J. Chem. Phys.* **2008**, *129*, 014105.
- (30) Zheng, L. Q.; Carbone, I. O.; Lugovskoy, A.; Berg, B. A.; Yang, W. A hybrid recursion method to robustly ensure convergence efficiencies in the simulated scaling based free energy simulations. *J. Chem. Phys.* **2008**, *129*, 034105.
- (31) Babin, V.; Roland, C.; Sagui, C. Adaptively biased molecular dynamics for free energy calculations. *J. Chem. Phys.* **2008**, *128*, 134101.
- (32) Zheng, H.; Zhang, Y. K. Determination of free energy profiles by repository based adaptive umbrella sampling: Bridging non-equilibrium and quasiequilibrium simulations. *J. Chem. Phys.* **2008**, *128*, 204106.
- (33) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
- (34) Dickson, B. M.; Legoll, F.; Lelievre, T.; Stoltz, G.; Fleurat-Lessard, P. Free energy calculations: An efficient biasing potential method. *J. Phys. Chem. B* **2010**, *114*, 5823–5830.
- (35) Zhang, Y.; Voth, G. A. Combined metadynamics and umbrella sampling method for the calculation of ion permeation free energy profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2277–2283.
- (36) Zheng, L. Q.; Chen, M. G.; Yang, W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20227–20232.
- (37) Zheng, L. Q.; Chen, M. G.; Yang, W. Simultaneous escaping of explicit and hidden free energy barriers: Application of the orthogonal space random walk strategy in generalized ensemble based conformational sampling. *J. Chem. Phys.* **2009**, *130*, 234105.
- (38) Marcus, R. A. On the theory of oxidation-reduction reactions involving electron transfer. *J. Chem. Phys.* **1956**, *24*, 966–978.
- (39) King, G.; Warshel, A. Investigation of the free-energy functions for electron-transfer reactions. *J. Chem. Phys.* **1990**, *93*, 8682–8692.
- (40) Min, D. H.; Zheng, L. Q.; Harris, W.; Chen, M. G.; Lv, C.; Yang, W. Practically efficient QM/MM alchemical free energy simulations: The orthogonal space random walk strategy. *J. Chem. Theory Comput.* **2010**, *6*, 2253–2266.
- (41) Lelievre, T.; Rousset, M.; Stoltz, G. Long-time convergence of an adaptive biasing force method. *Nonlinearity* **2008**, *21*, 1155–1181.
- (42) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted scaling: A new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994**, *100*, 9025–9031.
- (43) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.
- (44) Tidor, B. Simulated annealing on free-energy surfaces by a combined molecular-dynamics and Monte-Carlo approach. *J. Phys. Chem.* **1993**, *97*, 1069–1073.
- (45) Pitera, J.; Kollman, P. Designing an optimum guest for a host using multimolecule free energy calculations: Predicting the best ligand for Rebek's "tennis ball". *J. Am. Chem. Soc.* **1998**, *120*, 7557–7567.
- (46) Li, H.; Fajer, M.; Yang, W. Simulated scaling method for efficient localized conformational sampling and simultaneous "alchemical" free energy simulation: A general method for MM, QM, and QM/MM simulations. *J. Chem. Phys.* **2007**, *126*, 024106.
- (47) Kong, X. J.; Brooks, C. L. Lambda-dynamics: A new approach to free energy calculations. *J. Chem. Phys.* **1996**, *105*, 2414–2423.
- (48) Knight, J. L.; Brooks, C. L. Lambda-dynamics free energy simulation methods. *J. Comput. Chem.* **2009**, *30*, 1692–1700.
- (49) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. Constant-pH molecular dynamics using continuous titration coordinates. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 738–752.
- (50) Knight, J. L.; Brooks, C. L. Applying efficient implicit nongeometric constraints in alchemical free energy simulations. *J. Comput. Chem.* **2001**, *32*, 141–152.
- (51) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. Assessing the accuracy of metadynamics. *J. Phys. Chem. B* **2005**, *109*, 6714–6721.

- (52) Darve, E.; Rodriguez-Gomez, D.; Pohorille, A. Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* **2008**, *14*, 144120.
- (53) Kastner, J.; Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration". *J. Chem. Phys.* **2005**, *123*, 144104.
- (54) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (55) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Calfisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Feig; 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. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (56) Boresch, S.; Leitgeb, M.; Beselman, A.; MacKerell, A. D. Unexpected relative aqueous solubilities of a phosphotyrosine analogue and two phosphonate derivatives. *J. Am. Chem. Soc.* **2005**, *127*, 4640–4648.
- (57) Bruckner, S.; Boresch, S. Efficiency of alchemical free energy simulations. I. A practical comparison of the exponential formula, thermodynamic integration, and Bennett's acceptance ratio method. *J. Comput. Chem.* **2011**, *32*, 1303–1319.
- (58) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (59) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D. Jr. CHARMM general force field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (60) Schreiber, G.; Fersht, A. R. Interaction of barnase with its polypeptide inhibitor barstar studies by protein engineering. *Biochemistry* **1993**, *32*, 5145–5150.
- (61) MacKerell, A. D. Jr; Feig, M.; Brooks, C. L. III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- (62) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.
- (63) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An nlog(n) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (64) Nôse, S. A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (65) Nôse, S.; Klein, M. L. Constant pressure molecular-dynamics for molecular-systems. *Mol. Phys.* **1983**, *50*, 1055–1076.
- (66) Wescott, J. T.; Fisher, L. R.; Hanna, S. Use of thermodynamic integration to calculate the hydration free energies of n-alkanes. *J. Chem. Phys.* **2002**, *116*, 2361.
- (67) Li, J. B.; Zhu, T. H.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. Extension of the platform of applicability of the SMS.42R universal solvation model. *Theor. Chem. Acc.* **1999**, *103*, 9–63.
- (68) Lee, L. P.; Tidor, B. Barstar is electrostatically optimized for tight binding to barnase. *Nat. Struct. Biol.* **2001**, *8*, 73–76.
- (69) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.