Analysis of Transition Rates from Variational Flooding Using Analytical Theory

David Cummins, Carter Longstreth, and James McCarty

Department of Chemistry, Bellingham WA, 98225, USA

(*mccartj8@wwu.edu.)

(Dated: October 29, 2024)

Abstract

Variational flooding is an enhanced sampling method for obtaining kinetic rates from molecular dynamics simulations. The method is inspired by the idea of conformational flooding that employs a boost potential acting along a chosen reaction coordinate to accelerate rare events. In this work, we show how the empirical distribution of crossing times from variational flooding simulations can be modeled with analytical Kramers' time-dependent rate (KTR) theory. An optimized bias potential that fills metastable free energy basins is constructed from the variationally enhanced sampling (VES) method. This VES-derived flooding potential is then augmented by a switching function that determines the fill level of the boost. Having a prescribed time-dependent fill rate of the flooding potential gives an analytical expression for the distribution of crossing times from KTR theory that is used to extract unbiased rates. In the case of a static boost potential, the distribution of barrier crossing times follows an expected exponential distribution, and unbiased rates are extracted from a series of boosted simulations at discrete fill levels. Introducing a time-dependent boost that increases the fill level gradually over the simulation time leads to a simplified procedure for fitting the biased distribution of crossing times to analytical theory. We demonstrate the approach for the paradigmatic cases of alanine dipeptide in vacuum, the asymmetric S_N2 reaction, and the folding of chignolin in explicit solvent.

I. INTRODUCTION

Enhanced sampling methods are frequently used in atomistic simulations to accelerate conformational dynamics and to facilitate transitions between metastable states separated by large kinetic barriers[1]. Examples include umbrella sampling[2], metadynamics[3], steered molecular dynamics[4, 5], adaptive biasing force[6, 7], replica exchange[8], accelerated MD[9], variationally enhanced sampling[10], on-the-fly probability enhanced sampling[11], and many others. A large class of these methods, including metadynamics and its variants, use an external bias potential that acts along a chosen set of slow degrees of freedom called collective variables (CVs). The bias amplifies fluctuations of the CVs, enabling the system to overcome free energy barriers[12]. Given sufficient sampling of the configuration space, the free energy surface is reconstructed from the biased trajectory using a reweighting procedure[13], and thermodynamic properties are calculated as ensemble averages over the reweighted trajectory.

In addition to thermodynamic observables, one is often interested in calculating dynamic properties from MD simulations, including kinetic rate constants. Several methods have been introduced to obtain kinetic properties at a reduced computational cost relative to conventional MD. For example, algorithms that do not require a bias potential include the Weighted Ensemble (WE) method[14, 15], transition path sampling[16], Markov state models[17], or a combination of these approaches[18]. Another class of simulation approaches employs a bias or boost potential. These include hyperdynamics[19], conformational flooding[20], Gaussian accelerated molecular dynamics (GaMD)[21], infrequent metadynamics[22], variational flooding[23], and OPES flooding[24]. These approaches are similar in spirit to conventional enhanced sampling methods in that an external bias is employed; however, the aim is not to obtain a fully converged free energy surface but to push the system out of a metastable state while preserving the ensemble of transition states and maintaining rare-event kinetics. Tiwary and Parrinello introduced a way to obtain rare event kinetics from metadynamics by using transition state theory to compute an acceleration factor due to the accumulated bias potential[22]. In this approach, metadynamics is employed to fill the metastable basin with a suitable boost potential that will accelerate barrier-crossing events. A Gaussian bias kernel is deposited slowly so as to preserve rare-event statistics and a separation of time scales between the time spent in the free energy basin and the time spent at the transition state. Under the condition of a slow enough bias deposition, so-called infrequent metadynamics (iMetaD), the recovered unbiased escape times follow the statistics of a Poisson process[25]. iMetaD has been

successfully applied to calculate the residence time of small molecules for pharmacologically important protein drug targets[26, 27]. When combined with deep learning methods to learn approximate reaction coordinates, iMetaD can provide quantitatively accurate drug residence times[28].

One restriction of the iMetaD approach is that the bias must not be placed on the dividing surface, requiring a good estimate of the true reaction coordinate and a sufficiently slow hill deposition rate. This condition may be challenging to achieve in practice. A method inspired by single-molecule force spectroscopy was introduced to directly recover kinetic information from iMetaD trajectories by applying Kramers' time-dependent rate (KTR) theory[29]. Instead of rescaling the simulation time, the observed empirical cumulative distribution function (ECDF) of biased crossing times is fit to a model based on KTR theory. In the fitting procedure, the quality of the CV is explicitly taken into account through an additional parameter. This enables accurate unbiased transition times to be estimated directly from the biased ECDF even for cases where the CV is less than optimal or for cases that fail the assumption of Poisson statistics for the recovered escape times. For example, KTR theory was recently used to obtain product release rates for a chemotherapeutic prodrug/enzyme system from iMetaD simulations[30].

In iMetaD the maximum fill-level of the bias continues to increase on average due to the accumulating sum of Gaussian kernels deposited in the local free energy basin. An alternative approach to rare event kinetics, the variational flooding[23] procedure, employs a boost potential that is constructed using the variationally enhanced sampling (VES) formalism to sample a prescribed target probability distribution[10]. In variational flooding, a target probability distribution is chosen such that the system uniformly samples low-energy configurations while an energy cutoff ensures the fill-level of the boost potential remains below the free energy barrier ($\Delta F^{\dagger}$). A similarly inspired approach that uses a prescribed target distribution to enforce a cutoff in the accumulated bias is the recent OPES flooding[24]. The advantage of having a variational formalism is that one can construct a boost potential that fills the free energy basin in a controlled way. In this work, we demonstrate how the flexibility afforded by variational flooding allows one to design a boost potential that is amenable to analytical KTR theory. Having an analytical expression for the distribution of crossing times allows us to directly fit the biased ECDF to obtain kinetic rates, simplifying the analysis of kinetics from biased MD simulations.

We investigate several cases where the variational flooding potential is introduced in such a way as to give an analytical expression for the survival probability from KTR theory. In the first case, the bias is variationally optimized to fill the free energy basin up to a fixed prescribed maximum

value before being used as a time-independent boost potential. This is similar to the approach used in the original variational flooding paper[23], where the bias is first optimized with respect to a target distribution that enforces a cutoff in free energy space. The boost potential effectively "floods" the free energy surface as shown schematically in FIG. 1A. This boost potential is then used to accelerate barrier crossing events, while maintaining the statistics of an activated process. Because the bias is time-independent, the maximum fill-level, V_{MB}, effectively lowers the free energy barrier between the reactant and the transition state by a constant.

In the second case, by introducing a time-dependent switching function, we realize a boost potential that increases in a controlled manner during the simulation time. We demonstrate the case where the boost potential grows at a constant rate, leading to a maximum fill-level of the bias that increases linearly during the simulation, shown schematically in FIG. 1B. The idea is similar in spirit to single-molecule pulling experiments with a constant pulling rate, but in this case we are not pulling along the reaction coordinate but increasing the fill-level of the boost potential. This procedure leads to an analytical expression for the cumulative distribution of barrier crossing times that can be used to extract the unbiased transition rate. This procedure simplifies the extraction of rates from biased simulations and improves the computational efficiency as compared to iMetaD. Finally, in a related example, to make the connection to iMetaD, we show how a time-dependent switching function can be introduced that leads to a nonlinear increase of the boost potential such that the fill level grows logarithmically over the simulation time. The paper is organized as follows: In Section II we describe the theoretical background. In Section III we demonstrate the method on a few illustrative cases. A discussion and summary of the approach concludes the paper.

II. THEORY

In our procedure, a preliminary MD simulation is used to construct a suitable boost potential on-the-fly to fill a local metastable basin using VES in an approach termed variational flooding[23, 31]. In a subsequent step, multiple, independent MD simulations are run under the influence of the boost potential to accelerate transitions out of the metastable basin. The boost potential is augmented by a switching function that controls the fill-level. From these trajectories, Kramers' theory gives a convenient method to analyze the distribution of crossing times and extract unbiased rate constants. We now review the necessary theoretic background for each of these steps.

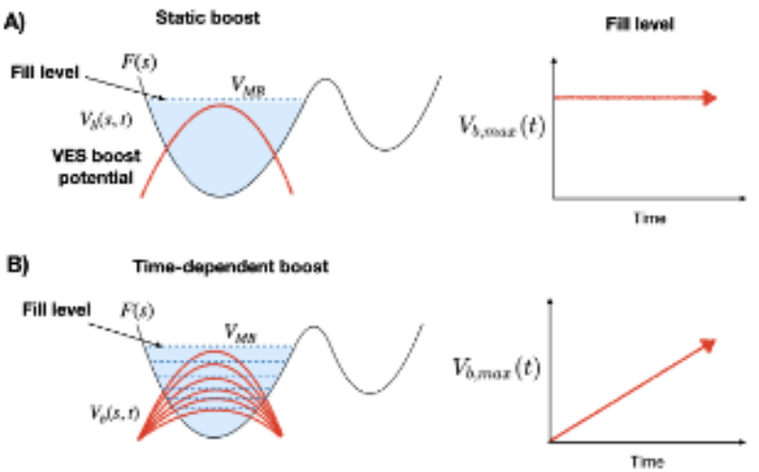

FIG. 1. A) Case 1: The boost potential is constructed on-the-fly using VES with a fixed cutoff that ensures only low-energy configurations are biased. The fill level is independent of the simulation time. The effect of the boost potential is to lower the barrier height, accelerating crossing events. B) Case 2: A boost potential is constructed with a time-dependent cutoff such that the low-energy basin is filled at a prescribed flooding rate. For example, in the case of a constant flooding rate, the maximum boost level increases linearly during the simulation.

A. Bias Construction Using Variational Flooding

Variational flooding is an application of the variationally enhanced sampling (VES) approach of Valsson and Parrinello[10]. As in metadynamics, the bias acts along a set of CVs, $\mathbf{s}(\mathbf{R})$, that are functions of the atomic coordinates, $\mathbf{R}$. The free energy surface along the reaction coordinate is defined up to an additive constant as

$$F(\mathbf{s}) = -\frac{1}{\beta}\log\int d\mathbf{R}\;\delta[\mathbf{s}-\mathbf{s}(\mathbf{R})]\exp(-\beta U(\mathbf{R})) \tag{1}$$

where $\beta = 1/k_B T$ with k_B being Boltzmann's constant and T the temperature, and $U(\mathbf{R})$ is the potential energy. In variational flooding, we seek a boost potential, $V_b(\mathbf{s})$, that fills the free energy basin such that the system will explore a modified free energy surface $\tilde{F}_b(\mathbf{s}) = F(\mathbf{s}) + V_b(\mathbf{s})$, with a lower activation energy. Such a boost potential takes the form:

$$V_b(\mathbf{s}) = -F(\mathbf{s}) - \frac{1}{\beta}\log p^{tg}(\mathbf{s}) \tag{2}$$

where $p^{tg}(\mathbf{s})$ is a target probability distribution chosen so that the boost potential destabilizes the local free energy basin but goes strictly to zero at the transition state region (designated by

regions of high free energy). Under the influence of a boost potential with the form of Equation 2, the simulation will sample the target distribution along the modified free energy surface $p^{tg}(\mathbf{s}) \propto e^{-\beta \tilde{F}_b(\mathbf{s})}$.

In variational flooding, the target distribution is chosen to uniformly fill low-energy regions of the free energy surface up to a prescribed cutoff and is taken to be a normalized sigmoidal function of the form:

$$p^{tg}(\mathbf{s}) = \frac{\sigma(F_\mathbf{s} - V_{MB})}{\int d\mathbf{s}'\ \sigma(F_\mathbf{s} - V_{MB})} \tag{3}$$

with $\sigma(F_\mathbf{s} - V_{MB})$ a Fermi-type switching function:

$$\sigma(F_\mathbf{s} - V_{MB}) = \frac{1}{1 + e^{\lambda(F_\mathbf{s} - V_{MB})}}. \tag{4}$$

The parameter, V_{MB}, determines the maximum height of the boost potential and can be adjusted to accelerate rare events without biasing high energy transition states. The parameter λ determines the softness of the switching function[23].

Having specified the target distribution, a boost potential that satisfies Equation 2 also requires knowledge of $F(\mathbf{s})$ which is not *a priori* known. The appropriate boost is obtained using a variational procedure by minimizing the functional of Valsson and Parrinello[10]

$$\Omega[V_b] = \frac{1}{\beta} \log \frac{\int d\mathbf{s} \exp(-\beta(F(\mathbf{s}) + V_b(\mathbf{s})))}{\int d\mathbf{s} \exp(-\beta(F(\mathbf{s})))} + \int d\mathbf{s}\, V_b(\mathbf{s}) p^{tg}(\mathbf{s}) \tag{5}$$

Minimization of $\Omega[V_b]$ with respect to V_b yields Equation 2. This is accomplished by writing the bias potential as a basis set expansion in terms of a set of basis functions f_k

$$V_b = \left(\sum_k^n c_k \cdot f_k(\mathbf{s}) \right) \cdot \sigma(x) \tag{6}$$

where c_k are a set of coefficients to be optimized according to a variational formalism. $\sigma(x)$ is the switching function that enforces the condition that the max fill-level is V_{MB} and has the form of Equation 4, calculated self-consistently with $x = -\sum_k^n c_k \cdot f_k(\mathbf{s}) - V_{MB}$. The variational coefficients are updated using a stochastic gradient descent algorithm[32, 33] that involves the gradient

$$\frac{\partial \Omega}{\partial c_k} = -\left\langle \frac{\partial V_b}{\partial c_k} \right\rangle_{V_b} + \left\langle \frac{\partial V_b}{\partial c_k} \right\rangle_{p^{tg}} \tag{7}$$

and the Hessian

$$\frac{\partial^2 \Omega}{\partial c_j \partial c_k} = \beta \cdot \mathrm{Cov}\left[\frac{\partial V_b}{\partial c_j}, \frac{\partial V_b}{\partial c_k} \right]_{V_b} - \left\langle \frac{\partial^2 V_b}{\partial c_j \partial c_k} \right\rangle_{V_b} + \left\langle \frac{\partial^2 V_b}{\partial c_j \partial c_k} \right\rangle_{p^{tg}} \tag{8}$$

where the expectation values and covariance are either obtained in a biased simulation employing the boost potential V_b or are computed numerically on a grid over the target distribution, p^{tg}.

B. Conformational Flooding and Hyperdynamics Time Rescaling

In its original implementation[23], the boost potential, V_b, acting on the free energy surface is introduced as a fixed bias after the variational coefficients have converged to a quasi-static boost potential. From transition state theory, the acceleration factor, $\alpha(t)$, due to the boost potential is given by[22]

$$\alpha(t) = \frac{k_b}{k_0} = \frac{Z_0}{Z_b} = \frac{\int_{\lambda<\lambda^*} d\mathbf{R}\, e^{-\beta U(\mathbf{R})}}{\int_{\lambda<\lambda^*} d\mathbf{R}\, e^{-\beta[U(\mathbf{R})+V_b(\mathbf{s})]}}$$
$$= \left\langle e^{\beta V_b(\mathbf{s})} \right\rangle_b \tag{9}$$
$$\tag{10}$$

where k_b is the biased rate measured in a MD simulation with the boost potential, V_b, and k_0 is the unbiased rate that would be measured for the same transition in the absence of the boost potential. The integrals are over all configurations in the reactant basin along the hypersurface, $\lambda(\mathbf{R})$, with λ^* representing the transition state. The brackets $< ... >_b$ indicate an average over a biased trajectory. The physical time, t^*, is recovered by rescaling the simulation time by the acceleration factor:

$$t^* = t \left\langle e^{\beta V_b(\mathbf{s})} \right\rangle_b$$
$$t^* = \Delta t_{MD} \sum_i^{n_{tot}} e^{\beta V_b^{(i)}(\mathbf{s})} \tag{11}$$

where Δt_{MD} is the MD integration time step, and the sum is over the discrete MD steps with $V_b^{(i)}(\mathbf{s})$ the boost potential at the i^{th} step. In this procedure the trajectory time is weighted at each step to obtain a rescaled trajectory from which unbiased rates are estimated from the distribution of crossing events. We will refer to this procedure as the hyperdynamics time rescaling approach, which involves scaling the biased simulation time according to Equation 11.

C. Kramers Time-dependent Rate Theory

Recently, a formulation based on KTR theory was introduced to estimate unbiased rates from iMetaD simulations[29]. Considering the case of diffusive dynamics and a high barrier, the unbiased rate of barrier crossing along a reaction coordinate is

$$k_0 = A e^{-\beta \Delta F^\dagger} \tag{12}$$

where $\Delta F^{\dagger}$ is the barrier height, and A is a pre-exponential factor. It is assumed that in a biased simulation, the time-dependent boost potential reduces the effective barrier height by the maximum level of the bias at time t, $V_{MB}(t)$. In multiple iMetaD simulations, $V_{MB}(t)$ is taken as the average over independent runs and is an estimate of the average fill-level of the metastable basin after some time, t. KTR theory estimates the biased escape rate as

$$k_b(t) = Ae^{-\beta\Delta F^{\dagger}+\beta\gamma V_{MB}(t)} = k_0 e^{\beta\gamma V_{MB}(t)} \tag{13}$$

The parameter, $\gamma \in [0,1]$, is added to account for the fact that the CVs along which the bias acts only approximate the true reaction coordinate. In the case where $\gamma = 0$, the CVs are orthogonal to the reaction coordinate and the bias does not contribute to the rate acceleration. On the other hand, when $\gamma = 1$, the CVs lie ideally along the reaction coordinate, and the bias lowers the effective barrier. From Equation 13, the survival probability is adapted from methods used in single-molecule force spectroscopy[34]

$$S(t) = \exp\left(-k_0 \int_0^t e^{\beta\gamma V_{MB}(t')} dt'\right) \tag{14}$$

where k_0 is the unbiased crossing rate.

D. Case 1: Variational Flooding with a Constant Fill-level

We first examine the case in which the variational flooding procedure is used to obtain a boost potential that floods the free energy surface uniformly up to a specified fixed cutoff value, V_{MB} as illustrated in FIG. 1A. This situation is similar to the approach used in Ref [23] where the boost potential was first optimized using VES in a preliminary simulation and then used as a static boost potential in independent simulations. Since the bias has already reached its maximum allowed height, the value of V_{MB} is time-independent and equal to the prescribed fill-level. In this case, Equation 14 gives the survival probability

$$S(t) = \exp\left(-k_0 t e^{\beta\gamma V_{MB}}\right) \tag{15}$$

which has the form of an exponential process with a cumulative probability distribution of barrier crossing times as

$$\begin{aligned} \mathrm{CDF} &= 1 - \exp\left(-k_0 t e^{\beta\gamma V_{MB}}\right) \\ &= 1 - \exp\left(-t/\tau_b\right) \end{aligned} \tag{16}$$

with the mean biased transition time,

$$\tau_b = \frac{e^{-\beta\gamma V_{MB}}}{k_0} \tag{17}$$

Thus, in the case of a time-independent boost potential, the ECDF from a biased simulation can be fit to that of an exponential distribution, Equation 16, as expected for a Poisson process. This is because a static boost potential lowers the activation barrier but still preserves the statistics of an activated process, in contrast to metadynamics, which introduces a time-dependent bias, leading to biased escape times that do not follow Poisson statistics[25]. If the CV is ideal (meaning $\gamma = 1$), the unbiased rate, k_0, can, in principle, be obtained directly from the mean biased transition time, τ_b, through Equation 17 because the value of the bias cutoff, V_{MB}, is a known parameter in the procedure. Generally, the CV will not be idea and even in the case of an ideal CV ($\gamma = 1$), the direct application of Equation 17 introduces some approximation error. It is instructive to compare Equation 17 with the acceleration factor from the hyperdynamics time rescaling approach given by Equation 10. From the hyperdynamics time rescaling approach, the result from Equation 17 is recovered when the acceleration factor is

$$\left\langle e^{\beta V_b(t)} \right\rangle_b \approx e^{\beta\gamma V_{MB}} \tag{18}$$

The ensemble-averaged acceleration factor on the left-hand side of Equation 18 can be approximated using a cumulant expansion

$$\left\langle e^{\beta V_b(t)} \right\rangle_b = \exp\left\{ \sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k \right\} \tag{19}$$

with the first two cumulants given by

$$\begin{aligned} C_1 &= \langle V_b(t) \rangle_b \\ C_2 &= \left\langle V_b(t)^2 \right\rangle_b - \langle V_b(t) \rangle^2 \end{aligned} \tag{20}$$

We see that the acceleration factor reduces to the result from Kramers' theory by setting the average bias everywhere to $\langle V_b(t) \rangle_b = V_{MB}$ and assuming all higher order cumulants vanish. In practice, however, the error from the Kramers' approach will be subsumed in the unknown parameter γ, when $\gamma \neq 1$.

Introducing γ as a second fit parameter, both unknown parameters, k_0 and γ, are obtained by linearization of Equation 17:

$$\ln \tau_b = -\beta\gamma V_{MB} - \ln k_0 \tag{21}$$

where τ_b is the mean transition time measured in a series of biased simulations at different fixed V_{MB} values.

E. Case 2: Variational Flooding with a Time-dependent Boost Potential

For a time-dependent boost, as in metadynamics, the biased ECDF will not follow a simple exponential distribution. Here, we exploit the flexibility afforded by variational flooding to construct a boost potential acting on the free energy surface that leads to a fill-level that increases in a controlled manner as depicted in FIG. 1B. As before, we perform a preliminary MD simulations using VES to obtain a boost potential, expressed in terms of a set of variational coefficients and basis set functions according to

$$V_b(\mathbf{s}) = \sum_k^n c_k \cdot f_k(\mathbf{s}) \tag{22}$$

The variational coefficients converge quickly because the simulation remains within the reactant basin. **In a second step**, we introduce a time-dependent switching function of the form of Equation 4 but now with a time-dependent cutoff $V_{MB}(t) = rt$:

$$\sigma(F_\mathbf{s}, t) = \frac{1}{1 + e^{\lambda(F_\mathbf{s} - rt)}}. \tag{23}$$

where r determines the rate of increase in the fill-level and has units of power. Introducing the boost potential in the form of Equation 6 with a time-dependent switching function Equation 23, ensures that the fill-level increases linearly as $V_{MB}(t) = rt$. In this step, the optimized variational coefficients that define the boost potential through Equation 22 are kept fixed and not optimized so that the only time-dependence of the bias is through the time-dependent switching function. Substituting $V_{MB}(t) = rt$ into Equation 14 from KTR theory gives an analytical expression for the cumulative probability distribution of barrier crossing times:

$$\text{CDF} = 1 - \exp\left(\frac{k_0}{\gamma\beta r}\left(1 - e^{\gamma\beta rt}\right)\right). \tag{24}$$

Note that Equation 24 is different from the CDF for a simple exponential distribution because of the time-dependence of the fill-level as the boost potential is increased. From a series of independent trajectories, the ECDF of crossing times can be fit to Equation 24 with two fitting parameters being k_0 and γ. This avoids the need of performing multiple simulations at varying fixed V_{MB} values since the fill-level is not static but steadily increasing during the simulation. Furthermore, there is no need to rescale the simulation time as is done in the hyperdynamics approach, since the

ECDF of the biased crossing times can be fit directly to the analytical result. However, to be consistent with Kramers' theory, the fill rate, r, should be chosen such that the boost always remains lower than the barrier height, which is not generally known *a priori*. In practice, a preliminary metadynamics or VES simulation can provide an estimate of the barrier height to check that the boost potential does not increase past the barrier height during the simulation.

As a closely related example, a logarithmic time dependence of the fill-level, $V_{MB} = a\log(1+bt)$, can be achieved by using a switching function of the form:

$$\sigma(F_{\mathbf{s}},t) = \frac{1}{1+e^{\lambda(F_{\mathbf{s}}-a\log(1+bt))}}. \tag{25}$$

where the parameters a and b determine the filling rate of the boost potential. From Equation 14, the cumulative distribution of crossing times is given from KTR theory as[29]

$$\text{CDF} = 1-\exp\left(\frac{k_0}{b(\beta\gamma a+1)}\left[1-(1+bt)^{\beta\gamma a+1}\right]\right) \tag{26}$$

The bias constructed using Equation 25 for the switching function in Equation 6 is comparable to metadynamics since the average max bias in iMetaD is approximately described by a logarithmic time dependence. However, the advantage here is that the fill rate can be precisely tuned by setting the parameters a and b in the switching function.

III. RESULTS

A. Alanine dipeptide in vacuum

We demonstrate the approach of using variational flooding to obtain kinetic rates from analytical theory on the well-studied alanine dipeptide (Ace-Ala-Nme) in vacuum (FIG. 2A). This system has a conformation transition from the C_{7eq} to the C_{7ax} state that can be distinguished by the two backbone dihedral angles, ϕ and ψ, and serves as a prototypical model for backbone dynamics of proteins. Simulations were performed with GROMACS-5.1.4[35–38] using the Amber99-SB force fields[39]. We used an integration time step of 0.002 ps with bonds to hydrogen atoms constrained using the LINCS algorithm[40]. The temperature was maintained at 260 K using the stochastic velocity rescaling thermostat[41]. To obtain a boost potential acting on the free energy surface, we performed a VES simulation as implemented in the VES module of PLUMED2[42] starting in the C_{7eq} state. We used a Fourier series expansion for the basis set that defines the bias potential acting only on the ϕ dihedral angle: $V_b(\phi) = \sum_k c_k e^{ik\phi}$ with periodicity $(-\pi,\pi)$ and

20 coefficients. The variational coefficients are updated every 1 ps using a stochastic gradient descent[32] with a fixed step size of 0.1. We used a committor so that the simulation stops after crossing into the C_{7ax} state. The final set of optimized coefficients are used as fixed coefficients to define a boost potential through Equation 6 that is employed in subsequent MD simulations to accelerate crossing events from the C_{7eq} to the C_{7ax} state.

1. *Variational Flooding with a Constant Fill-level.*

A boost potential with a constant fill-level is achieved by employing a switching function given by Equation 4 that ensures the bias defined through Equation 6 has a maximum value at V_{MB}. We used prescribed fill-levels of $V_{MB} = \{22, 24, 26, 28\}$ kJ/mole. The target probability distribution given by Equation 3 is self-consistently updated every 100 iterations during the bias optimization. The boost potential generated from VES simulations with different values of V_{MB} is shown in FIG. 2B. These boost potentials are sufficient to destabilize the C_{7eq} state and accelerate transitions into the C_{7ax} state. We subsequently performed 60 independent MD simulations for each value of V_{MB}. Each MD simulation was started from the same starting structure and the same initial set of bias coefficients from the prior VES simulation, but with different initial velocities randomly generated from a Maxwell-Boltzmann distribution. In each simulation the fill-level is time-independent since the value of V_{MB} in the switching function is set at a fixed value. From these 60 trajectories, the crossing time to reach the C_{7ax} state is recorded. FIG. 2C shows the cumulative distribution of biased crossing times from 60 independent trajectories at different V_{MB} values. The ECDF of crossing times shifts to shorter times as the fill-level is increased due to the increased boost potential. Each ECDF is fit to the theoretical CDF for an exponential distribution given by Equation 16 that is valid for a boost potential with a constant fill-level. A fit of the ECDF gives the mean biased transition time, τ_b, for each value of V_{MB}. FIG. 2D shows the linear plot of the log mean biased transition times, τ_b, as a function of the boost fill-level. A fit to Equation 21 gives an unbiased transition rate of $k_0 = 0.03 \pm 0.01\ \mu\text{s}^{-1}$ with $\gamma = 0.93$.

To compare this result with the hyperdynamics time rescaling approach, we scale the simulation time according to Equation 11. Upon rescaling the simulation time, the ECDFs for each set of simulations collapse to a single distribution shown in FIG. 2E. The ECDF of the rescaled times can be fit to the CDF for an exponential distribution, $1 - \exp(-t^*/\tau_0)$, where t^* is the rescaled time, to extract the mean unbiased transition time, $\tau_0 = k_0^{-1}$. The mean unbiased time is an order

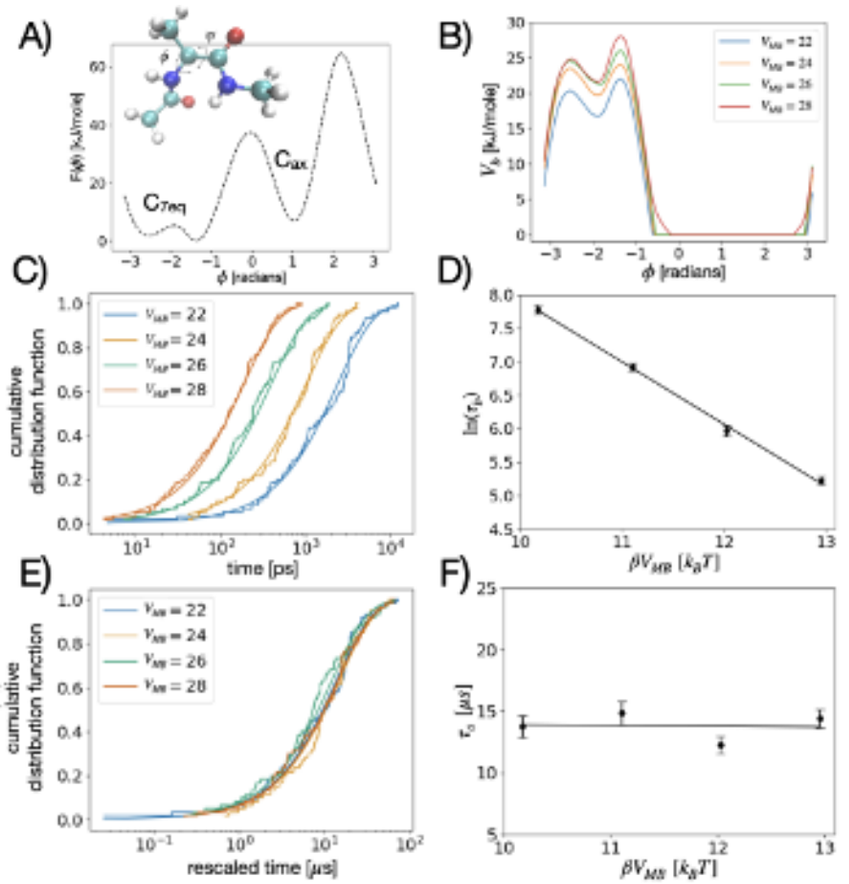

FIG. 2. A) Alanine dipeptide in vacuum. As a reference, the free energy surface along the ϕ torsion is shown with a dashed line. B) The optimized boost potential obtained from VES simulations with different prescribed fill-level values, V_{MB}, implemented through the target distribution to flood the C_{7eq} state. This leads to quasi-stationary VES boost potentials with $V_{mb} = 22$ kJ/mol (blue), $V_{mb} = 24$ kJ/mol (orange), $V_{mb} = 26$ kJ/mol (green), $V_{mb} = 28$ kJ/mol (red). C) The ECDF of crossing times from 60 biased simulations for different fill-level values, V_{MB}. The solid line is a fit to the CDF for an exponential distribution that gives the mean biased transition time, τ_b. D) Plot of the log of the mean biased transition time, τ_b, as a function of the fill-level of the boost potential. Extrapolation to the y-intercept gives the unbiased transition rate when the boost potential is zero. E) The ECDFs collapse to a single exponential distribution after rescaling the simulation time according to the hyperdynamics formalism. F) The unbiased mean transition times, τ_0, after time rescaling are independent of the fill-level, V_{MB}.

of magnitude larger than the biased transition times, showing the speed-up afforded by the boost potential. After rescaling the biased simulation time by Equation 11 the mean transition times, τ_0, are independent of the fill-level as shown in FIG. 2F. The average of these unbiased transition times gives $k_0 = 0.07 \pm 0.02\ \mu s^{-1}$ which is consistent with the above method of directly fitting the biased ECDF. However, it should be noted that the hyperdynamics approach of scaling the simulation time according to the acceleration factor does not introduce the parameter γ, and the single torsion, ϕ, is likely not an ideal CV. Prior variational flooding simulations biasing both ϕ and ψ torsion angles gave $k_0 = 0.035\ \mu s^{-1}$, in better agreement with results using Kramers' theory from this work[23].

2. *Variational Flooding with a Time-dependent Boost Potential*

To demonstrate a time-dependent boost, we performed a second set of 60 independent simulations with the boost potential now multiplied by a time-dependent switching function. First, a VES simulation was performed for 1.7 ns to optimize a boost potential through Equation 6 using the same optimization procedure as above. Subsequently, this boost potential was fit to a cubic spline and multiplied by the time-dependent switching function given by Equation 23 for a linear boost. This ensures that the fill-level of the bias increases at a constant rate during the simulation according to $V_{MB}(t) = rt$. We set $r = 5 \times 10^{-4}$ kJ mol^{-1} ps^{-1}. FIG. 3A shows the time-dependent boost potential at 5 ns increments during a single representative MD simulation, showing the steady increase of the fill-level with time. The inset shows the fill-level, $V_{MB}(t)$, increases linearly during the simulation. FIG. 3B shows the ECDF of biased crossing times using this time-dependent VES boost potential. The ECDF is fit to the analytical expression of Equation 24 from KTR theory giving an unbiased rate of $k_0 = 0.04 \pm 0.02\ \mu s^{-1}$ and $\gamma = 0.94$. This is in excellent agreement with unbiased simulations ($k_0 = 0.043 \pm 0.008\ \mu s^{-1}$) and the results using a fixed boost potential above. The advantage of this approach is that only a single set of simulations need to be performed instead of performing multiple simulations at different V_{MB} values as was done in the previous free energy flooding method that used a constant fill-level[23]. Using a VES-optimized boost potential with a time-dependent switching function and fitting the biased ECDF directly to analytical theory gives a simple method to extract unbiased rates from MD simulation.

Error bars from fits to the ECDF are estimated using a bootstrap analysis of the barrier crossing times taking 50 bootstrap samples, sampling with replacement. The size of the error bars represent

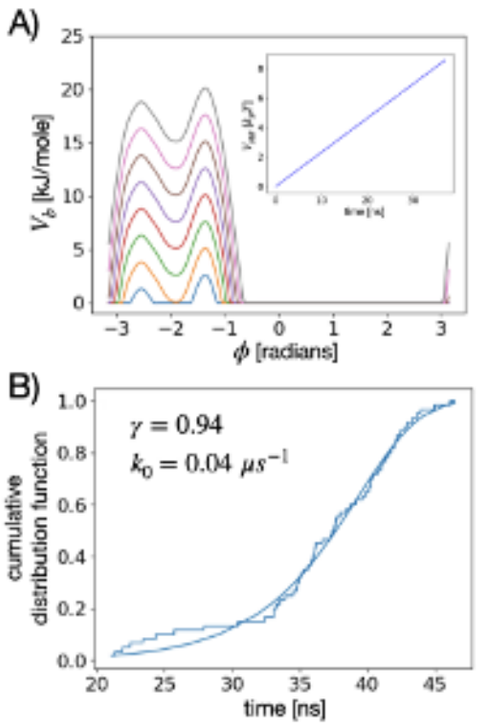

FIG. 3. A) Time series of the time-dependent boost potential from a single representative MD simulation. The instantaneous boost potential is shown every 5 ns. The flooding rate is controlled by a time-dependent switching function. The inset shows the fill-level increases linearly during the simulation time. B) The ECDF of crossing events from 60 independent simulations biased with the time-dependent boost potential. The solid line is a fit to the analytical CDF from KTR theory, giving the unbiased rate k_0 and γ parameter from a single set of simulations.

the 30 and 70 percentiles of the bootstrap samples. Table I presents a comparison of the unbiased crossing rates extracted using each method: analytical Kramers' theory with a static fill-level, analytical KTR theory with a time-dependent fill-level, and hyperdynamics time rescaling. All rate estimates are in agreement with the value obtained from long unbiased simulations reported in Ref [23], demonstrating that all these methods yield reliable kinetics of rare events using variational flooding. In this work we demonstrate the time-dependent boost only for a single reaction coordinate; however, the time-dependent switching function could be applied to a two dimensional boost and is shown in the Supporting Materials. Future work is needed to investigate the application of analytical KTR theory to a two dimensional boost potential in terms of efficiency and accuracy.

TABLE I. Comparison of theoretical CDF and kinetic parameters obtained from variational flooding simulations for alanine dipeptide

analysis method	CDF	k_0 (μs^{-1})	γ
analytical Kramers' theory with constant fill-level	$1-e^{-t/\tau_b}$	0.03 ± 0.01	0.93
analytical KTR theory with linear flooding	$1-\exp\left(\frac{k_0}{\gamma\beta r}\left(1-e^{\gamma\beta rt}\right)\right)$	0.04 ± 0.02	0.94
Hyperdynamics time rescaling	$1-e^{-t^*/\tau_0}$	0.07 ± 0.02	-
unbiased MD simulation [a]	$1-e^{-t/\tau_0}$	0.043 ± 0.008	-

[a] Ref [23]

B. S_N2 reaction in vacuum

As a second example, we consider the asymmetric nucleophilic S_N2 substitution reaction $CH_3F + Cl^- \longrightarrow CH_3Cl + F^-$ in vacuum, shown in FIG. 4A. This reaction has previously been studied by variational flooding[31], and serves as a prototypical model for the application to *ab initio* MD simulations to study reaction kinetics[43]. Simulations were performed using CP2K version 2023.1 (Development Version 2013.1)[44–46], patched with PLUMED2.9[42] using the semiempirical PM6 model[47] to reduce the computational cost. We used a supercell of 10 $\times$ 10 $\times$ 10 $Å^3$ without periodic boundary conditions, with a threshold accuracy of 10^{-5} Hartree in the self-consistent field iteration. We used a time step of 0.5 fs to update the atomic positions based on the Born-Oppenheimer approach. The temperature was set to 1200 K and maintained using the stochastic velocity rescaling thermostat[41]. In a first step, we performed a VES simulation to construct the boost potential using the CV, $s = d_1 - d_2$, where d_1 is the distance between fluorine and carbon atoms and d_2 is the distance between chlorine and carbon atoms (FIG. 4A). We used the Legendre polynomials as the basis set functions with 50 variational coefficients to be optimized. The coefficients are updated using the stochastic gradient descent with a step size of 10, and the coefficients that define the bias are updated every 1 ps. The VES simulation is stopped when the product is formed using a committor, which occurred after 50 iterations of the bias (50 ps).

We employed the VES-optimized bias as a time-dependent boost potential by using the switching function given by Equation 23 with a filling rate of $r = 0.4$ kJ mol^{-1} ps^{-1}. We performed 60 independent simulations using this time-dependent boost potential. Each simulation started from the same initial reactant configuration, but with a different initial seed for the velocity. The in-

stantaneous boost potential at intervals of 5 ps from a single representative simulation is shown in FIG. 4B. This implementation of the boost leads to a linearly increasing fill-level as shown in the inset of FIG. 4B. As a second example of a time-dependent boost, we also performed a set of 60 simulations using the switching function given by Equation 25 that leads to a logarithmic filling rate. Here, we used parameters $a = 120.0$ kJ mol^{-1} and $b = 0.01$ ps^{-1}. The instantaneous boost from a single representative simulation at intervals of 10 ps is shown in FIG. 4C, demonstrating a logarithmic growth of the fill-level.

As a control, we also performed 30 independent iMetaD simulations. Metadynamics was performed as implemented in PLUMED2[42] with a Gaussian width of 0.025 Å, a Gaussian height of 0.3 kJ mol^{-1}, and a deposition stride of 1 ps. From iMetaD simulations, the fill-level is obtained as an average over R total runs

$$V_{MB}(t) = \frac{1}{R}\sum_{r} \max_{t' \in [0,t]} V_B^r(t') \tag{27}$$

where $V_B^r(t)$ is the instantaneous bias (sum of Gaussian hills) at time t for simulation run r. As seen in FIG. 4D the fill-level determined from Equation 27 increases monotonically, resembling a logarithmic growth of the fill-level. However, because of the infrequent deposition stride required for iMetaD, the filling rate is slower than in our VES boosted simulation by about an order of magnitude.

FIG. 5A shows the ECDF of the biased crossing times for each set of simulations for the S_N2 reaction. The ECDF from simulations biased with a variational flooding boost potential with a linear fill rate is fit to the analytical expression for the CDF given by Equation 24, giving an unbiased rate of $k_0 = 26$ ms^{-1} with $\gamma \sim 1$. Confidence intervals are computed as the 30th and 70th percentiles from a bootstrap analysis. The accumulated crossing times were resampled with replacement, giving an upper and lower confidence interval for the transition rate of (10, 49) ms^{-1}. Similarly, the ECDF from simulations biased with a variational flooding boost potential with a logarithmic fill rate is fit to the analytical expression for the CDF given by Equation 26, giving an unbiased rate of $k_0 = 105$ ms^{-1} with $\gamma \sim 0.43$ and a confidence interval of (30, 250) ms^{-1}. The ECDF from iMetaD simulations was modeled using the procedure of Palacio-Rodriguez, *et al.*[29]. Briefly, the numerical $V_{MB}(t)$ from Equation 27 (FIG. 4D) is fit to a spline and the parameters k_0 and γ are determined by maximizing the likelihood function[29]

$$\mathscr{L} = \prod_{i \in \text{events}}^{M} -\frac{dS(t)}{dt}\bigg|_{t=t_i} \prod_{j \in \text{non-events}}^{N} S(t_j) \tag{28}$$

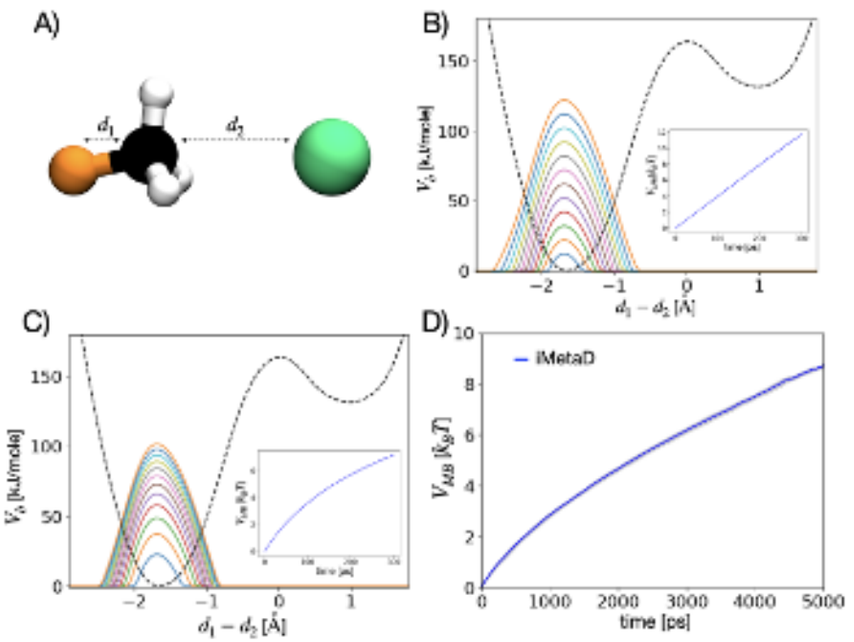

FIG. 4. A) Distances used to define the CV for the asymmetric S_N2 nucleophilic substitution reaction of fluoromethane to chloromethane. B) Time series of the time-dependent variational flooding boost potential from a representative MD simulation demonstrating linear flooding. The instantaneous boost potential is shown every 5 ps. For reference, the free energy landscape is shown as a dashed line. The inset shows the linear increase of the fill level. C) The time-dependent boost potential from a representative MD simulation demonstrating logarithmic flooding. The instantaneous boost potential is shown every 10 ps. The free energy landscape is shown as a dashed line. The inset shows a logarithmic increase of the fill level. D) The average fill level from 30 independent iMetaD simulations. The solid blue line represents the average and the gray shaded region represents the standard deviation of the maximum bias.

where $S(t)$ is the survival probability from KTR theory, given by Equation 14. In Equation 28, M is the number of crossing events observed, and N is the number of non-events observed up to time t_j. The parameters k_0 and γ are found from numerical maximization of $\ln \mathscr{L}(k_0^*(\gamma), \gamma)$ using the script provided in Ref. [29]. The quality of the fit is assessed using the Kolmogorov Smirnov test (KS-test)[25]. The ECDF of accumulated crossing times from iMetaD simulations is also presented in FIG. 5A with the fit to the numerical KTR theory giving a p-value of 0.39. Confidence intervals from iMetaD simulations are estimated from 50 bootstrap samples that pass the KS-test. The iMetaD procedure yields an unbiased rate of $k_0 = 351$ ms^{-1} and $\gamma \sim 1$, with a confidence interval of (346, 465) ms^{-1}, consistent with the rate obtained from our variational

TABLE II. Comparison of transition rates obtained from variational flooding and iMetad simulations for the asymmetric S_N2 reaction at 1200 K.

analysis method	CDF	k_0 (ms^{-1})
analytical KTR theory linear flooding	$1-\exp\left(\frac{k_0}{\gamma\beta r}\left(1-e^{\gamma\beta rt}\right)\right)$	26 (10,49)[a]
analytical KTR theory logarithmic flooding	$1-\exp\left(\frac{k_0}{b(\beta\gamma a+1)}\left[1-(1+bt)^{\beta\gamma a+1}\right]\right)$	105 (30,250)
iMetaD	numerical $1-S(t)$	351 (346,465)
Hyperdynamics time rescaling [b]	$1-e^{-t^*/\tau_0}$	154 ± 4

[a] Lower and upper confidence intervals represent the 30th and 70th percentiles from a bootstrap analysis of the transition times

[b] Ref [31]

flooding procedure.

While all ECDFs in FIG. 5A are well-described by KTR theory, the convenience of having an analytical expression afforded by the variational flooding approach simplifies the extraction of rate constants. Previously, variational flooding with a fixed boost potential found an unbiased transition rate of $k_0 = 154 \pm 4$ ms^{-1} at 1200 K[31]. Table II summarizes the results for the extracted rates from variational flooding and iMetaD simulations for the S_N2 reaction. Transition rates from variational flooding are in general agreement with the rate from iMetaD simulations, but obtained at a reduced computational cost. Finally, to compare with the previous hyperdynamics time rescaling approach[31], we scaled the times according to Equation 10. As seen in FIG. 5B, all CDF curves collapse to a single distribution for the rescaled transition times within the uncertainty range.

C. Chignolin folding

As a final example system we consider the folding of the 10-residue chignolin miniprotein[48] in explicit solvent (FIG.6A). Input files for the MD simulation were obtained from the PLUMED-NEST (plumID:22.031)[49], as provided by Ray, et al.[24]. Simulations were performed in GROMACS-2019.4[50] using the CHARMM22* force fields[51] with an integration time step of 0.002 ps and bonds to hydrogen atoms constrained using the LINCS algorithm[40]. The starting structure is taken as the unfolded chain and solvated in a box of 1907 CHARMM TIP3P water

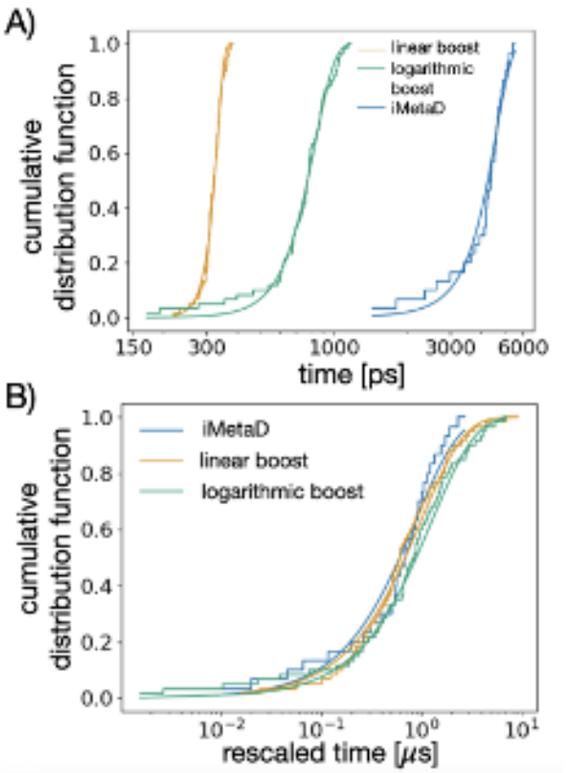

FIG. 5. A) The ECDF of crossing events from 60 independent simulations employing a linear variational flooding boost (orange), a logarithmic variational flooding boost (green), and iMetaD (blue). For variational flooding simulations, the solid line is a fit to analytical KTR theory. For iMetaD the solid line is a fit to KTR theory using the numerical $S(t)$ by maximizing the log likelihood function to estimate the optimal parameters. For iMetaD, the KS-test gives a p-value of 0.39. B) The ECDFs collapse to the same exponential distribution after rescaling the simulation time according to the hyperdynamics formalism.

molecules with two sodium ions to neutralize the system. The temperature was kept at 340 K using the stochastic velocity rescaling thermostat[41]. We used the harmonic linear discriminant analysis (HLDA) CV based on six interatomic contacts within the protein[52]. The interatomic distances and definition of the CV is described in the Supplemental Material. We first performed a VES simulation as implemented in the VES module of PLUMED2[42] using multiple walkers with four replicas to efficiently sample the unfolded state and converge the variational coefficients that define the bias. We used the Legendre polynomials as the basis set functions with 50 variational coefficients to be optimized. The coefficients are updated using the stochastic gradient descent with a step size of 0.1, and the coefficients that define the bias are updated every

TABLE III. Folding times for chignolin obtained from variational flooding

analysis method	k [μs^{-1}]	γ
time-dependent flooding	1.85 (1.35, 2.00)[a]	0.3
unbiased [b]	1.67	-

[a] Lower and upper confidence intervals represent the 30th and 70th percentiles from a bootstrap analysis of the transition times

[b] Ref [53]

4 ps. The VES simulation is stopped when the folded state is found using a committor, which occurred after 1940 iterations of the bias (7.78 ns). After the initial boost potential is generated using VES, the optimized bias is fit to a cubic spline and employed is subsequent MD simulations with a time-dependent switching function. We ran 60 independent trajectories using a linear time-dependent boost with a switching function defined by Equation 23. Each simulation started from the same unfolded structure and same bias coefficients but with velocities randomly generated from a Maxwell-Boltzmann distribution. For the linear time-dependent boost we used a constant fill rate of $r = 2.5 \times 10^{-4}$ kJ mol^{-1} ps^{-1}. The folding times are compared to the folding time of 0.6 μs reported from a 106 μs unbiased trajectory[53].

The instantaneous boost potential at intervals of 5 ns from a representative MD simulation is shown in FIG. 6B, leading to a linear flooding rate. The empirical distribution of crossing times is well described by the analytical KTR theory (FIG. 6C). A fit to Equation 24 gives the unbiased folding time of 0.54 μs in good agreement with unbiased MD simulation. The fit gives $\gamma \sim 0.3$ indicating that the single HLDA CV is suboptimal, yet the procedure still yields an accurate kinetic estimate. Table III summarizes the comparison of chignolin folding times from time-dependent flooding simulations with unbiased MD.

IV. DISCUSSION AND CONCLUSION

We have investigated an application of KTR theory to the variational flooding method for rate calculation from biased MD simulations. In a first step, a boost potential that is able to accelerate barrier crossing events is constructed using VES. Because we are only interested in accelerating transitions out of a metastable basin, and not converging the entire free energy surface, the VES

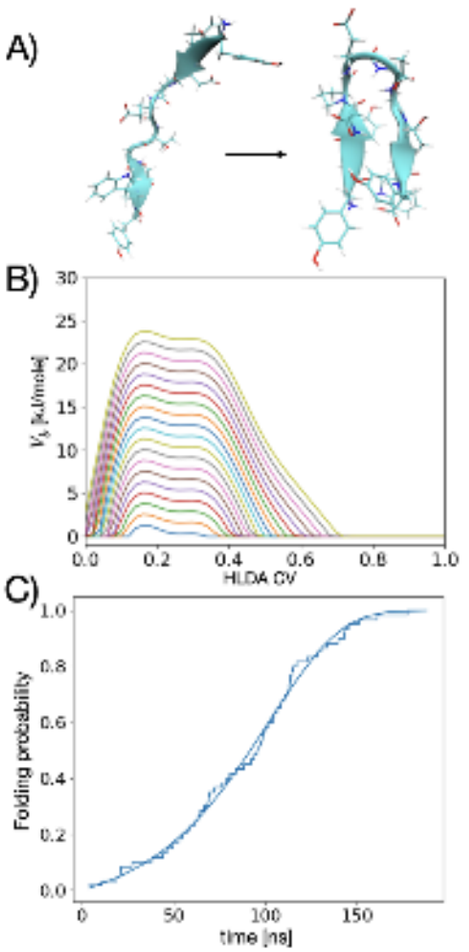

FIG. 6. A) Structures representing the unfolded and folded states of chignolin. B) Time series of the time-dependent boost potential from a representative MD simulation. The instantaneous boost potential is shown every 5 ns and increases linearly during the simulation. C) The biased ECDF of crossing events from 60 independent simulations employing a linear variational flooding boost. The solid line is a fit to analytical KTR theory Equation 24.

bias converges quickly and the optimization is stopped after a crossing event is observed. In a subsequent step, we employ the optimized bias with fixed coefficients and a switching function that determines the fill-level of the boost. The method lends itself to the construction of a boost potential that results in a distribution of crossing times that can be described by analytical KTR theory. This procedure greatly simplifies the extraction of unbiased rates from biased simulations since the ECDF of crossing times from multiple independent simulations is fit directly to an analytical expression. We have demonstrated the approach for three different switching functions

that lead to different time-dependencies of the boost potential. In the case of a time-independent switching function with a static cutoff, V_{MB}, the model recovers the expected Kramers' result of an exponential dependence of the transition rate on the barrier height. In this case, the ECDF of crossing times is consistent with a Poisson process for an activated process, leading to an exponential distribution of crossing times. However, because the fit to a single exponential function has only one parameter, the mean biased time, τ_b, it is not possible to extract both γ and k_0 from a single set of simulations with a fixed V_{MB}. Furthermore, we show that the error introduced by setting $\gamma = 1$ even for an ideal CV amounts to setting the average bias everywhere to its maximum $< V_b >= V_{MB}$ and assuming higher order cumulants vanish. For a fixed boost potential, it is better to perform multiple simulations with increasing V_{MB} to extract both γ and k_0 from a linear fit to the data. For example, when only the dihedral angle, ϕ, is used as a CV for the alanine dipeptide system, we obtained $\gamma = 0.93$. This procedure gives a kinetic rate in quantitative agreement with unbiased simulations of the same system. A consequence of introducing γ is that one only requires an approximate reaction coordinate to accelerate rare events. This is an advantage of the KTR approach as compared to the hyperdynamics approach of rescaling the simulation time by the acceleration factor.

Introducing a time-dependent value of $V_{MB}(t)$ into the variational flooding procedure through a time-dependent switching function, eliminates the need of performing multiple sets of MD simulations at different fixed V_{MB} values. Here, we investigated two time-dependent switching functions: one that results in linear flooding, and another that leads to logarithmic flooding. Both procedures are easily implemented through a Fermi-type switching function, and extracted kinetic rates are in quantitative agreement with rates determined using the hyperdynamics time rescaling approach. Parameters of the time-dependent switching function, including the fill rate of the boost, must be chosen carefully to ensure the boost remains below the barrier height. In this work, we monitor the bias value during the simulation to check that the boost has not increased past a threshold value. In this respect, the logarithmic flooding achieved through Equation 25 is useful for implementing a boost that fills at a less extreme filling rate near the barrier. Such a boost may be applicable for situations where the system must find the near attack conformation before reacting to the product state.

In contrast to iMetaD, which requires multiple MD simulations with a slow deposition of Gaussian hills, the variational flooding procedure uses a boost potential optimized using a prior VES simulation. This two-step procedure improves the computational efficiency of obtaining rates from

biased MD simulations. This is important for applications to *ab initio* simulations that are more computationally intensive, where the slow bias deposition from iMetaD becomes prohibitive. With variational flooding, one only needs to perform one preliminary simulation to construct a suitable boost potential, and then use the time-dependent switching function to accelerate transitions out of the basin. The speed up is afforded by the flexibility of the switching function to control the filling rate at a desired level that is introduced in a post-hoc manner after the VES optimization step. Although VES was chosen as the enhanced sampling method to construct the boost in this work, one could apply the switching function to any boost potential, such as the bias obtained from a preliminary metadynamics or OPES simulation. Another possible extension would be to perform infrequent metadynamics on top of the time-dependent flooding bias. Metadynamics encourages the exploration of conformational space and could prevent the system from getting stuck if the VES bias is not fully converged. For a sufficiently slow deposition stride, the filling rate would still be controlled by the time-dependent flooding potential.

One limitation of the present work is the focus on only a single CV. We envision that future applications may take advantage of data-driven machine learned CVs[54] or DeepTICA CVs[55] to enable a one dimensional boost for more complex chemical reactions or system with more conformational complexity.

In conclusion, we have demonstrated an application of variational flooding with a time-dependent boost to accelerate barrier crossing in MD simulations. We have illustrated how the fill rate of a VES-optimized bias can be controlled to achieve a boost potential that is amenable to analytical KTR theory. We term this method Kramers Rate Approach to Variationally Enhanced Sampling (KRAVES). This method that combines variational flooding with KTR theory leads to a convenient way to extract kinetic rates from an analytical expression for the distribution of crossing times. This procedure should be useful for studying reaction kinetics from MD simulations.

V. SUPPLEMENTAL MATERIAL

The Supplement Material includes details on the construction of a two dimensional VES bias for use with a time-dependent switching function and details of the HLDA CV used for chignolin folding.

VI. ACKNOWLEDGMENTS

This work was funded by the National Science Foundation NSF CHE-2102189.

VII. DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request. Simulation data files, input scripts, and analysis codes used to analyze the transition times from this work along with chignolin unfolding trajectories are available on zenodo at https://doi.org/10.5281/zenodo.14010449.

[1] J. Hénin, T. Lelièvre, M. R. Shirts, O. Valsson, and L. Delemotte, Living Journal of Computational Molecular Science **4**, 1583 (2022).

[2] G. M. Torrie and J. P. Valleau, Journal of Computational Physics **23**, 187 (1977).

[3] A. Laio and M. Parrinello, Proceedings of the national academy of sciences **99**, 12562 (2002).

[4] H. Grubmüller, B. Heymann, and P. Tavan, Science **271**, 997 (1996).

[5] F. Colizzi, R. Perozzo, L. Scapozza, M. Recanatini, and A. Cavalli, Journal of the American Chemical Society **132**, 7361 (2010).

[6] J. Comer, J. C. Gumbart, J. Hénin, T. Lelièvre, A. Pohorille, and C. Chipot, The Journal of Physical Chemistry B **119**, 1129 (2015).

[7] E. Darve and A. Pohorille, The Journal of chemical physics **115**, 9169 (2001).

[8] Y. Sugita and Y. Okamoto, Chemical physics letters **314**, 141 (1999).

[9] D. Hamelberg, J. Mongan, and J. A. McCammon, The Journal of chemical physics **120**, 11919 (2004).

[10] O. Valsson and M. Parrinello, Physical review letters **113**, 090601 (2014).

[11] M. Invernizzi and M. Parrinello, The journal of physical chemistry letters **11**, 2731 (2020).

[12] O. Valsson, P. Tiwary, and M. Parrinello, Annual review of physical chemistry **67**, 159 (2016).

[13] P. Tiwary and M. Parrinello, The Journal of Physical Chemistry B **119**, 736 (2015).

[14] B. W. Zhang, D. Jasnow, and D. M. Zuckerman, The Journal of chemical physics **132** (2010).

[15] D. M. Zuckerman and L. T. Chong, Annual review of biophysics **46**, 43 (2017).

[16] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, The Journal of chemical physics **108**, 1964 (1998).

[17] B. E. Husic and V. S. Pande, Journal of the American Chemical Society **140**, 2386 (2018).

[18] J. Copperman and D. M. Zuckerman, Journal of chemical theory and computation **16**, 6763 (2020).

[19] A. F. Voter, Physical Review Letters **78**, 3908 (1997).

[20] H. Grubmüller, Physical Review E **52**, 2893 (1995).

[21] Y. Miao, V. A. Feher, and J. A. McCammon, Journal of chemical theory and computation **11**, 3584 (2015).

[22] P. Tiwary and M. Parrinello, Physical review letters **111**, 230602 (2013).

[23] J. McCarty, O. Valsson, P. Tiwary, and M. Parrinello, Physical review letters **115**, 070601 (2015).

[24] D. Ray, N. Ansari, V. Rizzi, M. Invernizzi, and M. Parrinello, Journal of Chemical Theory and Computation **18**, 6500 (2022).

[25] M. Salvalaglio, P. Tiwary, and M. Parrinello, Journal of chemical theory and computation **10**, 1420 (2014).

[26] R. Casasnovas, V. Limongelli, P. Tiwary, P. Carloni, and M. Parrinello, Journal of the American Chemical Society **139**, 4780 (2017).

[27] J. M. L. Ribeiro, S.-T. Tsai, D. Pramanik, Y. Wang, and P. Tiwary, Biochemistry **58**, 156 (2018).

[28] S. Lee, D. Wang, M. A. Seeliger, and P. Tiwary, bioRxiv (2024).

[29] K. Palacio-Rodriguez, H. Vroylandt, L. S. Stelzl, F. Pietrucci, G. Hummer, and P. Cossio, The Journal of Physical Chemistry Letters **13**, 7490 (2022).

[30] K. A. Croney and J. McCarty, The Journal of Physical Chemistry B **128**, 3102 (2024).

[31] G. Piccini, J. J. McCarty, O. Valsson, and M. Parrinello, The journal of physical chemistry letters **8**, 580 (2017).

[32] F. Bach and E. Moulines, Advances in neural information processing systems **26** (2013).

[33] O. Valsson and M. Parrinello, Journal of chemical theory and computation **11**, 1996 (2015).

[34] G. Hummer and A. Szabo, Biophysical journal **85**, 5 (2003).

[35] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, SoftwareX **1**, 19 (2015).

[36] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, Journal of computational chemistry **26**, 1701 (2005).

[37] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. Van Der Spoel, *et al.*, Bioinformatics **29**, 845 (2013).

[38] S. Páll, M. J. Abraham, C. Kutzner, B. Hess, and E. Lindahl, in *Solving Software Challenges for Exascale: International Conference on Exascale Applications and Software, EASC 2014, Stockholm, Sweden, April 2-3, 2014, Revised Selected Papers 2* (Springer, 2015) pp. 3–27.

[39] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, Proteins: Structure, Function, and Bioinformatics **78**, 1950 (2010).

[40] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, Journal of computational chemistry **18**, 1463 (1997).

[41] G. Bussi, D. Donadio, and M. Parrinello, The Journal of chemical physics **126** (2007).

[42] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, Computer physics communications **185**, 604 (2014).

[43] K. L. Fleming, P. Tiwary, and J. Pfaendtner, The Journal of Physical Chemistry A **120**, 299 (2016).

[44] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, *et al.*, The Journal of Chemical Physics **152** (2020).

[45] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele, Wiley Interdisciplinary Reviews: Computational Molecular Science **4**, 15 (2014).

[46] T. D. Kühne, M. Krack, F. R. Mohamed, and M. Parrinello, Physical review letters **98**, 066401 (2007).

[47] J. J. Stewart, Journal of Molecular modeling **13**, 1173 (2007).

[48] S. Honda, T. Akiba, Y. S. Kato, Y. Sawada, M. Sekijima, M. Ishimura, A. Ooishi, H. Watanabe, T. Odahara, and K. Harata, Journal of the American Chemical Society **130**, 15327 (2008).

[49] Nature methods **16**, 670 (2019).

[50] E. Lindahl, M. Abraham, B. Hess, and D. Van Der Spoel, URL https://doi. org/10.5281/zenodo **2424363** (2022).

[51] A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, *et al.*, The journal of physical chemistry B **102**, 3586 (1998).

[52] D. Mendels, G. Piccini, Z. F. Brotzakis, Y. I. Yang, and M. Parrinello, The Journal of chemical physics **149** (2018).

[53] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, Science **334**, 517 (2011).

[54] L. Bonati, V. Rizzi, and M. Parrinello, The journal of physical chemistry letters **11**, 2998 (2020).

[55] L. Bonati, G. Piccini, and M. Parrinello, Proceedings of the National Academy of Sciences **118**, e2113533118 (2021).