

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/107945>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Exploring High Dimensional Free Energy Landscapes: Temperature Accelerated Sliced Sampling

Shalini Awasthi¹ and Nisanth N. Nair^{1, a)}

Department of Chemistry, Indian Institute of Technology, Kanpur, 208016, India

(Dated: 15 February 2017)

Biased sampling of collective variables is widely used to accelerate rare events in molecular simulations and to explore free energy surfaces. However, computational efficiency of these methods decreases with increasing number of collective variables, which severely limits the predictive power of the enhanced sampling approaches. Here we propose a method called Temperature Accelerated Sliced Sampling (TASS) that combines temperature accelerated molecular dynamics with umbrella sampling and metadynamics to sample the collective variable space in an efficient manner. The presented method can sample a large number of collective variables and is advantageous for controlled exploration of broad and unbound free energy basins. TASS is also shown to achieve quick free energy convergence and is practically usable with *ab initio* molecular dynamics techniques.

Keywords: Metadynamics, Umbrella Sampling, Temperature Accelerated Molecular Dynamics, Reweighting, Weighted Histogram Analysis, Free energy calculations

I. INTRODUCTION

In a canonical ensemble molecular dynamics (MD) simulation, configurations are sampled with the probability

$$P(\mathbf{R}) = \frac{e^{-\beta U(\mathbf{R})}}{Z} \quad (1)$$

where \mathbf{R} is the configuration of a molecular system with N number of atoms, $\beta = 1/k_B T$ with Boltzmann constant k_B and temperature T . Here U is the potential energy, and Z is the configurational partition function. Let the order parameter be $\zeta(\mathbf{R})$, then the probability along ζ is given by

$$P(\zeta') = \frac{1}{Z} \int d\mathbf{R} \delta(\zeta(\mathbf{R}) - \zeta') e^{-\beta U(\mathbf{R})} . \quad (2)$$

The Helmholtz free energy along ζ can then be computed as

$$F(\zeta) = -\frac{1}{\beta} \ln P(\zeta) + f , \quad (3)$$

where f is some constant. $F(\zeta)$ could be directly obtained from the probability distribution of ζ computed from a canonical ensemble MD simulation, provided a proper sampling of ζ is achieved.¹⁻⁷

Often it is more convenient to assume that ζ is a linear combination of a few collective variables $\{S_\alpha(\mathbf{R})\}$. In practice, probability distribution $P(\mathbf{S})$ for the set of selected collective variables is constructed as,

$$P(\mathbf{S}') = \frac{1}{Z} \int d\mathbf{R} e^{-\beta U(\mathbf{R})} \prod_\alpha \delta(S_\alpha(\mathbf{R}) - S'_\alpha) , \quad (4)$$

^{a)}Corresponding Author: nnair@iitk.ac.in

thus

$$F(\mathbf{S}) = -\frac{1}{\beta} \ln P(\mathbf{S}) + f \quad (5)$$

and the minimum energy pathway can be traced on the multi-dimensional surface $F(\mathbf{S})$. This assumes that we have the knowledge of \mathbf{S} for describing the process of our interest. The current work presumes that the set of collective variables $\{S_\alpha\}$ to describe and to sample the distribution is known, however, the number of collective variables is large. Although, the number of coordinates to describe a process is often small in number,^{8,9} several other orthogonal coordinates have to be enhanced-sampled for a quick convergence in probability distribution along the reactive coordinates and thus the free energy estimates.

The timescale at which a barrier crossing event takes place on a potential energy landscape during a canonical ensemble simulation is $\propto e^{\beta U(\mathbf{R})}$. Due to the limitation of small time steps in MD simulations, the simulation time to observe such processes becomes very large and computationally unfeasible for many interesting processes with free energy barrier $\Delta F^\ddagger \gg \beta^{-1}$. One of the ways in which this timescale bottleneck can be overcome is by modifying the Boltzmann weight through altering $U(\mathbf{R})$ as $U(\mathbf{R}) + U^{\text{bias}}(\mathbf{S})$ where $U^{\text{bias}}(\mathbf{S})$ is the bias potential. Metadynamics¹⁰⁻¹⁴ (MTD) and Umbrella Sampling (US)^{15,16} are two such popular biased sampling methods, among several others¹⁷⁻²⁴.

In MTD, a time dependent bias potential, $U^{\text{bias}} \equiv V^{\text{b}}(\mathbf{S}, t)$, is constructed by summing the Gaussian potentials deposited along the trajectory $\mathbf{S}(t)$:

$$V^{\text{b}}(\mathbf{S}, t) = \sum_{\tau < t} w_\tau \exp \left[-\frac{\{\mathbf{S} - \mathbf{S}(\tau)\}^2}{2(\delta s)^2} \right], \quad (6)$$

In the Well Tempered (WT-MTD)²⁵ variant of MTD,

$$w_\tau = w_0 \exp \left[-\frac{V^{\text{b}}(\mathbf{S}, t)}{k_B \Delta T} \right] \quad (7)$$

where w_0 is the initial Gaussian height and ΔT is a parameter. Free energy estimate can be obtained as²⁶

$$F(\mathbf{S}) = -\gamma \lim_{t \rightarrow \infty} V^{\text{b}}(\mathbf{S}, t) + f \quad (8)$$

where

$$\gamma = (T + \Delta T) / \Delta T \quad (9)$$

and f is some constant.

The main advantage of MTD is that it is capable of sampling the \mathbf{S} space in a self-guided manner, and thus the method can explore unprecedented minima and reaction pathways on high-dimensional free energy landscapes.^{9,12,14} In MTD simulations, the computational time required to fill a free energy basin exponentially depends on the number of collective variables. Thus, in practice, MTD simulations are carried out using limited number of collective variables (often up to 3). In order to increase the efficiency of sampling large number of coordinates, parallel tempering MTD²⁷, bias-exchange MTD^{28,29}, replica exchange with collective variable tempering³⁰, parallel bias MTD³¹, and variational MTD³² methods have been proposed.

In US simulations, a time independent harmonic restraint bias potential, $U^{\text{bias}} \equiv W_h^{\text{b}}$, is applied at chosen discrete values of \mathbf{S} , given by

$$W_h^{\text{b}}(\mathbf{S}) = \frac{1}{2} \kappa_h (\mathbf{S} - \mathbf{S}_h)^2, \quad h = 1, \dots, M \quad (10)$$

where \mathbf{S}_h is the position of the umbrella window h . To obtain $F(\mathbf{S})$, the distribution of \mathbf{S} from M windows are reweighted and stitched together

by the weighted histogram analysis (WHAM) method.^{33,34} The sampling of the collective variables are determined by the span of the windows, and thus US allows us to achieve a controlled sampling of collective variable space. Like in MTD, the computational cost increases with the number of dimensions and most of the applications using this technique have been limited to one or two collective variables only.

Another way to accelerate the sampling of collective variables is by modifying the Boltzmann factor using $\tilde{\beta} \ll \beta$, where $\tilde{\beta}$ corresponds to the temperature \tilde{T} , which is much greater than the system temperature T . This is achieved in Temperature Accelerated Molecular Dynamics (TAMD)/driven-Adiabatic Free Energy Dynamics (d-AFED)^{1,35,36} approach by defining an extended system where a set of auxiliary variables $\{s_\alpha\}$ is introduced that couple with $\{S_\alpha\}$ through a harmonic potential. Further, $\{s_\alpha\}$ is thermostated to $\tilde{\beta}$, while the physical system is thermostated to β , and the free energy at β can be computed as,^{1,36,37}

$$F(\mathbf{s}) = -\frac{1}{\tilde{\beta}} \ln \tilde{P}(\mathbf{s}) + f \quad (11)$$

where $\tilde{P}(\mathbf{s})$ is the probability distribution of $\{s_\alpha\}$ computed at $\tilde{\beta}$. Tuckerman and co-workers³⁸ have integrated TAMD/d-AFED with biased sampling approach to improve its efficiency and further extended this approach to build a free energy minimization procedure to locate saddle points and minimum energy pathways on complex free energy landscapes.³⁹ It may be noted that in their "heating and flooding" approach, both temperature acceleration and the bias potentials are applied simultaneously to all the collective variables. Recently Yang et al.⁴⁰ has combined Integrated Tempering Sampling approach with TAMD/d-

AFED.

We have recently introduced a method called Well-Sliced MTD (WS-MTD)⁴¹ to overcome the limitation of metadynamics in sampling broad and unbound free energy basins which are encountered often in the case of A+B type of chemical reactions, drug binding, protein folding etc. In this technique, we have combined US and MTD to sample orthogonal collective variables simultaneously. Such a combination approach was also proposed earlier by Filizola et al.⁴² US allows to achieve controlled sampling of collective variables, while MTD allows to sample orthogonal variables in a self-guided manner. However, the efficiency of this approach also decreases with increasing number of collective variables.

In the current work, we introduce a technique called Temperature Accelerated Sliced Sampling (TASS), which extends the WS-MTD approach to explore free energy landscape with large number of collective variables. The efficiency is improved by introducing temperature acceleration of collective variables in the spirit of TAMD/d-AFED. The method could be considered as an improvement to MTD and TAMD/d-AFED approaches to sample broad and unbound surfaces in an efficient manner. Furthermore, this method may also be looked at as an extension to the US for incorporating large number of orthogonal coordinates. At first, we will discuss the theory behind the TASS approach, and then demonstrate its efficiency for the following four problems: (a) exploring a three dimensional potential model; (b) computing the free energy landscape in the space of four backbone torsions for alanine tripeptide *in vacuo* using the AMBER force-field; (c) modeling cyclization reaction

of butadiene using *ab initio* Car–Parrinello MD by sampling three collective variables; (d) computing the free energy barrier for the hydrolysis reaction of an enzyme–drug complex by sampling four collective variables in a density functional theory (DFT) based QM/MM MD simulation.

II. THEORY

In the TASS approach, we use the Hamiltonian

$$H_h(\mathbf{R}, \mathbf{P}, \mathbf{s}, \mathbf{p}) = H^0(\mathbf{R}, \mathbf{P}) + \sum_{\alpha=1}^n \left[\frac{p_\alpha^2}{2\mu_\alpha} + \frac{k_\alpha}{2} (S_\alpha(\mathbf{R}) - s_\alpha)^2 \right] + W_h^b(s_1) + V_h^b(s_2, t) + \text{bath}(\mathbf{P}; T) + \text{bath}(\mathbf{p}; \tilde{T}) , \quad (12)$$

where $h = 1, \dots, M$ and $n \geq 2$. Here H^0 is the system Hamiltonian, \mathbf{R} and \mathbf{P} are the set of all atomic positions and momenta, and $\mathbf{S}(\mathbf{R})$ is the set of n collective variables. Importantly, n number of auxiliary variables $\{s_\alpha\}$ with masses $\{\mu_\alpha\}$ and momenta $\{p_\alpha\}$ are introduced that couple to the collective variables $\{S_\alpha\}$ by a harmonic potential with coupling constants $\{k_\alpha\}$. Along s_1 and s_2 , umbrella and metadynamics bias potentials $W_h^b(s_1)$ and $V_h^b(s_2, t)$ are added, respectively. The atomic system is coupled to a thermal bath at temperature T and auxiliary variables are coupled to a thermostat at temperature \tilde{T} . Also, $\{k_\alpha\}$ and $\{\mu_\alpha\}$ values are chosen such that the dynamics of $\{s_\alpha\}$ is close to $\{S_\alpha\}$ and they are adiabatically decoupled as done in a regular TAMD/d–AFED simulation.^{1,35,36} Our aim is to construct the free energy landscape $F(\mathbf{s})$ at temperature T . The Hamiltonian in Equation (12), allows one to sample the collective variables space by a combination of US, MTD, and TAMD/d–AFED. Especially, the

temperature accelerated sampling in the spirit of TAMD/d–AFED allows one to choose larger number of collective variables compared to other biased sampling techniques. In Equation (12), only one collective variable is biased using US and MTD, however, a larger number of variables could be biased in a straightforward manner.

Following the reweighting equations for WS–MTD as used in our previous work⁴¹, we first reweight the metadynamics bias potential^{37,43,44} as

$$\tilde{P}_h(\mathbf{s}') = \frac{\int d\tau A_h(\tau) \prod_\alpha^n \delta(s_\alpha(\tau) - s'_\alpha)}{\int d\tau A_h(\tau)} , \quad (13)$$

where

$$A_h(\tau) = \exp \left[\tilde{\beta} \left\{ V_h^b(s_2(\tau), \tau) - c_h(\tau) \right\} \right]$$

with

$$c(t) = \frac{1}{\tilde{\beta}} \ln \left[\frac{\int ds_2 \exp[\tilde{\beta} \gamma V^b(s_2, t)]}{\int ds_2 \exp[\tilde{\beta} (\gamma - 1) V^b(s_2, t)]} \right] ,$$

and the constant γ is given by Equation (9). In the subsequent step, we reweight $\tilde{P}_h(\mathbf{s})$ for the umbrella bias potential and combine distributions of all umbrella windows using the standard WHAM approach. In this procedure, the reweighted distribution $\tilde{P}(\mathbf{s})$ is obtained from M number of $\tilde{P}_h(\mathbf{s})$ using a self-consistent approach using

$$\tilde{P}(\mathbf{s}) = \frac{\sum_{h=1}^M n_h \tilde{P}_h(\mathbf{s})}{\sum_{h=1}^M n_h \exp[\tilde{\beta} f_h] \exp[-\tilde{\beta} W_h(s_1)]} , \quad (14)$$

with

$$\exp[-\tilde{\beta} f_h] = \int ds_1 \exp[-\tilde{\beta} W_h^b(s_1)] \tilde{P}(\mathbf{s}) , \quad (15)$$

and W_h^b is given by Equation (10). Here n_h is the number of configurations sampled in the h^{th} window of the umbrella potential. If the collective variables and the auxiliary variables are adiabatically separated, the distribution $\tilde{P}(\mathbf{s})$ at higher

temperature \tilde{T} is related to $P(\mathbf{s})$ at temperature T as³⁶

$$P(\mathbf{s}) \propto \tilde{P}(\mathbf{s})^{(\beta/\tilde{\beta})}. \quad (16)$$

Then, the free energy surface at temperature T can be obtained using Equation (11). However, in TAMD/d-AFED, special care should be taken for obtaining a canonical distribution for $\{s_\alpha\}$; Friction coefficient for the Langevin thermostat is chosen appropriately to achieve the desired target temperature for the extended variables (monitoring the running average of their temperature).

III. RESULTS AND DISCUSSION

A. Three Dimensional Model System

For testing the method, we considered a three-dimensional model system that has four minima:

$$U(x, y, z) = \sum_{i=1}^3 U_i^0 \exp \left(-w_i \left[(x - x_i^0)^2 + b_i (y - y_i^0)^2 + c_i (z - z_i^0)^2 \right] \right) \quad (17)$$

Parameters for the potential are given in Table SI1, and the plot of $U(x, y, z)$ is shown in Figure 1a. The four minima are labeled as **A**, **B**, **C** and **D**, and the barriers in this potential energy landscape are tabulated in Table SI2. Mass of the system was taken as 1.0 a.m.u. and MD time step was chosen as 0.24 fs.

In the TASS simulation, x , y , and z coordinates were chosen as collective variables; i.e. $S_1 \equiv x$, $S_2 \equiv y$, and $S_3 \equiv z$. Masses of auxiliary variables $\{s_\alpha\}$ were taken as 40.0 a.m.u. and values of k_α were taken as 3.14×10^3 kcal mol⁻¹ Bohr⁻².

Temperature acceleration was then invoked along all the auxiliary variables $\{s_\alpha\}$. The system temperature was set to 300 K, while that of

the auxiliary variables was set to 600 K. Temperature of the system and that of the auxiliary variables were maintained using two separate Langevin thermostats with friction coefficients 0.02 fs⁻¹ and 0.04 fs⁻¹, respectively. Auxiliary variables s_1 and s_2 were (arbitrarily) chosen for applying US and MTD biases, respectively. Umbrella potentials were placed along s_1 from -0.5 Bohr to 6.5 Bohr at intervals of 0.5 Bohr. The initial structure for any given umbrella window was generated by setting the s_1 coordinate to that corresponding to the equilibrium value of the umbrella window, while the other coordinates were having the same values as in the minimum **A**. Restraining potential κ_i used for all the umbrella potentials was 31.4 kcal mol⁻¹ Bohr⁻². The initial Gaussian height (w_0) was set to 0.6 kcal mol⁻¹ and the Gaussian width parameter δs was 0.5 Bohr. The parameter ΔT was taken as 1200 K. MTD bias potential was updated every 200 MD steps.

The convergence of free energy barriers as a function of simulation length (per umbrella window) is shown in Table SI2 and Figure 1c. From Figure 1d, it is clear that the free energy estimates converge to the exact result with increase in simulation time. The converged free energy surface is also plotted in Figure 1b. Positions of these minima and the topology of the potential energy surface are correctly reproduced in the reconstructed free energy surface.

These results show that a free energy surface with multiple minima and complex topology can be efficiently explored by the TASS method. Moreover, the free energy estimates systematically converge to the exact results.

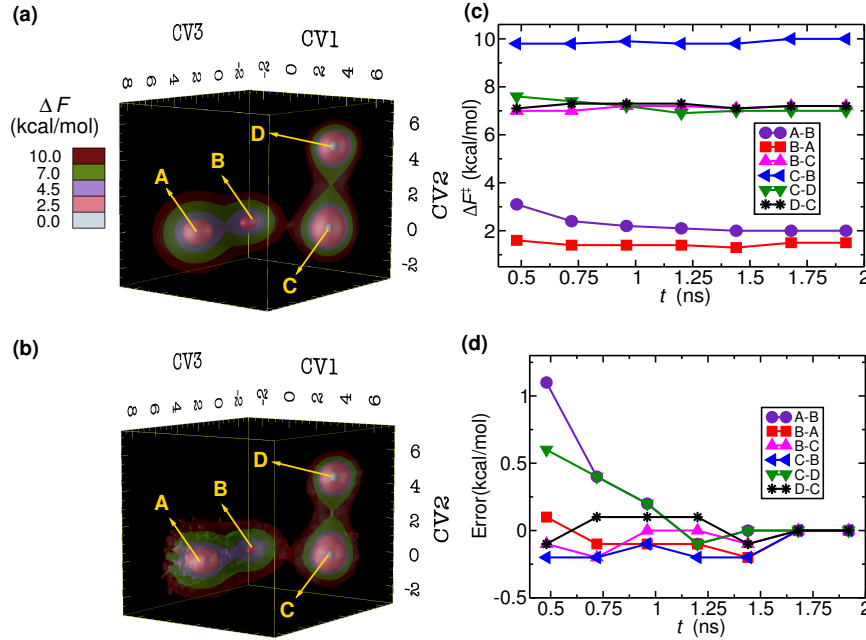


FIG. 1: (a) Three-dimensional model potential with four minima used to test the TASS method, visualized as contour surfaces; (b) Free energy surface reconstructed using the TASS method; (c) Free energy barriers computed from TASS as a function of simulation time per umbrella window; (d) Exact error in the free energy barrier estimates as a function of simulation time per umbrella window.

B. Alanine Tripeptide

The free energy surface of alanine tripeptide (Figure 2a) *in vacuo* as a function of four backbone angles ($\phi_1, \psi_1, \phi_2, \psi_2$) is explored here. Alanine tripeptide was modeled using the ff14SB force-field⁴⁵ and MD simulations were carried out using the PLUMED-AMBER interface.^{46,47} The time step was chosen as 1.0 fs.

Here, umbrella bias was applied along the ϕ_1 while MTD bias was applied along the ϕ_2 . All the four coordinates were sampled using high temperature. MTD bias potentials were updated every 500 fs and the parameters $w_0 = 0.6 \text{ kcal mol}^{-1}$, $\delta s = 0.05$ radians and $\Delta T =$

900 K were taken. Umbrella potentials were placed from $-\pi$ to π at an interval of 0.2 radians with $\kappa_h = 1.2 \times 10^2 \text{ kcal mol}^{-1} \text{ rad}^{-2}$, $k_\alpha = 1.2 \times 10^3 \text{ kcal mol}^{-1} \text{ rad}^{-2}$, and a mass of 50 a.m.u. $\text{\AA}^2 \text{ rad}^{-2}$ was assigned to all the auxiliary variables.

The initial structure for any given umbrella window was generated arbitrarily by setting the ϕ_1 internal coordinate to the equilibrium of the umbrella window, while the other collective variables were corresponding to the minimum P. Langevin thermostat with a friction coefficient of 0.001 fs^{-1} was used for maintaining the temperature of physical system at 300 K. An overdamped Langevin thermostat with a friction coefficient of

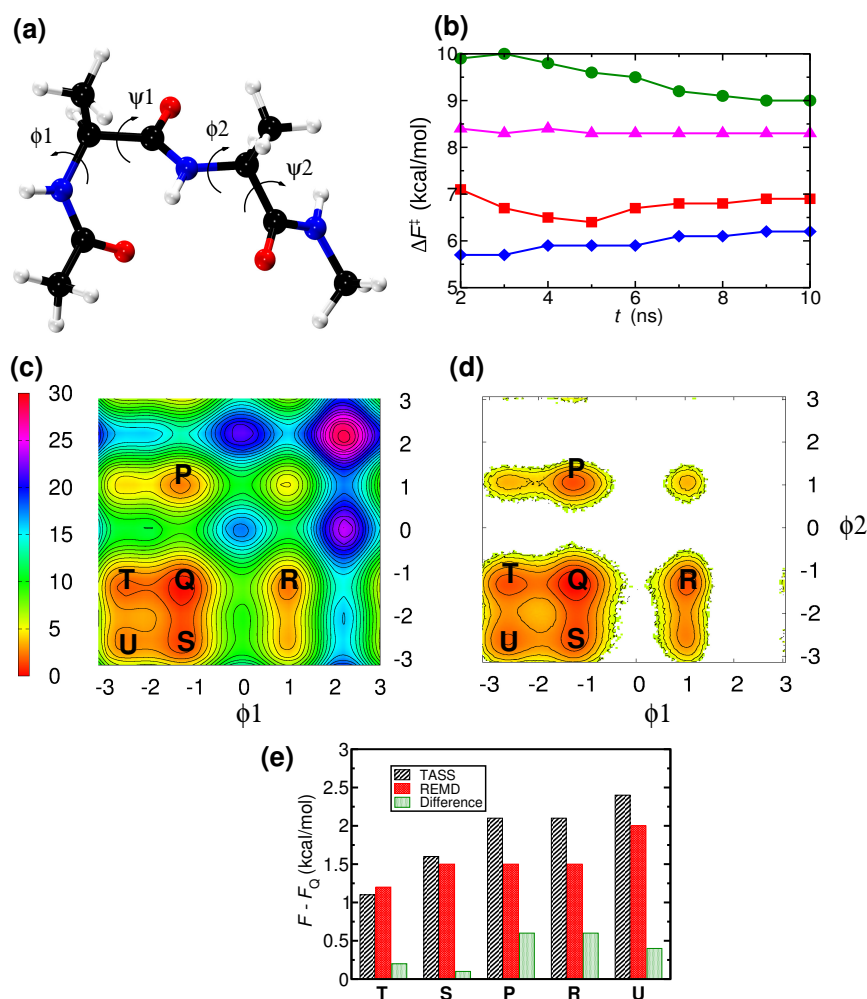


FIG. 2: (a) Ball and stick representation of alanine tripeptide. ϕ and ψ are defined as dihedral angle between atoms C–N–C $_{\alpha}$ –C and N–C $_{\alpha}$ –C–N, respectively as shown in the figure. color code: H (white), C (black), O (red), and N (blue). (b) Convergence of free energy barriers as a function of simulation time per umbrella window. Here symbols \blacksquare , \bullet , \blacklozenge , \blacktriangle represent free energy barriers for P \rightarrow Q, Q \rightarrow P, R \rightarrow Q, and Q \rightarrow R, respectively. Projection of the reconstructed five-dimensional free energy surface on (ϕ_1 , ϕ_2) plane as obtained from (c) TASS simulation, and (d) REMD simulation. Contour values are shown for every 1 kcal mol $^{-1}$. Free energy is in kcal mol $^{-1}$. (e) Converged free energy of all the minima with respect to that of minimum Q from TASS and REMD simulations are shown together with their difference.

0.1 fs $^{-1}$ was used to maintain the CV temperature at 900 K. Before starting a TASS simulation, we carried out equilibration at 300 K for a particular

umbrella window for about 100 ps.

For the purpose of comparison, we performed about 1 μ s long replica exchange molecular dy-

namics (REMD) using AMBER 12. Four replicas at temperatures 300 K, 365 K, 440 K, and 535 K were chosen. Each replica was first equilibrated at its target temperature for 1 ns. An exchange attempt between replica was made at every 10 ps.

The free energy surface along the (ϕ_1, ϕ_2) coordinates computed from the REMD simulations is given in Figure 2d. Six major minima were obtained, labeled as **P**, **Q**, **R**, **S**, **T**, and **U**. Subsequently, we carried out TASS simulation with four collective variables as mentioned before $(\phi_1, \psi_1, \phi_2, \psi_2)$. Computed free energy barriers separating these minima as a function of simulation time are plotted in Figure 2b. The barriers systematically converge, with an error less than $0.1 \text{ kcal mol}^{-1}$, after 10 ps long simulation per umbrella window. The converged high dimensional surface is then projected to the (ϕ_1, ϕ_2) space; Figure 2c. Clearly, the positions of the minima and the saddles are very well reproduced from TASS. Moreover, the diagonal symmetry of the landscape can also be noticed, showing that the exploration of the high dimensional free energy landscape has been performed very efficiently. Similar observations were also made when the free energy surface was projected along the (ϕ_1, ψ_1) and (ϕ_2, ψ_2) ; see Figure SI1. As free energy barriers could not be accurately computed from the REMD results (due to the insufficient sampling near the saddle points), we compare the free energy difference between the minima obtained from REMD and TASS; see Figure 2e and Table SI3. After convergence, the maximum difference between the REMD and the TASS results is only $0.6 \text{ kcal mol}^{-1}$, and this difference is likely due to the insufficient sampling in REMD. These results further support that TASS can effi-

ciently explore the high dimensional free energy landscapes and provide converged free energy estimates in a computationally efficient way.

C. 1,3-Butadiene to Cyclobutene Reaction

Here we explore the broad free energy surface for the conversion of 1,3-butadiene to cyclobutene which occurs via an electrocyclic reaction (see also Figure 3a).

We have chosen the following collective variables to model this reaction: a) distance C_1-C_4 , $d[C_1 - C_4]$; b) the distance C_1-C_2 , $d[C_1 - C_2]$; c) the distance C_2-C_3 , $d[C_2 - C_3]$.

In TASS simulations, umbrella bias was applied along the $d[C_1-C_4]$ coordinate, realizing that the free energy landscape will be broad along this coordinate, and MTD bias was applied along $d[C_1 - C_2]$. Auxiliary variables corresponding to all the three coordinates were sampled using high temperature. Simulations were carried out using *ab initio* MD employing plane-wave Kohn-Sham density functional theory (DFT) as available in the CPMD program.⁴⁸ PBE exchange correlation functional⁴⁹ with ultrasoft pseudopotential⁵⁰ was used here. A cutoff of 30 Ry was used for the plane-wave expansion of wavefunctions. System was taken in a cubic supercell of side length 15 Å. Car-Parrinello⁵¹ MD at 300 K was carried out with a time step of 0.096 fs and fictitious masses of orbitals were taken as 600 a.u.

The parameter k_α was set to $1.2 \times 10^3 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and μ_α was 50.0 a.m.u. Langevin thermostat with a friction coefficient of 0.4 fs^{-1} was used to maintain the temperature of the extended degrees of freedom to 600 K.

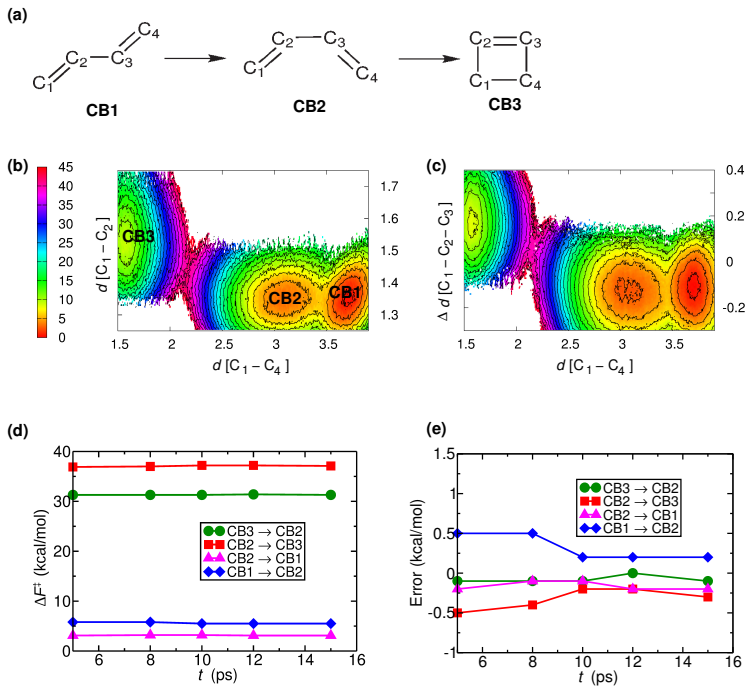


FIG. 3: (a) Structures of *trans*-1,3-butadiene (**CB1**), *cis*-1,3-butadiene (**CB2**), and cyclobutene (**CB3**); (b) Projected free energy surface computed from TASS after 15 ps of the simulation per umbrella window; (c) Converged free energy surface computed from WS-MTD; Free energy values are in kcal mol⁻¹ relative to the free energy of the minimum (**CB1**); Contour values are drawn between 1.0 and 45.0 kcal mol⁻¹ at every 2 kcal mol⁻¹ intervals; CVs are in Å; (d) Free energy barriers computed from TASS simulation as a function of simulation time per umbrella window; (e) Difference in the free energy barriers computed from TASS and WS-MTD (“Error”) as a function of simulation time per umbrella window.

In our simulations, $w_0 = 0.6$ kcal mol⁻¹ and $\delta s = 0.05$ Bohr were taken. MTD bias was updated every 19 fs. In US, the umbrella windows were placed from 1.5 Å to 3.9 Å at an interval of 0.05 Å with $\kappa_h = 4.4 \times 10^2$ kcal mol⁻¹ Å⁻². Before starting the TASS simulation, each umbrella was equilibrated for about 2 ps, and the initial structure for each umbrella window was obtained arbitrarily, as done in the case of alanine tripeptide.

We compare the results of the TASS simula-

tion with the free energy surface and the barriers computed using the WS-MTD approach from our earlier work.⁴¹ Free energy barriers converge to less than 0.5 kcal mol⁻¹ (in comparison with the WS-MTD barriers) within 10 ns per umbrella window; see Figure 3d,e. Simulation for 5 ps seems enough to compute the free energy barriers with an error less than 0.5 kcal mol⁻¹ (see Table SI4). The converged difference in the barriers of about 0.25 kcal mol⁻¹ could be ascribed to the differences in the type and the number of collective vari-

ables used in TASS and WS-MTD.

These results show that the TASS approach could efficiently sample a high dimensional free energy landscape in three collective variable space of a chemical reaction. The method seems to be as accurate as the WT-MTD, and is much efficient than the ordinary well-tempered MTD approach where free energy barriers for the same reaction was found not to converge even after 1000 ps.⁴¹

D. Tetrahedral Intermediate Formation during Hydrolysis of Aztreonam and Class-C β -Lactamase complex

To further demonstrate the application of the TASS method, we have applied this to model an enzymatic reaction in a DFT based QM/MM MD simulation. Here we model the formation of a tetrahedral intermediate during the hydrolysis of the covalent complex formed by aztreonam drug and Class-C β -Lactamase; see Figure 4. Four collective variables were chosen for simulating this hydrolysis reaction (see Figure 4 for labeling): a) coordination number of $\text{Tyr}_{150}\text{O}_\eta$ to hydrogens of W_1 , $C[\text{Tyr}_{150}\text{O}_\eta - W_1\text{H}]$; b) distance between AztC2 and $W_1\text{O}$, $d[\text{AztC2} - W_1\text{O}]$; c) the distance $\text{Tyr}_{150}\text{O}_\eta$ to $\text{Lys}_{67}\text{N}_\zeta$, $d[\text{Tyr}_{150}\text{O}_\eta - \text{Lys}_{67}\text{N}_\zeta]$; d) the distance $\text{Tyr}_{150}\text{O}_\eta$ to $\text{Lys}_{315}\text{N}_\zeta$, $d[\text{Tyr}_{150}\text{O}_\eta - \text{Lys}_{315}\text{N}_\zeta]$. Here

$$C[\text{Tyr}_{150}\text{O}_\eta - W_1\text{H}] = \sum_{J \in W_1\text{H}} \frac{1}{\left(1 + \left(\frac{d_J}{d_0}\right)^6\right)}$$

where $d_J \equiv d[\text{Tyr}_{150}\text{O}_\eta - W_1\text{H}]$ and $d_0 = 1.3 \text{ \AA}$. The auxiliary variables corresponding to all the four collective variables were sampled using high temperature (1000 K), while the physical system was sampled at 300 K. Here $C[\text{Tyr}_{150}\text{O}_\eta - W_1\text{H}]$

was chosen as a collective variable to accelerate proton transfer from water to $\text{Tyr}_{150}\text{O}_\eta$ and MTD bias was applied along this collective variable. To enhance the nucleophilic attack of OH^- on the carbonyl carbon of the drug molecule $d[\text{AztC2} - W_1\text{O}]$ coordinate was chosen as a collective variable which was sampled using the US bias. Along this coordinate we expect that the free energy landscape is broad and thus is more efficient to sample using US than MTD or TAMD alone. The collective variables $d[\text{Tyr}_{150}\text{O}_\eta - \text{Lys}_{67}\text{N}_\zeta]$ and $d[\text{Tyr}_{150}\text{O}_\eta - \text{Lys}_{315}\text{N}_\zeta]$ were considered for sampling different conformations of Tyr_{150} , Lys_{67} , and Lys_{315} . We treated these coordinates as in TAMD/d-AFED as we anticipated that the free energy surface along these coordinates has small barriers and the surface is not flat and unbound.

The hybrid QM/MM simulations were performed using the CPMD/GROMOS interface⁵² as implemented in the CPMD package. Aztreonam drug, side chains of Lys_{67} , Tyr_{150} , Ser_{64} , Lys_{315} , Thr_{316} ; backbone of Lys_{315} , Thr_{316} , Gly_{317} , and two water molecules near the active site were treated quantum mechanically. Rest of the protein and the solvent molecules were treated by molecular mechanics (MM). For more computational details, see SI Section I.

Constant temperature Car-Parrinello⁵¹ MD at 300 K was carried out using the Nosè-Hoover chain thermostats for the nuclei and orbital degrees of freedom.⁵³ A time step of 0.14 fs was used to integrate the equations of motion and the fictitious masses of orbitals were taken as 600 a.u. We assigned $k_\alpha = 1.2 \times 10^3 \text{ kcal mol}^{-1}$ and $\mu_\alpha = 50.0 \text{ a.m.u.}$ for all the auxiliary variables. An overdamped Langevin thermostat with a friction

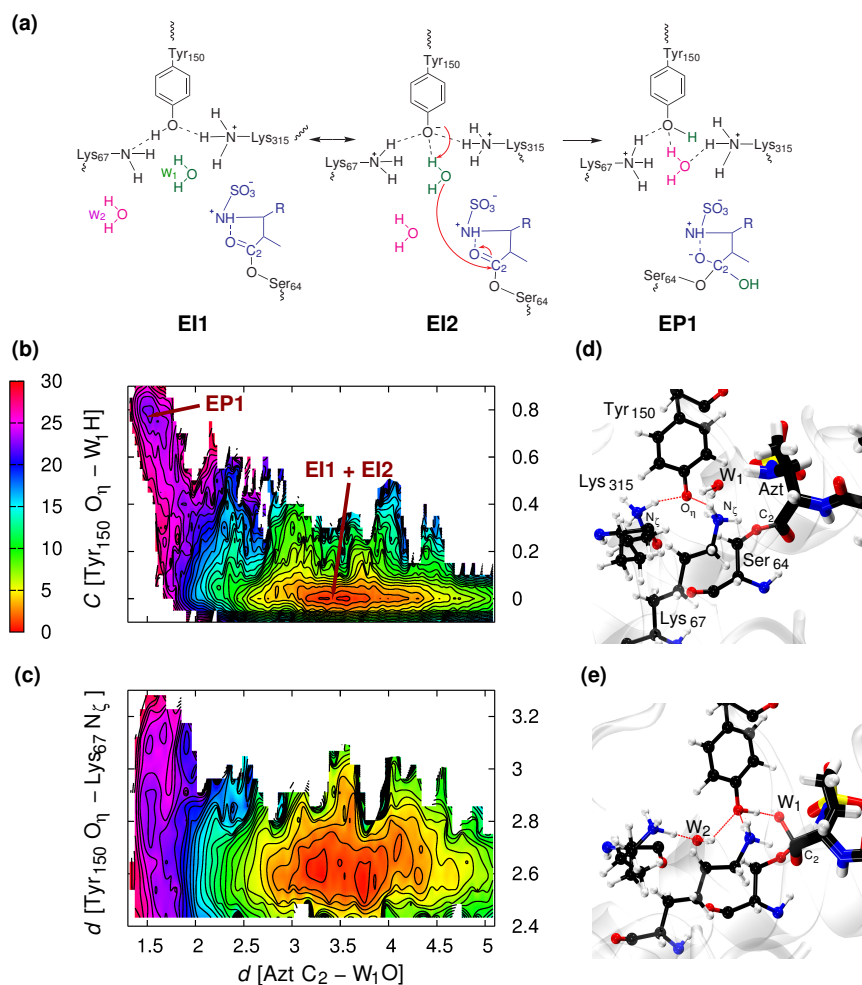


FIG. 4: (a) Mechanism of formation of EP1 from the enzyme–drug covalent complex EI formed by aztreonam (blue color) and Class C β -lactamase. Here W₁ molecule is activated by Tyr₁₅₀ and the former attacks C₂ resulting in EP1; (b) and (c) are the two different projections of the five dimensional free energy landscape; (d) and (e) show snapshots of EI2 and EP1 from the QM/MM trajectory; atom colors: S (yellow), O (red), N (blue), C (black), H (white) ; protein backbone is represented as transparent ribbons.

coefficient of 0.4 fs^{-1} was used to maintain temperature of the auxiliary variables at 1000 K. In US, windows were placed from 1.3 \AA to 5.1 \AA at an interval of 0.1 \AA with $\kappa_h = 4.5 \times 10^2 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. The MTD parameters were $w_0 = 0.6 \text{ kcal mol}^{-1}$, $\delta s = 0.05 \text{ Bohr}$ and $\Delta T = 2000 \text{ K}$ were taken. MTD

bias was updated every 19 fs.

Before starting the TASS simulations, each umbrella window was equilibrated for about 4 ps. Initial structure for an umbrella window was taken from the adjacent equilibrated window. The whole protein, including the QM part, and the solvent

molecules were free to move during the MD simulations.

We could successfully simulate the reaction **EI1**→**EP1** using the TASS method, and the converged reconstructed free energy surface is given in Figure 4b,c. Unlike in the previous benchmark cases, we have used varying simulation lengths (4-8 ps each) for different umbrella windows till a convergence in the free energy barrier was achieved. In the reactant basin, we have noticed proton transfer between Tyr₁₅₀ and Lys₆₇, i.e. **EI1**↔**EI2**. The hydrogen bonding interactions between the two residues were maintained throughout the reaction. However, distance between the Tyr₁₅₀O_η and Lys₆₇N_ζ increases as proton transfer occurs from W₁ to Tyr₁₅₀O_η; see Figure 4c. On the other hand, the hydrogen bonding interaction between Lys₃₁₅ and Tyr₁₅₀ was broken in the initial stages of the chemical reaction, as a result of which, another water molecule (W2) moved into the active site, and Lys₃₁₅ formed interactions with Glu₂₇₂.

The free energy barrier for the reaction was computed from the projected free energy surface on $d[\text{Tyr}_{150}\text{O}_\eta - \text{Lys}_{67}\text{N}_\zeta]$ and $C[\text{Tyr}_{150}\text{O}_\eta - \text{W}_1\text{H}]$ coordinates, and is 24.5 kcal mol⁻¹. From experimental studies⁵⁴ it is known that aztreonam is a slowly hydrolyzing drug and from the measured rate constants for deacylation, we estimate the corresponding free energy barrier as 23 kcal mol⁻¹ (using the transition state theory). This agrees well with our computed free energy barrier of 24.5 kcal mol⁻¹.

The same reaction was failed to simulate in an ordinary MTD run, likely due to the broad nature of the basin along the $d[\text{Tyr}_{150}\text{O}_\eta - \text{Lys}_{67}\text{N}_\zeta]$ coor-

inate (see also Figure 4)c, and thus large computational time would be required to build the sufficient bias potential in the relevant parts of the free energy surface. Moreover, W₁ water molecule was also driven out of the active site, which was then replaced by an MM water molecule from the bulk (data not shown). Both these difficulties are overcome in the TASS simulation since US was carried out along the $d[\text{Tyr}_{150}\text{O}_\eta - \text{Lys}_{67}\text{N}_\zeta]$ coordinate. Additionally, sampling of different conformations of the active site residues needed four collective variables which is also practically difficult to sample properly with conventional MTD, especially in DFT based QM/MM simulations.

IV. CONCLUSIONS

In this paper, we presented a method called TASS that combines MTD, US, and TAMD/d-AFED to sample large number of collective variables and explore high dimensional free energy landscapes. Free energy estimates using TASS is shown to converge systematically to the exact values. We have demonstrated the efficiency of TASS in sampling four and five dimensional free energy landscapes with multiple minima along different coordinates. Moreover, the method is also shown to be practically usable for computing free energy surfaces of chemical reactions in *ab initio* and DFT based hybrid QM/MM MD simulations.

The method is well suited for exploring free energy surfaces that are broad and unbound, where conventional enhanced sampling approaches such as MTD and TAMD/d-AFED become inefficient. The coordinate along which the free energy is flat, broad or unbound is cho-

sen for US; for instance for a A+B type reaction, the distance between the reacting species is taken in US. The chosen MTD collective variables are along which we expect large free energy barriers while the free energy surface is bound and not flat. However, all the collective variables are kept at high temperature, as in TAMD/d-AFED. The collective variables other than the biased ones are maintained at high temperature, and we prefer to choose these collective variables along which small barriers are present or the surface is rugged, for e.g. sampling hydrogen bond breaking and formation, and changes in torsions.

Controlled exploration of free energy surfaces, for instance along certain reaction pathways, can thus be achieved in TASS, by an appropriate choice of the US coordinate. The method permits one to add and remove collective variables in different umbrella windows, giving flexibility and computational efficiency in exploring complex high dimensional free energy landscapes. Although, we have used US and MTD biases, the method can be straightforwardly extended to the cases where MTD bias is not required (either in selected or for all the umbrella windows), by setting $A(\tau) = 1$ in Equation (13). Replica exchange based algorithms can also be combined with TASS to further improve the sampling efficiency. The TASS Hamiltonian in Equation (12) can be realized effortlessly in simulations using MD plugins like PLUMED⁴⁶ which has been interfaced with several popular MM and QM programs.

SUPPLEMENTARY INFORMATION

Supplementary information contains parameters for three-dimensional model potential, computational details, and details of free energy profiles for different simulations.

ACKNOWLEDGMENTS

Authors thank IIT Kanpur for availing the HPC facility. SA thanks UGC for Ph. D fellowship.

¹M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, 1st ed. (Oxford University Press, Oxford, 2010).

²T. Lelièvre, M. Rousset, and G. Stoltz, *Free Energy Computations: A Mathematical Perspective* (Imperial College Press, London, 2010).

³E. Vanden-Eijnden, *J. Comput. Chem.* **30**, 1737 (2009).

⁴C. D. Christ, A. E. Mark, and W. F. van Gunsteren, *J. Comput. Chem.* **31**, 1569 (2010).

⁵C. Abrams and G. Bussi, *Entropy* **16**, 163 (2014).

⁶T. Morishita, S. G. Itoh, H. Okumura, and M. Mikami, *Phys. Rev. E* **85**, 066702 (2012).

⁷T. Lelièvre, M. Rousset, and G. Stoltz, *J. Chem. Phys.* **126**, 134111 (2007).

⁸R. D. Miller, *Annu. Rev. Phys. Chem.* **65**, 583 (2014).

⁹O. Valsson, P. Tiwary, and M. Parrinello, *Annu. Rev. Phys. Chem.* **67**, 159 (2016).

¹⁰A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci.* **99**, 12562 (2002).

¹¹M. Iannuzzi, A. Laio, and M. Parrinello, *Phys. Rev. Lett.* **90**, 238302 (2003).

¹²A. Barducci, M. Bonomi, and M. Parrinello, *WIREs Comput. Mol. Sci.* **1**, 826 (2011).

¹³L. Sutto, S. Marsili, and F. L. Gervasio, *WIREs: Comput. Mol. Sci.* **2**, 771 (2012).

¹⁴A. Laio and F. L. Gervasio, *Rep. Prog. Phys.* **71**, 126601 (2008).

¹⁵G. M. Torrie and J. P. Valleau, *Chem. Phys. Lett.* **28**, 578 (1974).

¹⁶J. Kästner, *WIREs Comput. Mol. Sci.* **1**, 932 (2011).

¹⁷T. Huber, A. E. Torda, and W. F. van Gunsteren, *J. Comput. Aided Mol. Des.* **8**, 695 (1994).

¹⁸H. Grubmüller, *Phys. Rev. E* **52**, 2893 (1995).

- ¹⁹F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
- ²⁰E. Darve and A. Pohorille, *J. Chem. Phys.* **115**, 9169 (2001).
- ²¹U. H. E. Hansmann and L. T. Wille, *Phys. Rev. Lett.* **88**, 068105 (2002).
- ²²M. Yang, L. Yang, Y. Gao, and H. Hu, *J. Chem. Phys.* **141**, 044108 (2014).
- ²³J. Comer, J. C. Gumbart, J. Hénin, T. Lelièvre, A. Pohorille, and C. Chipot, *J. Phys. Chem. B* **119**, 1129 (2015).
- ²⁴Y. I. Yang, J. Zhang, X. Che, L. Yang, and Y. Q. Gao, *J. Chem. Phys.* **144**, 094105 (2016).
- ²⁵A. Barducci, G. Bussi, and M. Parrinello, *Phys. Rev. Lett.* **100**, 020603 (2008).
- ²⁶J. F. Dama, M. Parrinello, and G. A. Voth, *Phys. Rev. Lett.* **112**, 240602 (2014).
- ²⁷G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello, *J. Am. Chem. Soc.* **128**, 13435 (2006).
- ²⁸S. Piana and A. Laio, *J. Phys. Chem. B* **111**, 4553 (2007).
- ²⁹F. Marinelli, F. Pietrucci, A. Laio, and S. Piana, *PLoS Comput. Biol.* **5**, e1000452 (2009).
- ³⁰A. Gil-Ley and G. Bussi, *J. Chem. Theory Comput.* **11**, 1077 (2015).
- ³¹J. Pfaendtner and M. Bonomi, *J. Chem. Theory Comput.* **11**, 5062 (2015).
- ³²P. Shaffer, O. Valsson, and M. Parrinello, *Proc. Natl. Acad. Sci.* **113**, 1150 (2016).
- ³³A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).
- ³⁴S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, *J. Comput. Chem.* **13**, 1011 (1992).
- ³⁵L. Maragliano and E. Vanden-Eijnden, *Chem. Phys. Lett.* **426**, 168 (2006).
- ³⁶J. B. Abrams and M. E. Tuckerman, *J. Phys. Chem. B* **112**, 15742 (2008).
- ³⁷M. A. Cuendet and M. E. Tuckerman, *J. Chem. Theory Comput.* **10**, 2975 (2014).
- ³⁸M. Chen, M. A. Cuendet, and M. E. Tuckerman, *J. Chem. Phys.* **137**, 024102 (2012).
- ³⁹M. Chen, T.-Q. Yu, and M. E. Tuckerman, *Proc. Natl. Acad. Sci.* **112**, 3235 (2015).
- ⁴⁰L. Xie, L. Shen, Z.-N. Chen, and M. Yang, *J. Chem. Phys.* **146**, 024103 (2017).
- ⁴¹S. Awasthi, V. Kapil, and N. N. Nair, *J. Comput. Chem.* **37**, 1413 (2016).
- ⁴²J. M. Johnston, H. Wang, D. Provasi, and M. Filizola, *PLoS Comput. Biol.* **8**, e1002649 (2012).
- ⁴³M. Bonomi, A. Barducci, and M. Parrinello, *J. Comput. Chem.* **30**, 1615 (2009).
- ⁴⁴P. Tiwary and M. Parrinello, *J. Phys. Chem. B* **119**, 736 (2014).
- ⁴⁵J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, *J. Chem. Theory Comput.* **11**, 3696 (2015).
- ⁴⁶M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broaglia, and M. Parrinello, *Comput. Phys. Commun.* **180**, 1961 (2009).
- ⁴⁷D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman, *AMBER 12* (University of California, San Francisco, 2012).
- ⁴⁸Version 13.2, *CPMD Program Package*, IBM Corp 1990-2011, MPI für Festkörperforschung Stuttgart 1997-2001.
- ⁴⁹J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, *Phys. Rev. B* **46**, 6671 (1992).
- ⁵⁰D. Vanderbilt, *Phys. Rev. B* **41**, 7892 (1990).
- ⁵¹R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985).
- ⁵²A. Laio, J. VandeVondele, and U. Rothlisberger, *J. Chem. Phys.* **116**, 6941 (2002).
- ⁵³G. J. Martyna, M. L. Klein, and M. Tuckermann, *J. Chem. Phys.* **97**, 2635 (1992).
- ⁵⁴D. Monnaie, R. Virden, and J. M. Frère, *FEBS Letters* **306**, 108 (1992).