Published in final edited form as:

J Chem Theory Comput. 2023 December 12; 19(23): 8886–8900. doi:10.1021/acs.jctc.3c00526.

tICA-Metadynamics for Identifying Slow Dynamics in Membrane Permeation

Myongin Oh,
Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States

Gabriel C. A. da Hora,
Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States

Jessica M. J. Swanson
Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States

Abstract

Molecular simulations are commonly used to understand the mechanism of membrane permeation of small molecules, particularly for biomedical and pharmaceutical applications. However, despite significant advances in computing power and algorithms, calculating an accurate permeation free energy profile remains elusive for many drug molecules because it can require identifying the rate-limiting degrees of freedom (i.e., appropriate reaction coordinates). To resolve this issue, researchers have developed machine learning approaches to identify slow system dynamics. In this work, we apply time-lagged independent component analysis (tICA), an unsupervised dimensionality reduction algorithm, to molecular dynamics simulations with well-tempered metadynamics to find the slowest collective degrees of freedom of the permeation process of trimethoprim through a multicomponent membrane. We show that tICA-metadynamics yields translational and orientational collective variables (CVs) that increase convergence efficiency ~1.5 times. However, crossing the periodic boundary is shown to introduce artifacts in the translational CV that can be corrected by taking absolute values of molecular features. Additionally, we find that the convergence of the tICA CVs is reached with approximately five membrane crossings and that data reweighting is required to avoid deviations in the translational CV.

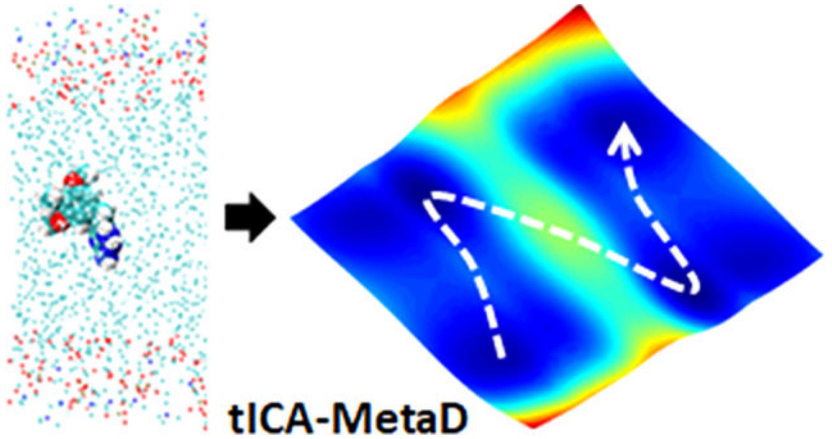
Corresponding Author: Jessica M. J. Swanson – Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States; j.swanson@utah.edu.

Author Contributions
M.O. and J.M.J.S. designed the research. G.C.A.d.H. and M.O. carried out the simulations and analyses. All authors interpreted the results and wrote the manuscript.

The authors declare no competing financial interest.

ASSOCIATED CONTENT
Supporting Information
The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.3c00526.
Linear relationship between the orientational CV and the orientation angle of trimethoprim; comparisons of eigenvalues and autocorrelations; time evolution of the location of trimethoprim when suboptimal and optimal CVs are biased; analysis of the 1-μs-long TTMetaD simulation; tICA results from the WTMetaD simulations with the PBCs and with the absolute values of the molecular features; average 2D PMF along the tICA CVs in the presence of the harmonic walls; definitions and examples of subtrajectories; molecular structures of trimethoprim and POPC; probability distributions of the orientation angle of trimethoprim close to and inside the membrane; analysis of the number of contacts and interaction energy between trimethoprim and lipids; and schematic diagram that summarizes the key steps of the tICA-MetaD protocol (PDF)

Graphical Abstract



INTRODUCTION

Lipid membranes function as the physical barrier that controls the exchange of matter, energy, and information between cells and organelles in biological systems. Membrane permeation of small molecules is often an activated process that takes place on time scales inaccessible by standard molecular simulations. The dynamics of large biomolecular systems is governed by a complex high-dimensional free energy landscape characterized by a hierarchy of energy barriers.[1–3] Transitions between metastable states become rare when they are separated by large energy barriers while thermal fluctuations are the only driving force for barrier crossing.[4] These kinetic bottlenecks restrict the time scale that can be explored (which is generally in the range of microseconds or shorter) by conventional molecular dynamics (MD) simulations, thereby introducing substantial statistical errors in the measurement of structural, thermodynamic, and kinetic properties. To circumvent this sampling issue, many of the existing computational methods used to evaluate membrane permeation rely on enhanced free energy sampling techniques.[5–9] Impressively, membrane permeation has been directly observed for a few small molecules in canonical MD simulations on the nano- and microsecond time scales. For instance, Kramer̈ et al.[10] performed unbiased MD simulations to evaluate the permeability coefficients of oxygen, water, and ethanol using counting methods and maximum likelihood estimation for the inhomogeneous solubility-diffusion (ISD) model.[6,11,12] They found that counting methods[13] yield nearly model-free estimates for all of the three permeants, whereas the ISD model causes large uncertainties for water due to insufficient sampling and overestimates for ethanol due to collective effects in the membrane.[10,14] For larger molecules with slower permeation, however, enhanced sampling can be essential to increase the occurrence of the slowest dynamic motions that enable permeation. This can be done by biasing the potential energy surface or altering the probability density of sampled conformations.[15,16] For example, an external bias potential can be added to the Hamiltonian (as in umbrella sampling and metadynamics (MetaD)), or the system can be coupled to higher temperatures (as in replica exchange MD) to effectively reduce energy barriers and thus sample transition regions, or the transition ensemble can be selectively sampled with path sampling approaches, such as transition interface sampling or its combination with replica exchange with or without memory effects.[14,17]

The success of enhanced free energy sampling methods involving Hamiltonian bias requires the selection of a proper set of collective variables (CVs). CVs are user-defined functions of atomic (Cartesian) coordinates that provide a low-dimensional projection of conformational phase space while, in principle, retaining “important” information. Dimensionality reduction is an essential consequence of our inability to work in or visualize high-dimensional spaces. Only in a reduced number of dimensions (typically 2–4) can we define effective bias potentials to alter dynamics, visualize complex free energy landscapes, efficiently sample conformational phase space that crosses high energy barriers, and simplify simulation data for noise omission and better inference (i.e., escape the curse of dimensionality).[18] Ideally, CVs are translationally and rotationally invariant, include all relevant slow molecular motions, and distinguish local minima (metastable states) and activation barriers (transition states).[19–22] If properly obtained, CVs provide lower variance estimators of the properties of interest when the system diffuses on the free energy surface (or potential of mean force, PMF) spanned by these CVs. However, it is highly nontrivial to intuit good CVs for complex systems (large biomolecules in particular), and accelerating dynamically irrelevant CVs may give rise to inaccurate profiles and decreased efficiency compared to unbiased MD.[21] In recent years, the availability of large data sets (obtained directly from unbiased and biased simulations) in conjunction with advances in computing hardware and algorithms has led to the automated design of CVs inspired by machine learning, data science, and information theory. A comprehensive review of machine learning approaches for CV discovery is presented in refs 4,18,21,23, and 24. Herein, we aim to test the limitations of one such approach.

Data-driven CVs are typically coincident with collective degrees of freedom that either have a high variance or evolve slowly. High-variance CVs can be identified by the principal component analysis (PCA), an unsupervised linear transformation that finds a subspace that maximally preserves the configurational variance contained within a molecular simulation trajectory.[21] Since the orthogonal eigenvectors, or the principal components (PCs), represent large-amplitude collective motions in terms of variance, they are often called essential dynamics[25–30] and used as collective variables for enhanced sampling. However, a major problem inherent in PCA is that it is not generally guaranteed that large-amplitude motions are associated with slow motions that enable transitions between metastable states. For example, Naritomi and Fuchigami[31] found that a closure motion of the lysine–arginine–ornithine binding protein described by the largest-amplitude mode determined by PCA does not represent the slowest mode, a twist motion that takes place on a time scale of tens of nanoseconds.

Generally, identifying slow motions rather than large-amplitude motions is essential to address the sampling problem. Time-lagged independent component analysis (tICA), initially introduced as a signal decomposition algorithm,[32] is an unsupervised linear transformation that finds a subspace that maximally preserves the kinetic content (i.e., minimizes the loss of kinetic information) by maximizing the autocorrelation function.[21,33–35] The resulting eigenvectors, or independent components (ICs), represent the slowestrelaxing degrees of freedom in time series data and approximate the eigenfunctions of the underlying Markovian dynamics. In other words, tICA provides the optimal linear approximation to the variational approach to conformational dynamics, a systematic approach for modeling the slow parts of Markov processes by approximating the dominant eigenfunctions of the MD propagator (or transfer operator).[36–39] Concisely, tICA is a special case of the linear variational approach which uses mean-free input descriptors as a basis set and empirical estimates of their covariance matrices in an eigenvalue problem.[33,36] More details on tICA can be found in the Methods section.

Once slowly decorrelating modes are discovered, they can be biased in CV-based enhanced sampling techniques to accelerate the occurrence of rare events. In tICA-MetaD, tICA is performed on MD simulation trajectories to identify the ICs which are then directly used as CVs in MetaD to obtain highly diffusive behavior in CV space and fast convergence of PMF calculations. For instance, Sultan and Pande[40] applied linear and nonlinear tICA-MetaD on unbiased MD simulations of alanine dipeptide and bovine pancreatic trypsin inhibitor to explicitly sample their slowest modes. In principle, the combined method can drive slow transitions even when no such transitions take place in the original unbiased simulations. However, this only happens when the slow motions captured in the original simulations are the same as those that enable the sought-after transition(s). This can be a major limitation of their method since it requires unbiased sampling of the relevant slow transitions (e.g., via very long aggregate MD sampling or enough MD runs starting from high and low free energy states), which is often challenging or even unfeasible depending on the system.[40] A clear example of this would be the permeation of a highly polar molecule for which unbiased sampling never entered the hydrophobic membrane midplane. To resolve this problem, McCarty and Parrinello[41] proposed to start with a biased simulation with suboptimal CVs, reweight the trajectory to recover unbiased distributions (i.e., Boltzmann statistics) of the system, and perform tICA to derive slow CVs for MetaD. This CV optimization method has been applied to different systems ranging from conformational transitions of alanine dipeptide, alanine tetrapeptide,[41] and L99A T4 lysozyme[42] to homogeneous crystallization of sodium and aluminum.[43] However, their method also faces several challenges. First, data reweighting is not a trivial task; for example, their reweighting scheme may not work properly on the trajectory generated by transition-tempered metadynamics (TTMetaD).[8,44] Second, reweighting does not necessarily converge accurately and rapidly to the underlying free energy depending on the reweighting technique, the enhanced sampling technique, and the stage of the simulation.[45]

Recovering the slow modes of an unbiased system from a biased trajectory is nontrivial and thus often approximate. Extensive efforts have been made to mitigate this challenge. For example, Mehdi et al.[46] employed the State Predictive Information Bottleneck (SPIB) approach to investigate the permeation of a small benzoic acid (BA) molecule across a symmetric phospholipid bilayer, specifically aiming to identify reaction coordinates (RCs) for enhanced sampling algorithms. Bonati et al.[47] developed the deep-tICA method by introducing a nonlinear variant of VAC to the approach initially proposed by McCarty and Parrinello[41] and utilizing on-the-fly probability-enhanced sampling (OPES) to construct the bias potential. Chen and Chipot[48] conducted an in-depth investigation into the use of classical autoencoders (AEs), time-lagged AEs (TAEs), modified TAEs, VAMPnets, and state-free reversible VAMPnets (SRVs) for deep learning-based CV discovery in molecular processes. Donati and Keller[49] introduced a methodology that applies the Girsanov reweighting scheme to metadynamics simulations, allowing for the recovery of accurate dynamic properties of a molecular system from an enhanced sampling simulation. Another noteworthy approach is the Girsanov Reweighting Enhanced Sampling Technique (GREST) proposed by Shmilovich and Ferguson,[50] which is an adaptive sampling scheme that alternates rounds of data-driven slow CV discovery and enhanced sampling along these coordinates.

Previously, we studied the molecular mechanism of membrane permeation of trimethoprim, an antibacterial agent primarily used in the treatment of urinary tract infections,[51] using an alternative implementation of tICA-MetaD.[52] First, we performed a short TTMetaD simulation with nonoptimal CVs to generate an initial trajectory that involved at least one transition of interest (permeation). Second, we identified the slowest decay modes using tICA without data reweighting. Lastly, we carried out a new TTMetaD simulation using a permeation CV consistent with the original bias and the slow CV obtained from tICA. We then examined the accuracy and convergence of PMF calculations. Interestingly, the use of tICA CVs was shown to accelerate the convergence of PMF calculations while also revealing a subtle effect of cholesterol on the permeation of the small drug molecule through the heterogeneous model membrane. Perhaps more importantly, this work brought to light several outstanding questions regarding the optimal use of tICA-MetaD. For example, it remains unclear what effect trajectory reweighting has on tICA CVs, whether this approach is applicable to other forms of tempering, and how many rare events are needed in the original trajectories in order to obtain consistent (converged) tICA CVs. Moreover, the influence of the periodic boundary conditions (PBCs) on tICA-like analyses of permeation has, to the best of our knowledge, been largely overlooked.

In this work, we apply several variations of tICA-MetaD to membrane permeation to address these outstanding questions. We opt for a pair of molecular features, specifically the z-positions of trimethoprim relative to the lipid bilayer (or signed perpendicular distances), and design a straightforward yet suboptimal initial CV. This approach aims to initiate rare-event crossings to obtain trajectory data that can be fed into tICA-MetaD protocols in order to identify optimal CVs. The use of a simple and largely transferable initial CV (such as the z-position relative to the membrane) is particularly relevant when the selection of good CVs is not known a priori. We find that tICA-MetaD distinctly identifies the translational and orientational modes as slow CVs, and the use of the machine-learned tICA CVs leads to ~1.5 times faster convergence of our PMF calculations. However, the tICA-MetaD scheme that we employed cannot be applied to TTMetaD trajectories. Also, PBCs can be detrimental to tICA for membrane permeation trajectories. The orientational CV is not significantly affected by the PBCs, while taking absolute values of molecular features corrects the error in the translational CV. Additionally, convergence of the tICA CVs is attained even when only five membrane crossings are included in the trajectory regardless of data reweighting, and the first eigenvector may deviate significantly from the translational CV when data reweighting is omitted. Collectively, we hope that these findings will help inform future implementation best practices for tICA-MetaD applied to membrane permeation.

METHODS

All-Atom Molecular Dynamics Simulation.

We performed all-atom molecular dynamics simulations using GROMACS 2019.4[53] patched with PLUMED 2.5.3[54] for well-tempered metadynamics (WTMetaD) and TTMetaD. For optimal comparison, the heterogeneous bilayer system used in our previous work selected for this study, including 36 phosphatidylcholine (POPC) and 4 cholesterol molecules ($n_{\mathrm{POPC}}/n_{\mathrm{CHOL}} = 9:1$). The bilayer was placed on the xy-plane (i.e., with the surface normal along the z-axis) and solvated by 2200 water molecules in a unit cell with dimensions of approximately 3.6 × 3.6 × 9.5 nm^3. The PBC was then applied in every direction. CHARMM36[55] and the CHARMM general force field (CGenFF)[56] were employed to model the lipid molecules and trimethoprim, respectively. The water molecules were modeled with the TIP3P[57] potential. We prepared the initial structure of the lipid bilayer using the CHARMM-GUI membrane builder[58] and equilibrated it for 150 ns in water. After equilibration, we placed the drug molecule randomly into the aqueous region using PACKMOL.[59] To generate an isobaric–isothermal (NPT) ensemble of the system at 323.15 K and under 1 bar, the lipid and the water plus the drug molecule were separately coupled to two heat baths using velocity rescaling with a stochastic term,[60] while the system was coupled semi-isotropically to a Berendsen barostat[61] such that the simulation box was rescaled every 5 ps. The cutoff distance for the short-range neighbor list was set to 12 Å, and the neighbor list was updated every 40 steps. Fast smooth Particle-Mesh Ewald (SPME)[62] was used to treat long-range electrostatic interactions. All covalent bonds including hydrogen atoms were constrained by linear constraint solver (LINCS),[63] and the integration time step was set to be 2 fs. The initial velocities were randomly sampled from a Maxwell–Boltzmann distribution at 323 K which is well above the transition temperature of the lipids. We generated the images from the simulations and analyzed the MD trajectories using Visual Molecular Dynamics (VMD) version 1.9.3.[64] Details of each simulation are given in Table 1.

To investigate the effect of the PBC on tICA CVs, we applied lower and upper walls that limit the phase space accessible to the system during the simulation. In PLUMED 2.5.3, these walls are implemented in the form of restraining potentials $\eta(s)$ given by

$$\eta(s) = \sum_i \kappa_i \left(\frac{s_i - a_i + o_i}{r_i} \right)^e \tag{1}$$

where κ_i, s_i, a_i, o_i, and r_i denote a force constant, a CV, the position of the wall, an offset, and a rescaling factor, respectively, and e is the exponent that determines the power law. In our simulations, set $o_i = 0$, $r_i = 1$, and $e = 2$ to realize harmonic restraints and added both upper and lower walls on z_1 and z_2 (i.e., suboptimal CVs) at 3.6 and −3.6 nm with $\kappa_i = 1500 \ \mathrm{kJ/mol}$.

Well-Tempered Metadynamics.

WTMetaD[19,65,66] is an enhanced sampling technique that discourages the system from revisiting configurations that have previously been sampled in the CV space by periodically adding to the Hamiltonian of the system small repulsive Gaussians whose amplitude decreases exponentially as the simulation progresses. The sampling of rare events is accelerated due to the ergodic and diffusive motion of the system along the selected CVs. The instantaneous bias potential $V(\boldsymbol{s},\ t)$ at point $\boldsymbol{s}$ in CV space and at time t is determined by the sum of Gaussian hills deposited over the past trajectory of the system

$$V(s,t) = \sum_{t' = n\tau_G}^{t' < t} w(t') \prod_{i=1}^{N_{\mathrm{CV}}} \exp\left(-\frac{[s_i(t) - s_i(t')]^2}{2\sigma_i^2}\right) \tag{2}$$

where σ_i is the width of the Gaussian hill for s_i, the i th CV; N_{CV} is the number of CVs; τ_{G} is the stride of Gaussian deposition; and w is the adaptive height of the Gaussian hill given by

$$w(t') = w_0 \exp\left(-\frac{V(\boldsymbol{s},\ t')}{k_{\mathrm{B}}\Delta \boldsymbol{T}}\right) \tag{3}$$

where w_0 is the initial height of the Gaussian hill, k_{B} is the Boltzmann constant, and $\Delta \boldsymbol{T}$ is a parameter that tunes how quickly the Gaussian height is reduced. Ideally, $\Delta \boldsymbol{T}$ should be proportional to the free energy barrier to be crossed, which is generally not known in advance. If the parameter is too small, then the system fails to escape local minima; if the parameter is too large, unphysical instability can be introduced causing significant errors in the PMF. The adjustment of the Gaussian deposition allows smooth convergence and tunable error of the bias potential so that the unbiased free energy $F(\boldsymbol{s})$ of the system can be estimated as the limit

$$\lim_{t \to \infty} V(\boldsymbol{s}, t) = -\left(1 - \frac{1}{\gamma}\right) F(\boldsymbol{s}) \tag{4}$$

where γ is a bias factor at temperature $\boldsymbol{T}$ given by

$$\gamma = \frac{\boldsymbol{T} \ + \ \Delta \boldsymbol{T}}{\boldsymbol{T}}$$

An advantage of WTMetaD is that after a transient time, the simulation enters a quasi-stationary limit in which the expectation value of any observable $\langle O(\boldsymbol{R})\rangle$ can be estimated as a running average according to

$$\langle O(\boldsymbol{R})\rangle = \lim_{\tau \to \infty} \frac{\int_0^\tau O(\boldsymbol{R}(t))\mathrm{e}^{\beta[V(s,t) - c(t)]}\,\mathrm{d}t}{\int_0^\tau \mathrm{e}^{\beta[V(s,t) - c(t)]}\,\mathrm{d}t} \tag{6}$$

where $\boldsymbol{R}(t)$ is atomic position at time t, β is the reciprocal of the product of k_B and $\boldsymbol{T}$, and $c(t)$ is a time-dependent bias offset defined as

$$c(t) = -\frac{1}{\beta}\log\frac{\int \mathrm{e}^{-\beta[F(\boldsymbol{s}) + V(\boldsymbol{s},t)]}\mathrm{d}\boldsymbol{s}}{\int \mathrm{e}^{-\beta F(\boldsymbol{s})}\mathrm{d}\boldsymbol{s}} \tag{7}$$

The function $c(t)$ asymptotically tends toward the reversible work done on the system by the external bias.

In our WTMetaD simulations, the Gaussian bias was deposited every 500 steps with a height of 0.05 kJ/mol and a width of 0.2 nm. The height of the Gaussian hill was tempered with a bias factor of 15. Multiple replicas were prepared, initialized randomly, and run for 3–5 μs, and the resulting PMFs were obtained by averaging over replicas. To generate 1D PMFs, we further symmetrized the free energy profiles with respect to the center of the horizontal axis. The minimum free energy paths (MFEPs) on 2D PMFs were calculated with a zero-temperature string method[67,68] which represent the most probable transition paths in the ensemble of the permeation processes. The 1D PMF was directly obtained by using the average MFEP from independent replicas. Error bars were computed using the standard errors across replicas.

CV Selection.

In the tICA-MetaD process, an initial suboptimal CV (e.g., the signed perpendicular distance of the molecule's center of mass (COM) from the membrane) is first used to induce rare events (i.e., membrane crossings), and subsequently, optimal CVs are derived as linear combinations of sets of features. Typical features might include multiple interatomic distances or dihedral angles. The primary aim of tICA is to linearly combine these features into a few CVs that efficiently sample the system's slow dynamics when biased. Generally, many features could be tested, and those with significant weights are retained. We, however, used two features in order to compare with previous work that used the two distances between the z-position of the COM of each ring of trimethoprim and that of the lipid membrane COM.[69,70] These features are denoted by z_1 for the trimethoxybenzyl (TMB) group and z_2 for the diaminopyrimidine (DAP) group (Figure 1c). Although the previous work demonstrated that biasing z_1 and z_2 reduced the total simulation time required for convergence and captured orientational changes, it is not guaranteed that the two CVs are associated with slow transitions or optimal compared with many other degrees of freedom that mediate the permeation process. In addition, we chose a naïve linear combination of z_1 and z_2 as initial suboptimal CV to bias just one CV (mimicking the anticipated common use of the molecular COM distance) and to test tICA's ability to identify a superior linear combination from a naïve linear combination. Ultimately biasing the tICA-derived optimal CVs should outperform biasing either the initial suboptimal CV or the molecular features (z_1 and z_2) directly.

Time-Lagged Independent Component Analysis.

Once the covariance matrices central to tICA were computed, these features were then linearly combined to generate two CVs that optimally describe the slow modes of the system by solving a generalized eigenvalue problem. The robustness of the tICA CVs was examined over a range of lag times from 1 to 9 ns. Finally, tICA CVs were biased as new CVs in WTMetaD simulations to investigate any significant improvement in the sampling efficiency. To understand the effect of biasing on tICA CVs, we applied the method to both biased and unbiased subtrajectories, which involve different numbers of drug permeation events. We obtained unbiased trajectories from biased ones through a reweighting scheme, as described in the next section.

Here, we provide a brief introduction to the method, as the theoretical background of tICA can be found in many other works. tICA is an unsupervised dimensionality reduction algorithm designed to find a maximally slow subspace $\boldsymbol{U}$ via a linear transformation of molecular features $\boldsymbol{\chi}$ by maximizing the autocorrelation function of their projections. In other words, it finds the degrees of freedom where the correlation decays slowly and provides an optimal linear solution to the variational approach to conformational dynamics that approximates the slow components of reversible Markov processes. Mathematically, tICA determines the slowest independent collective degrees of freedom $\boldsymbol{u}_i$ onto which the projections $s_i(t) = \boldsymbol{u}_i \,.\, \boldsymbol{\chi}(t)$ have the largest autocorrelation function

$$\frac{\langle s_i(t)s_i(t + \tau)\rangle}{\left\langle s_i(t)^2\right\rangle} \tag{8}$$

for a chosen lag time τ. Equivalently, tICA maximizes the ratio $\boldsymbol{J}(\boldsymbol{U})$

$$\boldsymbol{J}(\boldsymbol{U}) = \frac{\boldsymbol{U}^{\mathrm{T}}\boldsymbol{C}(\tau)\boldsymbol{U}}{\boldsymbol{U}^{\mathrm{T}}\boldsymbol{C}(0)\boldsymbol{U}} \tag{9}$$

with respect to $\boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_d]$. This is equivalent to finding the solution of the generalized eigenvalue problem

$$\boldsymbol{C}(\tau)\boldsymbol{U} = \boldsymbol{C}(0)\boldsymbol{U}\boldsymbol{\Lambda} \tag{10}$$

where τ is the lag time, $\boldsymbol{C}$ is the covariance matrix (i.e., $\boldsymbol{C}(\tau) = \left\langle \chi(t)\chi^{\mathrm{T}}(t+\tau)\right\rangle_t$ where the data χ is mean-free), $\boldsymbol{U}$ is the eigenvector matrix that contains linearly independent components (ICs), and $\boldsymbol{\Lambda}$ is a diagonal eigenvalue matrix that contains autocorrelations. The dominant ICs are used to reduce the dimensionality of the phase space and capture the slowest modes of the molecular system. A direct estimation of $\boldsymbol{C}(\tau)$ from finite trajectories generally yields an asymmetric matrix and complex-valued eigenvectors and eigenvalues that are inconsistent with reversible molecular dynamics. To circumvent this problem, a symmetric estimator is often used to approximate the covariance matrices by averaging over all pairs $(\boldsymbol{x}_t, \boldsymbol{x}_{t+\tau})$ and the reverse $(\boldsymbol{x}_{t+\tau}, \boldsymbol{x}_t)$, which is equivalent to averaging time-forward and time-reversed trajectories, where x is a set of mean-free molecular coordinates.[34]

tICA suffers from one significant pitfall: the choice of a proper value of τ that determines which kinetic processes are considered to construct the new subspace. Ideally, τ must be sufficiently long that the dynamics described by ICs satisfy the Markovian property of being memoryless but also sufficiently short that the significant dynamics are not neglected. Currently, there is no recipe to choose an appropriate value of the parameter.[18] Therefore, ICs are ideally robust within a range of τ smaller than their time scales, but it is practically inevitable that they change to some extent as the parameter value changes.[71] In this work, we examine the components of the eigenvectors as a function of τ, as done previously by Parrinello and colleagues.[41,72]

Reweighting Algorithm.

Information about the unbiased state of the system can be recovered directly from a metadynamics simulation by means of a reweighting procedure.[45] We approximated the unbiased slow modes of the system from our WTMetaD trajectories using the reweighting scheme proposed by McCarty and Parrinello.[41] Here, we summarize the key steps involved in the reweighting procedure. For the column vector $\chi(t)$ of the molecular features at time t, the two covariance matrices in eq 10 are calculated as follows

$$\boldsymbol{C}(0) = \sum_t w(t)\chi(t)\chi^{\mathrm{T}}(t) \tag{11}$$

$$\boldsymbol{C}(\tau) = \sum_t w(t)\chi(t)\chi^{\mathrm{T}}(t+\tau) \tag{12}$$

where the statistical weight $w(t)$ of the configuration at time t is given by

$$w(t) = \frac{\mathrm{e}^{\beta[V(\boldsymbol{s},t) - c(t)]}}{\sum_t \mathrm{e}^{\beta[V(\boldsymbol{s},t) - c(t)]}} \tag{13}$$

In the case where walls are employed, the potentials are not included in the calculation of the statistical weights. Furthermore, the time scale in biased simulations needs to be properly rescaled as follows

$$\mathrm{d}\tilde{t} = \mathrm{e}^{\beta[V(\boldsymbol{s},t) - c(t)]}\mathrm{d}t \tag{14}$$

so that τ is given by the sum of the rescaled time steps

$$\tau = \sum_{t = t_0}^{\tau'} \mathrm{e}^{\beta[V(s,t) - c(t)]}\Delta t \tag{15}$$

If N molecular features are chosen to define the basis functions, then $\boldsymbol{M}$ eigenvectors that correspond to the slowest relaxation times (i.e., largest eigenvalues) are selected to be biased as optimal CVs with WTMetaD

$$s_i(\boldsymbol{R}) = \sum_{k = 1}^{M \leq N} u_{ik}\chi_k(\boldsymbol{R}) \tag{16}$$

where $\boldsymbol{u}_i$ is the ith eigenvector in $\boldsymbol{U}$ (or the set of expansion coefficients) in eq 10.

Contact Analysis.

Contact analysis was performed for the DAP and TMB moieties interacting with different components of the system, including water, cholesterol, POPC, and the choline-phosphate (PO_4) group or ester region of POPC (Figure S11). A contact was defined as any atom present within a cutoff radius of 3.5 Å of the reference group. The analysis was performed using the MDAnalysis Python package.[73] The same groups were used to calculate the average interaction energy using gmx energy.

RESULTS AND DISCUSSION

Application of tICA-MetaD to Membrane Permeation Simulations.

We employed the tICA-MetaD scheme (Figure S16) proposed by McCarty and Parrinello[41] to identify unbiased slow CVs for the passive diffusion of trimethoprim across a model heterogeneous lipid membrane composed of POPC and cholesterol. The procedure starts from the design of any arbitrary suboptimal CV s_0 that enhances the occurrence of rare events. Typically, the molecule's COM distance from the membrane would be selected. In this work, we selected instead a naïve linear combination of two molecular features, z_1 and z_2 (Figure 1c), that were previously demonstrated to improve the convergence of permeation calculations (see the CV Selection section[8]). We defined s_0 as the sum of z_1 and z_2 with equal weights, namely, $s_0 = 0.7071z_1 + 0.7071z_2$. Mathematically, this is simply the unit vector of

(1,1) in the descriptor space of z_1 and z_2. Physically, s_0 carries the same information as z_{CM} (i.e., the z-position of the COM of the drug molecule relative to that of the lipid bilayer) but is rescaled; when we plot s_0 as a function of z_{CM} using one of the WTMetaD trajectories biasing s_0 (see the following paragraph), we observe a linear relationship between the two variables, namely, $s_0 = 1.4153 z_{\mathrm{CM}}$ with $R^2 = 0.9997$ (Figure 1a). Interestingly, s_0 is also the first PC that retains 99.84% of the total variance when PCA is applied to the (z_1, z_2) points obtained from a 5-μs-long unbiased trajectory of the system (Figure 1b). The selection of s_0 as our suboptimal initial CV is ideal for the following three reasons: (1) it enables direct comparison with previous permeation simulations in which z_1 and z_2 were biased, (2) it allows us to bias a single CV mimicking the anticipated scenario wherein a user selects a single molecular COM, and (3) following the protocol of McCarty and Parrinello[41] it is a direct test of the method's efficacy at defining an optimal linear combination based on a suboptimal linear combination.

Next, we performed WTMetaD simulations that bias s_0 and thus induce multiple permeation events. In this step, we generated 10 different biased trajectories for reliable statistical analysis, and each trajectory was 3 μs long. Figure 2a shows how suboptimal CV s_0 evolves with time in one of the biased trajectories. The dashed lines indicate the average positions of the phosphorus atoms of POPC. It is clear that s_0 induces transitions between the local minima. The average rate of complete permeation (i.e., mean recrossing frequency) was calculated to be 3.3 ± 1.3 μs^{-1}. (To determine the average rate, we divided the total number of membrane crossings by the total length of the simulation after omitting the transient time.) Figure 2d shows the difference $\gamma(t)$ between the instantaneous bias potential $V(s, t)$ (Figure 2b) and the bias offset $c(t)$ (Figure 2c) as a function of time, which determines the statistical weight of each time frame in the reweighting procedure. The gray region in Figure 2c indicates the transient time to be eliminated for tICA.

The average 1D PMF for the permeation process is presented in Figure 3. Two local minima are found at $s_0 = \pm 2.3\ \mathrm{nm}$, which are separated from the aqueous regions (corresponding to the regions where s_0 is in the range of [−5, −4] or [4, 5] nm) by small energy barriers located at $s_0 = \pm 3.3\ \mathrm{nm}$. Relative to the bulk region, the depth of the minima is ~1.0 kJ/mol, and the height of the barriers is approximately 4.0 kJ/mol. The system reaches the local minima when the TMB group resides in the hydrophobic region of lipid tails, while the DAP group interacts with the polar headgroups (Figure 3a,c). A large, broad energy barrier of 32.4 kJ/mol, which connects the two local minima, is observed at $s_0 = 0\ \mathrm{nm}$. The energy barrier arises when trimethoprim moves and flips inside the hydrophobic core of the lipid bilayer. s_0 can describe the translational motion but not the orientational change of the drug molecule. For example, we observed from unbiased simulations less populated conformations in which the DAP group dwells in the hydrophobic region while the TMB group is exposed to the polar region of the lipid bilayer, but the corresponding values of s_0 to those conformations are not distinguishable. From this observation, we anticipate that the optimal slow CVs should account for both translational and orientational modes of the permeant.

To identify the unbiased slow modes, we reweighted the biased trajectories and performed tICA on them as proposed by McCarty and Parrinello.[41] The results from tICA on the first reweighted trajectory are summarized in Figure 4. The red and yellow curves indicate the first and second components $c_i^{(1)}$ of the first eigenvector (or IC) in Figure 4a; likewise, the green and chartreuse curves represent the first and second components $c_i^{(2)}$ of the second eigenvector in Figure 4b. The components of the eigenvectors are fairly stable with respect to τ. The slow CVs are expressed as a linear combination of the chosen molecular features: e.g., $s_1 = 0.9994z_1 - 0.0358z_2$ and $s_2 = 0.7193z_2 - 0.6947z_1$ at $\tau = 4$ ns. (Note that the effect of the PBCs on tICA CVs is not negligible for membrane permeation trajectories, and thus tICA should be performed carefully as we discuss in the next section. For this analysis, we selected one of the trajectories whose tICA CVs were least affected by the PBCs.) Intriguingly, the first and second CVs represent the translational and orientational modes, respectively, as the ICs can be approximated as $s_1 \approx z_1$ and $s_2 \approx z_2 - z_1$. The first CV carries the same information as those of s_0 and z_{CM} because its value simply represents the vertical position of the COM of the TMB group. The second CV, which is the weighted difference between z_1 and z_2, is strongly correlated with the orientation angle θ of trimethoprim with respect to the surface normal of the membrane (i.e., the z-axis). This is clearly shown in Figure S1 where the inner product of the directional vector $\hat{s}_2$ of trimethoprim and the surface normal $\hat{n}$ of the membrane shows a linear relationship with the value of s_2: $\cos\ \theta = -2.028s_2$ with $R^2 = 0.9989$. Hence, $\hat{s}_2$ is oriented toward the membrane if $s_2 < 0$ when $s_1 > 0$ (near the upper leaflet) or if $s_2 > 0$ when $s_1 < 0$ (near the lower leaflet). The width of the curve simply reflects conformational fluctuations (particularly, elongation and contraction in the direction of $\hat{s}_2$) of the drug molecule since the value of s_2 will change significantly when $\hat{s}_2 \parallel \hat{n}$ (i.e., $\cos\ \theta = \pm 1$) and minimally when $\hat{s}_2 \perp \hat{n}$ (i.e., $\cos\ \theta = 0$) since both z_1 and z_2 are vertical positions with respect to the membrane COM.

Figure 4c displays the eigenvalues λ_i corresponding to the first (red) and second (green) eigenvectors as a function of τ. A clear separation of the two time scales is detected. The eigenvalues quantify the largest autocorrelation functions of the projections of the molecular feature vector onto the slowest independent collective degrees of freedom at a given τ (Figure S2). Physically, they are directly related to relaxation times t_i^* associated with the slow modes by

$$t_i^* = -\frac{\tau}{\ln|\lambda_i|} \tag{17}$$

The dominant eigenvalue λ_1 (and thus the slowest relaxation mode) is associated with the translational kinetics, and the associated relaxation time is $t_1^* = 7.78\ \mathrm{ns}$ at $\tau = 4\ \mathrm{ns}$. The other eigenvalue λ_2 (and thus the second slowest relaxation mode) is associated with the orientational kinetics, and the associated relaxation time is $t_2^* = 2.28\ \mathrm{ns}$ at $\tau = 4\ \mathrm{ns}$. Figure 4d shows the ratio of the kinetic variance KV_i retained by each IC at a given τ after dimensionality reduction, which is given by

$$\mathrm{KV}_i = \frac{\lambda_i^2}{\lambda_1^2 + \lambda_2^2}$$

(18)

in this study. The kinetic variance is mostly retained by s_1; for example, $\mathrm{KV}_1 = 0.92$ and $\mathrm{KV}_2 = 0.08$ at $\tau = 4\ \mathrm{ns}$.

2D PMF along tICA CVs.

The 2D PMF spanned by our new CVs, $s_1 = 0.9964z_1 - 0.0851z_2$ and $s_2 = 0.7197z_2 - 0.6942z_1$ and the MFEP, calculated with the zero-temperature string method,[67,68] more clearly captures the permeation process (Figure 5a). The two metastable states exist at at $A = (-1.5,\ -0.3)$ and $B = (1.5,\ 0.3)$, and the transition state is located at (0, 0). Importantly, the PMF shows the preferential orientation of trimethoprim in the membrane; $\hat{s}_2$ points away from the membrane center (i.e., the DAP group in the polar region and the TMB group in the hydrophobic region of the lipid bilayer) when the system is in the metastable states—see Figure 3a,c. The MFEP is a curve γ connecting critical points on an energy landscape V that satisfies $(\nabla V)^{\perp}(\gamma) = 0$ and represents the most probable transition path in a large ensemble of the permeation processes.[74] The MFEP starts with the DAP group oriented toward the lipid headgroups, while trimethoprim is in the aqueous region and above the lipid membrane (A in Figure 5a), which is slightly favored over the opposite orientation (by ~2.0 kJ/mol). From there, a barrier is encountered while TMB flips down into the hydrophobic region of phospholipids, reversing the orientation of trimethoprim (B in Figure 5a). The drug then crosses a larger barrier, passing through the lipid tails and flipping again to orient DAP once again to the polar glycerol, phosphate, and headgroup region (C in Figure 5a). The drug molecule then escapes from the lipid bilayer to the aqueous region with TMB flipping into the aqueous region (D in Figure 5a). The 1D PMF along the MFEP (Figure 5b) is definitely different from the 1D PMF from our suboptimal CV (Figure 3d). The heights of the small and large energy barriers are ~17.9 and ~47.1 kJ/mol, respectively, and the depth of the metastable states is ~3.2 kJ/mol. The green points in Figure 5c indicate the conformational space sampled in the 5-μs-long unbiased simulation. Note that only the metastable states and aqueous regions are sampled. Our visual inspection reveals one or two flipping events per 1 μs on average at the region of lipid headgroups in the unbiased simulation. We analyzed the orientation of trimethoprim when it approaches the membrane surface from the aqueous region (Figure S12a) and when it is buried in the membrane at the metastable states (marked with B and C in Figure 5a) (Figure S12b). This shows the probability distribution of the orientation angle θ of trimethoprim with respect to the surface normal of the nearest leaflet. We defined the orientation of the molecule as the vector connecting from the COM of the TMB group to the COM of the DAP group, or simply $\hat{s}_2$. This confirms that the drug molecule reverses its orientation at the polar headgroup region such that the TMB flips into the hydrophobic tail region.

To confirm our observation that the TMB group penetrates the lipid tails more readily, a series of analyses were also performed. First, tracking the z-coordinate of the COM of each group in the unbiased simulation (Figure 6a) shows that when trimethoprim begins to enter the tail region $(-2\ \mathrm{nm} < Z_{\mathrm{COM}} < 2\ \mathrm{nm})$, TMB goes deeper than DAP. Concurrently, the failed attempts to penetrate more frequently show DAP deeper, consistent with its preferential interaction with the lipid headgroups. We observed the same behavior in the biased simulations; TMB dives deeper each time trimethoprim attempts to penetrate, and successful crossings are marked by a rapid flip (Figure 6b).

This information led to an investigation of which lipids the DAP and TMB groups interact with. Contact analysis was performed to explore different regions of the unbiased simulation system, focusing on the membrane interactions. While both DAP and TMB have a similar number of contacts with the hydrophilic groups (i.e., choline and phosphate groups), TMB surpasses DAP in contact with the deeper ester groups (Figure S13a–c). This group is not only deeper in the membrane but also includes the first carbons of the hydrophobic tails. The methoxy groups of TMB, due to their carbons having a slightly negative charge ($q \sim -0.1\ e$), facilitate burial in this region, where the carbons have a similar but positive charge ($q \sim 0.2\ e$). The oxygen atoms are also less negative ($q \sim -0.39\ e$) than the nitrogen atoms of the DAP group ($q \sim -0.75\ e$), thus having less electrostatic repulsion with ester oxygens. This trend (related to the carbons) is also observed in the contacts of each component with cholesterol, resulting in more contacts of TMB with POPC/cholesterol overall (Figure S15a,b). Finally, the contacts of TMB and DAP with the water molecules support the flip mechanism of trimethoprim when entering the membrane: when trimethoprim is in water, TMB has more total interactions with water molecules than DAP (equivalent when normalized per atom, Figure S14d). When the molecule penetrates more deeply into the membrane (consistent with Figure 6a), it flips, directing the TMB portion into the more buried, hydrophobic tails, dropping water contacts to 0 in Figure S13d. Similar observations are apparent in the number of contacts per atom (Figure S14).

The average interaction energies of DAP and TMB with each system component were also calculated to verify the described preferential interactions (Table S1). Based again on the unbiased simulation, TMB shows a larger average interaction energy than DAP with each component, simply because it is a larger molecule. But the relative increase is substantially larger for the ester group and cholesterol, and significantly reduced for the phosphate and choline groups.

Interestingly, our triple-flip mechanism is not consistent with the triple-flip model previously proposed by Sun et al.[8] based on their TTMetaD simulations of trimethoprim permeating a homogeneous lipid bilayer composed only of POPC. In this model, the TMB group preferentially stays in the polar region of lipid headgroups, while the DAP group flips into the hydrophobic region of lipid tails due to its hydrophobicity. The mechanistic difference may be attributed to the effect of membrane cholesterol. However, there is no convincing evidence of the relative hydrophobicity of the two moieties. For example, Masoud et al.[75] estimated the dipole moment of 2,4-diaminopyridimine to be ~2.25–2.48 D based on their quantum-mechanical calculations, whereas the experimental dipole moment of 1,2,3-trimethoxybenzene was reported to be 2.25 D.[76] In addition, the TMB group possesses a benzene ring with only three hydrogen bond acceptors, whereas the DAP group contains a pyrimidine ring with two hydrogen bond donors and four hydrogen bond acceptors. Consequently, it is more energetically favorable for DAP to interact with the polar headgroups, while TMB flips to water or the hydrophobic region of the lipid bilayer. Collectively, TMB delves deeper when trimethoprim attempts to penetrate the lipid bilayer, forming more contacts with the ester groups and POPC/cholesterol than DAP. Additionally, TMB has less electrostatic repulsion with the ester oxygens and, in water, interacts more with water molecules than DAP. The average interaction energy of TMB significantly increases with the ester group and cholesterol but reduces with the polar headgroups. TMB's benzene ring forms fewer hydrogen bonds than DAP's pyrimidine ring. As a result, DAP prefers interacting with the polar headgroups, whereas TMB tends to flip toward water or the hydrophobic region of the lipid bilayer.

In the study of Mehdi et al.,[46] the authors considered two pivotal order parameters (OPs) pertaining to BA permeation, namely, the membrane-BA z-distance (d_{1Z}) and the angle formed between BA and the z-axis (θ_Z). Analysis of the SPIB RCs identifies key steps involved in the entry and exit mechanisms of BA that remarkably bear a striking resemblance to the triple-flip model observed in the permeation process of trimethoprim.

The mean recrossing frequency (or the average rate of membrane crossings) is an important metric to determine the sampling efficiency of the chosen CVs in membrane permeation simulations. Comparing recrossing frequencies serves as a reliable means to assess whether a crucial variable (slow mode) has been overlooked. It is widely known that if a slow mode is forgotten, the sampling process may suffer from hysteresis, leading to a lack of convergence in FES calculations.[22,66,77] Figure S3 displays how z_1 changes over time when (a) s_0 and (b) s_1 and s_2 are biased in biased trajectories. We found that the mean recrossing frequency of s_0, $3.3 \pm 1.3\ \ \mu\mathrm{s}^{-1}$, increases to $5.0 \pm 2.2\ \ \mu\mathrm{s}^{-1}$ (~1.5 times increase) when s_1 and s_2 are used as CVs. This is only a modest increase in efficiency, likely due to the translational position captured by both s_0 and s_1 being the dominant slow mode.

Lastly, we found that the same tICA-MetaD procedure cannot be directly applied to the TTMetaD trajectories. TTMetaD was developed to prevent undesirable situations that may arise from the inappropriate choice of the bias factor, which controls the speed at which the Gaussian height decreases. TTMetaD aggressively tempers the height of Gaussian hills only after basins, whose locations are roughly defined prior to simulations, are relatively full. We generated one 1-μs-long TTMetaD trajectory with the same parameters as in our previous study[52] and plotted how rescaled z_1, $V(s,\ \ t)$, $c(t)$, and $\gamma(t)$ evolve with time in Figure S4. We observed that the difference between $V(s,t)$ and $c(t)$ is extremely large in TTMetaD compared to WTMetaD; $\gamma(t)$ fluctuates mostly between 10 and −50 kJ/mol in Figure 2d and around −160 kJ/mol in the lower right panel of Figure S4. Therefore, the time step becomes extremely small when rescaled according to eq 14, and consequently, tICA cannot be performed for a reasonable range of τ. One potential approach to resolve this issue could be to shift the bias potential, which is defined up to a constant. However, this theoretical possibility requires further investigation.

Effect of the Periodic Boundary Conditions on the Slow CVs.

A significant problem appears when the time series analysis is done directly on the data points obtained from membrane permeation simulations because the PBCs make the behavior of the permeant "unphysical" when it crosses the boundaries (particularly in the z-dimension for our systems). The tICA algorithm would recognize the boundary conditions as the unrealistic transfer (or “teleportation”) of the molecule from one end of the simulation box to the other without traversing the physical space between them. We performed tICA on 10 different reweighted trajectories to compute the slow CVs at 10 different lag times ranging from 1 to 10 ns (see Figures 7a and S5). We observed that the first eigenvectors are scattered on the domain defined by $c_1^2 + c_2^2 = 1$ and $-\pi \leq \theta \leq \pi/2$ where θ is the angle between the positive c_1-axis and the vector. The data points lie on the circumference of a unit circle, as the eigenvectors are normalized. Interestingly, the second eigenvectors are slightly scattered around (−0.7, 0.7) and show a high level of precision compared to the first eigenvectors since the orientational change of trimethoprim is less affected by the PBCs than its translational motion. To identify the center of each eigenvector in the data, we used the Partitioning Around Medoids (PAM) clustering algorithm, which is similar to k-means clustering but requires the centroid of each cluster to be one of the input data points. We found that the centers of the first and second eigenvectors are (0.639, −0.769) and (−0.707, 0.707), respectively. Thus, when the PBCs are applied, the translational mode is not well captured by tICA as the drug molecule can reach the other side of the membrane rapidly without crossing the membrane.

We examined two different approaches to address the PBC issue. In the first approach, we applied harmonic potentials to construct lower and upper walls in the water region below and above the bilayer and thus trap trimethoprim inside the space between the two boundaries in the z-direction. The free energy surface along z_1 and z_2 in the presence of the walls is shown in Figure S6, and we did not see any significant difference from the one obtained in the absence of the walls in previous works.[8,52] We applied tICA to 5 different reweighted trajectories to calculate the slow CVs at 10 different lag times ranging from 1 to 10 ns (see Figures 7b and S7). As in the case of the PBCs, the first eigenvectors are scattered on the circumference of a unit circle within $-\pi \leq \theta \leq \pi/2$ but are characterized by a higher degree of dispersion. This is likely because the behavior of the drug molecule is unnatural due to its collisions with the walls in the aqueous region. However, the second eigenvectors capture the orientational mode more precisely compared to the case in which the walls are absent since boundary crossings are physically prevented by the walls. Our k-medoids analysis reveals that the centroids are (0.179, −0.984) and (−0.712, 0.702) for the first and second eigenvectors, respectively. The centroid of the first eigenvectors still captures the translational motion of the drug molecule to a considerable extent with the coefficient of z_2 much larger in magnitude than that of z_1, but their level of dispersion does not make it ideal to be selected as the translational CV. In addition, the mean recrossing frequency is 3.2 ± 0.3 μs^{-1} when z_1 and z_2 are biased as CVs in the presence of the harmonic walls, which is comparable to that of s_0.

In the second approach, we took absolute values of the molecular features z_1 and z_2 in the time series data to make them continuous even when trimethoprim crosses the periodic boundaries in z-direction. The molecular features are any real numbers in the range $[-d_z/2, d_z/2]$ where d_z is the length of the simulation box in the z-direction. The molecular features are signed perpendicular distances and thus invariariant under translation but suffer from discontinuity at the periodic boundaries. Their absolute values lose the information about which leaflet of the membrane is closer to trimethoprim but only retain the information about how far the drug molecule is away from the membrane as they can take any real numbers in the range $[0, d_z/2]$. The tICA results are summarized in Figures 7c and S8. Surprisingly, the eigenvectors are highly stable with respect to lag times and consistent over different trajectories. Furthermore, the levels of dispersion for both the first and second eigenvectors are significantly reduced. The medoids are (0.989, −0.148) and (−0.638, 0.770) for the first and second eigenvectors, respectively. Therefore, the second approach achieves the highest level of precision for both the first and second eigenvectors and the highest level of accuracy for the first eigenvector, but does not predict well the orientational mode of the drug molecule. However, the sampling efficiency was not improved when we used the absolute values either as direct CVs or in the definitions of s_1 and s_2. The former is expected since $|z_1|$ and $|z_2|$ separately do not distinguish the upper and lower regions of the lipid bilayer (upper when $z_1, z_2 > 0$, and lower when $z_1, z_2 < 0$); taking the absolute values removes the locational information and thus biasing them cannot induce trimethoprim to cross the membrane. When we biased $s_1' = 0.989|z_1| - 0.148|z_2|$ and $s_2' = 0.770|z_2| - 0.6388|z_1|$, the mean recrossing frequency over three replicas was only 1.3 ± 0.5 μs^{-1} (Figure S9). Thus, taking the absolute values extracts the coefficients of the translational CV in a more reliable manner, but it poorly describes the orientational motion of the drug molecule. Also, using the absolute values in the definitions of CVs does not necessarily improve the sampling performance of WTMetaD due to a loss of the locational information on the drug molecule.

Effect of Data Reweighting.

Data reweighting recovers information about the unbiased dynamics of the system by using the conformations collected from biased simulations and their statistical weights. However, fast and accurate convergence to the underlying unbiased free energy is not always guaranteed depending on the reweighting procedure, the enhanced sampling technique, and the stage of the simulation.[45] In our previous work,[52] we obtained tICA CVs directly from short biased trajectories that involve at least one permeation event without data reweighting. Initially, a TTMetaD simulation lasting 70 ns was performed, using z_1 and z_2 as CVs to generate an initial trajectory involving at least one membrane crossing. Subsequently, the z-positions of five heavy atoms in the drug molecule, namely, Z_1 to Z_5, were calculated relative to the COM of the lipid bilayer. tICA was then applied to this trajectory to identify the most dominant eigenvector corresponding to the slowest mode. Another TTMetaD simulation was conducted, incorporating the tICA CV and the z-position of the COM of the drug molecule (denoted as Z). Comparisons between the results obtained from simulations utilizing tICA-based CVs and traditional CVs $(z_1$ and $z_2)$ revealed that the tICA CVs achieved faster convergence and yielded more accurate outcomes. Essentially, the use of tICA CVs enhances the efficiency of potential of mean force (PMF) calculations, while concurrently providing additional insights into the permeation mechanism. Specifically, the tICA CV unveiled a subtle influence of cholesterol on the resistance of the lipid headgroup region to permeation, which was not observed when employing the canonical CVs. It is, therefore, necessary to understand the effect of data reweighting on tICA CVs and determine the minimum number of rare events required for tICA CVs to converge. This information can be used to further optimize the tICAMetaD procedure for future membrane permeation simulations.

We first extracted from our WTMetaD trajectories the subtrajectories (or segments of trajectories of specified duration) that contain only one to seven membrane crossings, and grouped them into seven sets of 10 subtrajectories according to the number of membrane crossings included (see Figure S10 for an example of each set). We then performed tICA on each subtrajectory with and without data reweighting. We chose $|z_1|$ and $|z_2|$ as the molecular features for tICA because the eigenvectors show the highest degree of convergence when the absolute values are used (Figure 7c). Figure 8 summarizes the tICA results with data reweighting. Here, we adopted the same color scheme as in Figure 4 and displayed in order the 100 eigenvector pairs we obtained for each set (because the analysis was performed at 10 different lag times on 10 subtrajectories) to evaluate their converging behavior. We also included the results from the 10 fully biased trajectories (Figure 8h) for the purpose of comparison. We observed that tICA does not capture the slow modes properly when subtrajectories contain only one permeation event because τ is relatively large for the length of the subtrajectories. We found that at least five permeation events are needed for the tICA CVs to achieve a high level of convergence (Figure 8e). The medoids are (0.996, −0.087) and (−0.635, 0.772) for the first and second eigenvectors, respectively, in Figure 8e, and (0.989, −0.148) and (−0.638, 0.770) for the full trajectories in Figure 8h. The large fluctuations of the second component of the first eigenvector (yellow curves) arise mostly from the effect of vector normalization. For example, considering vectors only in the first quadrant, when the first component changes from 1 to 0.99, the second component changes from 0 to 0.14; when the first component changes from 0.70 to 0.69, the second component changes from only 0.71 to 0.72.

Next, we omitted data reweighting and applied tICA directly to each subtrajectory and full biased trajectories (Figure 9). The eigenvectors are well converged when at least five membrane crossings are involved. The medoids are (0.968, −0.250) and (−0.678, 0.735) for the first and second eigenvectors, respectively, in Figure 9e, and (0.965, −0.260) and (−0.676, 0.737) for the full trajectories in Figure 9h. However, the eigenvectors are quite different from those obtained from the reweighted trajectories. The first eigenvectors deviate more from the coefficients of the translational CV ($s_1 \approx z_1$), whereas the second eigenvectors get slightly closer to the coefficients of the orientational CV ($s_2 \approx 0.7z_2 - 0.7z_1$) upon removal of data reweighting. We also observed that the orientational mode is not significantly affected by data reweighting even when the absolute values of the molecular features are not taken and the PBCs are still present, as clearly seen in Figure S5 (dark and light lines for the first and second components of the second eigenvector, respectively); the second eigenvectors are highly consistent and converge to s_2 even without reweighting. These observations can be attributed to the fact that the original bias was applied to the suboptimal CV s_0 simply to increase the translational kinetics and, thus, the rate of membrane crossing of trimethoprim. In addition, since the orientational kinetics is much faster than the translational kinetics, we may assume that the two modes are not strongly coupled to each other. Consequently, without data reweighting the Boltzmann statistics of the translational mode cannot be recovered, whereas the orientational CV can be obtained with an acceptable degree of accuracy and convergence.

CONCLUSIONS

In this work, we applied the tICA-MetaD procedure proposed by McCarty and Parrinello[41] to trimethoprim membrane permeation to better understand the effectiveness, limitations, and best practices of this methodology for membrane permeation simulations. Specifically, we sought to identify the slowest collective degrees of freedom and see if they improved simulation efficiency while also investigating the effects of the PBCs, the number of rare events in the original biased simulations, and data reweighting on the accuracy and convergence of tICA CVs. We found that the same tICA-MetaD scheme cannot be directly applied to TTMetaD trajectories due to extremely small reweighting factors and rescaled time steps. We observed that tICA-MetaD captures the translational and orientational modes separately, and the use of the tICA CVs accelerates the convergence of WTMetaD PMF calculations compared to the suboptimal CV initially selected. However, the PBCs may be detrimental to tICA in membrane permeation trajectories, particularly for the identification of the translational CV. Adding harmonic restraints to prevent PBC crossing does not solve the problem because it causes unnatural, rapid diffusive behavior of the drug molecule in the aqueous region, whereas taking absolute values of the molecular features during tICA analysis can reliably recover the translational CV. In contrast, the orientational CV is not significantly affected by the PBCs and the walls but is also not well captured when the absolute values are taken. Thus, taking absolute values is only useful for extracting the translational CV, and a method that captures both modes without PBC artifacts would be an important contribution to the field. Interestingly, we found that only five permeation events are sufficient for the tICA CVs to achieve convergence regardless of data reweighting, but the first eigenvector is quite different from the translational CV when data reweighting is omitted. Based on our results, we suggest that (1) the tICA-MetaD procedure can be applied to a short initial WTMetaD trajectory (after the transient time) that involves at least five membrane crossings and (2) absolute values of the molecular features should be used along with data reweighting for correction of the translational CV.

Future work should investigate how to select a minimal set of molecular features for optimal CVs, how to incorporate deep learning algorithms to add flexibility in feature selection and evade the linearity problem of tICA, and how the results may change when different reweighting algorithms are used for WTMetaD and TTMetaD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from NIH NIGMS (R35GM143117) and computational support from EXPANSE at the San Diego Supercomputing Center through the ACCESS program (allocation MCB200018) supported by NSF (grants #2138259, #2138286, #2138307, #2137603, and #2138296), as well as the Center for High Performance Computing at the University of Utah.

ABBREVIATIONS

COM	center of mass
CGenFF	CHARMM general force field
CV	collective variable
DAP	diaminopyrimidine
IC	independent component
ISD	inhomogeneous solubilitydiffusion
KV	kinetic variance
LINCS	linear constraint solver
MetaD	metadynamics
MFEP	minimum free energy path
MD	molecular dynamics
PBC	periodic boundary condition
PO4	phosphate
POPC	phosphatidylcholine
PMF	potential of mean force
PC	principal component
PCA	principal component analysis
SPME	smooth particlemesh Ewald
tICA	time-lagged independent component analysis
TTMetaD	transition-tempered metadynamics
TMB	trimethoxybenzyl
VMD	visual molecular dynamics
WTMetaD	well-tempered metadynamics

REFERENCES

(1). Buchenberg S; Schaudinnus N; Gerhard S Hierarchical Biomolecular Dynamics: Picosecond Hydrogen Bonding Regulates Microsecond Conformational Transitions. J. Chem. Theory Comput. 2015, 11 (3), 1330–1336. [PubMed: 26579778]

(2). Henzler-Wildman K; Kern D Dynamic personalities of proteins. Nature 2007, 450 (7172), 964–972. [PubMed: 18075575]

(3). Lewandowski JR; Halse ME; Blackledge M; Emsley L Direct observation of hierarchical protein dynamics. Science 2015, 348 (6234), 578–581. [PubMed: 25931561]

(4). Chen M Collective variable-based enhanced sampling and machine learning. Eur. Phys. J. B 2021, 94 (10), No. 211, DOI: 10.1140/epjb/s10051-021-00220-w. [PubMed: 34697536]

(5). Ghaemi Z; Minozzi M; Carloni P; Laio A A Novel Approach to the Investigation of Passive Molecular Permeation through Lipid Bilayers from Atomistic Simulations. J. Phys. Chem. B 2012, 116 (29), 8714–8721. [PubMed: 22540377]

(6). Awoonor-Williams E; Rowley CN Molecular simulation of nonfacilitated membrane permeation. Biochim. Biophys. Acta, Biomembr. 2016, 1858 (7, Part B), 1672–1687.

(7). Sugita M; Sugiyama S; Fujie T; Yoshikawa Y; Yanagisawa K; Ohue M; Akiyama Y Large-Scale Membrane Permeability Prediction of Cyclic Peptides Crossing a Lipid Bilayer Based on Enhanced Sampling Molecular Dynamics Simulations. J. Chem. Inf. Model. 2021, 61 (7), 3681–3695. [PubMed: 34236179]

(8). Sun R; Dama JF; Tan JS; Rose JP; Voth GA Transition-Tempered Metadynamics Is a Promising Tool for Studying the Permeation of Drug-like Molecules through Membranes. J. Chem. Theory Comput. 2016, 12 (10), 5157–5169. [PubMed: 27598403]

(9). Pokhrel N; Maibaum L Free Energy Calculations of Membrane Permeation: Challenges Due to Strong Headgroup–Solute Interactions. J. Chem. Theory Comput. 2018, 14 (3), 1762–1771. [PubMed: 29406707]

(10). Krämer A; Ghysels A; Wang E; Venable RM; Klauda JB; Brooks BR; Pastor RW Membrane permeability of small molecules from unbiased molecular dynamics simulations. J. Chem. Phys. 2020, 153 (12), No. 124107. [PubMed: 33003739]

(11). Marrink S-J; Berendsen HJC Simulation of water transport through a lipid membrane. J. Phys. Chem. A 1994, 98 (15), 4155–4168.

(12). Venable RM; Krämer A; Pastor RW Molecular Dynamics Simulations of Membrane Permeability. Chem. Rev. 2019, 119 (9), 5954–5997. [PubMed: 30747524]

(13). Davoudi S; Ghysels A Sampling efficiency of the counting method for permeability calculations estimated with the inhomogeneous solubility–diffusion model. J. Chem. Phys. 2021, 154 (5), No. 054106, DOI: 10.1063/5.0033476. [PubMed: 33557559]

(14). Vervust W; Zhang DT; van Erp TS; Ghysels A Path sampling with memory reduction and replica exchange to reach long permeation timescales. Biophys. J. 2023, 122, 2960–2972, DOI: 10.1016/j.bpj.2023.02.021. [PubMed: 36809877]

(15). Swenson DWH; Prinz J-H; Noe F; Chodera JD; Bolhuis PG OpenPathSampling: A Python Framework for Path Sampling Simulations. 1. Basics. J. Chem. Theory Comput. 2019, 15 (2), 813–836. [PubMed: 30336030]

(16). Yang YI; Shao Q; Zhang J; Yang L; Gao YQ Enhanced sampling in molecular dynamics. J. Chem. Phys. 2019, 151 (7), No. 070902. [PubMed: 31438687]

(17). Ghysels A; Roet S; Davoudi S; van Erp TS Exact non-Markovian permeability from rare event simulations. Phys. Rev. Res. 2021, 3 (3), No. 033068.

(18). Kaptan S; Vattulainen I Machine learning in the analysis of biomolecular simulations. Adv. Phys.: X 2022, 7 (1), No. 2006080.

(19). Bussi G; Laio A Using metadynamics to explore complex free-energy landscapes. Nat. Rev. Phys. 2020, 2 (4), 200–212.

(20). Trapl D; Horvacanin I; Mareska V; Ozcelik F; Unal G; Spiwok V Anncolvar: Approximation of Complex Collective Variables by Artificial Neural Networks for Analysis and Biasing of Molecular Simulations. Front. Mol. Biosci. 2019, 6, No. 25, DOI: 10.3389/fmolb.2019.00025. [PubMed: 31058167]

(21). Sidky H; Chen W; Ferguson AL Machine learning for collective variable discovery and enhanced sampling in biomolecular simulation. Mol. Phys. 2020, 118 (5), No. e1737742.

(22). Valsson O; Tiwary P; Parrinello M Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. Annu. Rev. Phys. Chem. 2016, 67 (1), 159–184. [PubMed: 26980304]

(23). Bernetti M; Bertazzo M; Masetti M Data-Driven Molecular Dynamics: A Multifaceted Challenge. Pharmaceuticals 2020, 13 (9), No. 253, DOI: 10.3390/ph13090253. [PubMed: 32961909]

(24). Glielmo A; Husic BE; Rodriguez A; Clementi C; Noé F; Laio A Unsupervised Learning Methods for Molecular Simulation Data. Chem. Rev. 2021, 121 (16), 9722–9758. [PubMed: 33945269]

(25). Lange OF; Grubmüller, H. Can Principal Components Yield a Dimension Reduced Description of Protein Dynamics on Long Time Scales? J. Phys. Chem. B 2006, 110 (45), 22842–22852. [PubMed: 17092036]

(26). Sittel F; Jain A; Stock G Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates. J. Chem. Phys. 2014, 141 (1), No. 014111. [PubMed: 25005281]

(27). Amadei A; Linssen ABM; de Groot BL; van Aalten DMF; Berendsen HJC An Efficient Method for Sampling the Essential Subspace of Proteins. J. Biomol. Struct. Dyn. 1996, 13 (4), 615–625. [PubMed: 8906882]

(28). Amadei A; Linssen ABM; Berendsen HJC Essential dynamics of proteins. Proteins 1993, 17 (4), 412–425. [PubMed: 8108382]

(29). Daidone I; Amadei A Essential dynamics: foundation and applications. Wiley Interdiscip. Rev.: Comput. Mol. Sci. 2012, 2 (5), 762–770.

(30). David CC; Jacobs DJ; Livesay DR Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. Protein Dynamics: Methods and Protocols; Humana Press: Totowa, NJ, 2014; pp 193–226.

(31). Naritomi Y; Fuchigami S Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis. J. Chem. Phys. 2013, 139 (21), No. 215102. [PubMed: 24320404]

(32). Molgedey L; Schuster HG Separation of a mixture of independent signals using time delayed correlations. Phys. Rev. Lett. 1994, 72 (23), 3634–3637. [PubMed: 10056251]

(33). Noé F; Clementi C Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods. Curr. Opin. Struct. Biol. 2017, 43, 141–147. [PubMed: 28327454]

(34). Wu H; Nüske F; Paul F; Klus S; Koltai P; Noé F Variational Koopman models: Slow collective variables and molecular kinetics from short off-equilibrium simulations. J. Chem. Phys. 2017, 146 (15), No. 154104. [PubMed: 28433026]

(35). Schultze S; Grubmüller H Time-Lagged Independent Component Analysis of Random Walks and Protein Dynamics. J. Chem. Theory Comput. 2021, 17 (9), 5766–5776. [PubMed: 34449229]

(36). Pérez-Hernández G; Paul F; Giorgino T; Fabritiis GD; Noé F Identification of slow molecular order parameters for Markov model construction. J. Chem. Phys. 2013, 139 (1), No. 015102. [PubMed: 23822324]

(37). Sittel F; Stock G Perspective: Identification of collective variables and metastable states of protein dynamics. J. Chem. Phys. 2018, 149 (15), No. 150901. [PubMed: 30342445]

(38). Nüske F; Keller BG; Pérez-Hernández G; Mey ASJS; Noé F Variational Approach to Molecular Kinetics. J. Chem. Theory Comput. 2014, 10 (4), 1739–1752. [PubMed: 26580382]

(39). Noé F; Nüske F A Variational Approach to Modeling Slow Processes in Stochastic Dynamical Systems. Multiscale Model. Simul. 2013, 11 (2), 635–655.

(40). Sultan M; Pande VS tICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. J. Chem. Theory Comput. 2017, 13 (6), 2440–2447. [PubMed: 28383914]

(41). McCarty J; Parrinello M A variational conformational dynamics approach to the selection of collective variables in metadynamics. J. Chem. Phys. 2017, 147 (20), No. 204109. [PubMed: 29195289]

(42). Brotzakis ZF; Parrinello M Enhanced Sampling of Protein Conformational Transitions via Dynamically Optimized Collective Variables. J. Chem. Theory Comput. 2019, 15 (2), 1393–1398. [PubMed: 30557019]

(43). Zhang Y-Y; Niu H; Piccini G; Mendels D; Parrinello M Improving collective variables: The case of crystallization. J. Chem. Phys. 2019, 150 (9), No. 094509. [PubMed: 30849916]

(44). Dama JF; Rotskoff G; Parrinello M; Voth GA Transition-Tempered Metadynamics: Robust, Convergent Metadynamics via On-the-Fly Transition Barrier Estimation. J. Chem. Theory Comput. 2014, 10 (9), 3626–3633. [PubMed: 26588507]

(45). Schäfer TM; Settanni G Data Reweighting in Metadynamics Simulations. J. Chem. Theory Comput. 2020, 16 (4), 2042–2052. [PubMed: 32192340]

(46). Mehdi S; Wang D; Pant S; Tiwary P Accelerating All-Atom Simulations and Gaining Mechanistic Understanding of Biophysical Systems through State Predictive Information Bottleneck. J. Chem. Theory Comput. 2022, 18 (5), 3231–3238. [PubMed: 35384668]

(47). Bonati L; Piccini G; Parrinello M Deep learning the slow modes for rare events sampling. Proc. Natl. Acad. Sci. U.S.A. 2021, 118 (44), No. e2113533118. [PubMed: 34706940]

(48). Chen H; Chipot C Chasing collective variables using temporal data-driven strategies. QRB Discovery 2023, 4, No. e2. [PubMed: 37564298]

(49). Donati L; Keller BG Girsanov reweighting for metadynamics simulations. J. Chem. Phys. 2018, 149 (7), No. 072335, DOI: 10.1063/1.5027728. [PubMed: 30134671]

(50). Shmilovich K; Ferguson AL Girsanov Reweighting Enhanced Sampling Technique (GREST): On-the-Fly Data-Driven Discovery of and Enhanced Sampling in Slow Collective Variables. J. Phys. Chem. A 2023, 127 (15), 3497–3517. [PubMed: 37036804]

(51). Crellin E; Mansfield KE; Leyrat C; Nitsch D; Douglas IJ; Root A; Williamson E; Smeeth L; Tomlinson LA Trimethoprim use for urinary tract infection and risk of adverse outcomes in older patients: cohort study. Br. Med. J. 2018, 360, No. k341, DOI: 10.1136/bmj.k341. [PubMed: 29438980]

(52). Aydin F; Durumeric AEP; Hora G. C. A. d.; Nguyen JDM; Oh MI; Swanson JMJ Improving the accuracy and convergence of drug permeation simulations via machine-learned collective variables. J. Chem. Phys. 2021, 155 (4), No. 045101. [PubMed: 34340389]

(53). Abraham MJ; Murtola T; Schulz R; Páll S; Smith JC; Hess B; Lindahl E GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2015, 1–2, 19–25.

(54). Tribello GA; Bonomi M; Branduardi D; Camilloni C; Bussi G PLUMED 2: New feathers for an old bird. Comput. Phys. Commun. 2014, 185 (2), 604–613.

(55). Klauda JB; Venable RM; Freites JA; O’Connor JW; Tobias DJ; Mondragon-Ramirez C; Vorobyov I; MacKerell AD; Pastor RW Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. J. Phys. Chem. B 2010, 114 (23), 7830–7843. [PubMed: 20496934]

(56). Vanommeslaeghe K; Hatcher E; Acharya C; Kundu S; Zhong S; Shim J; Darian E; Guvench O; Lopes P; Vorobyov I; Mackerell AD Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem. 2010, 31 (4), 671–690. [PubMed: 19575467]

(57). Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 1983, 79 (2), 926–935.

(58). Jo S; Lim JB; Klauda JB; Im W CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. Biophys. J. 2009, 97 (1), 50–58. [PubMed: 19580743]

(59). Martínez L; Andrade R; Birgin EG; Martínez JM PACKMOL: A package for building initial configurations for molecular dynamics simulations. J. Comput. Chem. 2009, 30 (13), 2157–2164. [PubMed: 19229944]

(60). Bussi G; Donadio D; Parrinello M Canonical sampling through velocity rescaling. J. Chem. Phys. 2007, 126 (1), No. 014101. [PubMed: 17212484]

(61). Berendsen HJC; Postma JPM; Gunsteren W. F. v.; DiNola A; Haak JR Molecular dynamics with coupling to an external bath. J. Chem. Phys. 1984, 81 (8), 3684–3690.

(62). Essmann U; Perera L; Berkowitz ML; Darden T; Lee H; Pedersen LG A smooth particle mesh Ewald method. J. Chem. Phys. 1995, 103 (19), 8577–8593.

(63). Hess B; Bekker H; Berendsen HJC; Fraaije JGEM LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. 1997, 18 (12), 1463–1472.

(64). Humphrey W; Dalke A; Schulten K VMD: Visual molecular dynamics. J. Mol. Graphics 1996, 14 (1), 33–38.

(65). Laio A; Parrinello M Escaping free-energy minima. Proc. Natl. Acad. Sci. U.S.A. 2002, 99 (20), 12562–12566. [PubMed: 12271136]

(66). Laio A; Gervasio FL Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Rep. Prog. Phys. 2008, 71 (12), No. 126601.

(67). E W; Ren W; Vanden-Eijnden E Finite Temperature String Method for the Study of Rare Events. J. Phys. Chem. B 2005, 109 (14), 6688–6693. [PubMed: 16851751]

(68). Cameron M; Kohn RV; Vanden-Eijnden E The String Method as a Dynamical System. J. Nonlinear Sci. 2011, 21 (2), 193–230.

(69). Sun R; Dama JF; Tan JS; Rose JP; Voth GA Transition-Tempered Metadynamics is a Promising Tool for Studying the Permeation of Drug-like Molecules through Membranes. J. Chem. Theory Comput. 2016, 12, 5157–5169. [PubMed: 27598403]

(70). Sun R; Han Y; Swanson JMJ; Tan JS; Rose JP; Voth GA Molecular transport through membranes: Accurate permeability coefficients from multidimensional potentials of mean force and local diffusion constants. J. Chem. Phys. 2018, 149 (7), No. 072310. [PubMed: 30134730]

(71). Naritomi Y; Fuchigami S Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: The case of domain motions. J. Chem. Phys. 2011, 134 (6), No. 065101. [PubMed: 21322734]

(72). Yang YI; Parrinello M Refining Collective Coordinates and Improving Free Energy Representation in Variational Enhanced Sampling. J. Chem. Theory Comput. 2018, 14 (6), 2889–2894. [PubMed: 29715017]

(73). Michaud-Agrawal N; Denning EJ; Woolf TB; Beckstein O MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. J. Comput. Chem. 2011, 32 (10), 2319–2327. [PubMed: 21500218]

(74). E W; Ren W; Vanden-Eijnden E Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. J. Chem. Phys. 2007, 126 (16), No. 164103. [PubMed: 17477585]

(75). Masoud MS; Awad MK; Shaker MA; El-Tahawy MMT The role of structural chemistry in the inhibitive performance of some aminopyrimidines on the corrosion of steel. Corros. Sci. 2010, 52 (7), 2387–2396.

(76). Exner O; Jehlička V Conformation around several equivalent bonds. Polymethoxy derivatives of benzene. Collect. Czech. Chem. Commun. 1983, 48 (4), 1030–1041.

(77). Barducci A; Bonomi M; Parrinello M Metadynamics. Wiley Interdiscip. Rev.: Comput. Mol. Sci. 2011, 1 (5), 826–843.

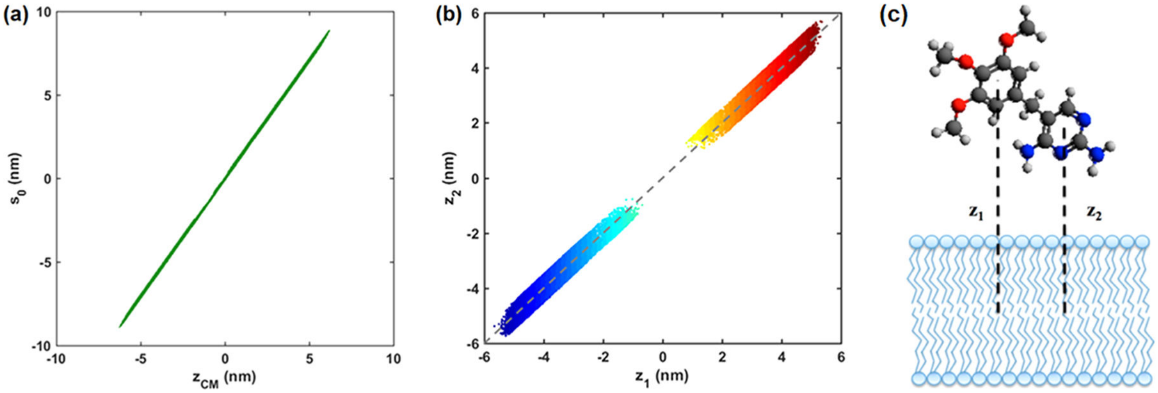


Figure 1.

(a) Linear relationship between the suboptimal CV s_0 and the z-position z_{CM} of the COM of trimethoprim with respect to that of the lipid bilayer. (b) Scatter plot displaying all of the (z_1, z_2) pairs accessed in 5 μs unbiased MD simulation. The red and yellow colors indicate the region above the bilayer, and the dark and light blue colors represent the region below the bilayer. The gap between the two regions signifies the interior of the membrane. The dashed gray line shows the first PC, s_0. (c) Molecular structure of trimethoprim and definitions of molecular features z_1 and z_2.

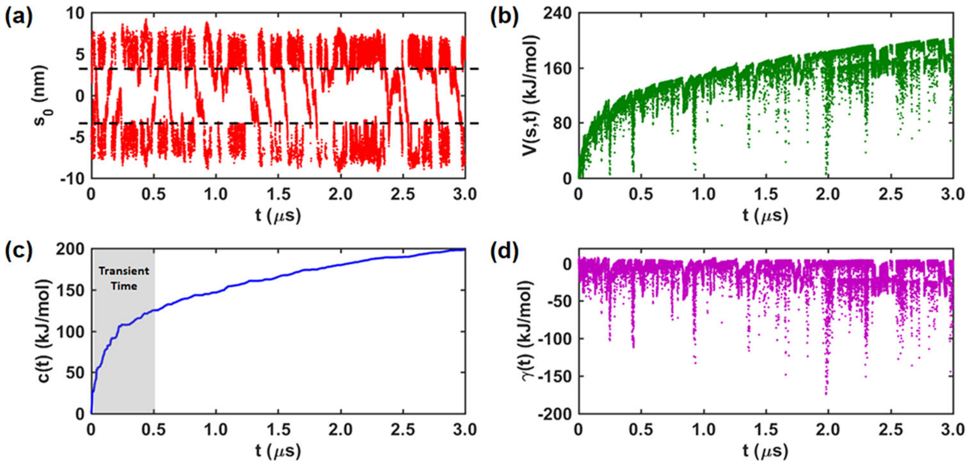


Figure 2.

Time evolution of (1) the suboptimal CV s_0, (2) the instantaneous potential $V(s,t)$, (3) the bias offset $c(t)$, and (d) the difference $\gamma(t)$ between $V(s,t)$ and $c(t)$ in one of the WTMetaD simulations that we performed. The dashed black lines in (a) indicate the average position of the phosphorus atoms of the lipid bilayer. The gray region in (c) indicates the transient time to be removed for tICA.

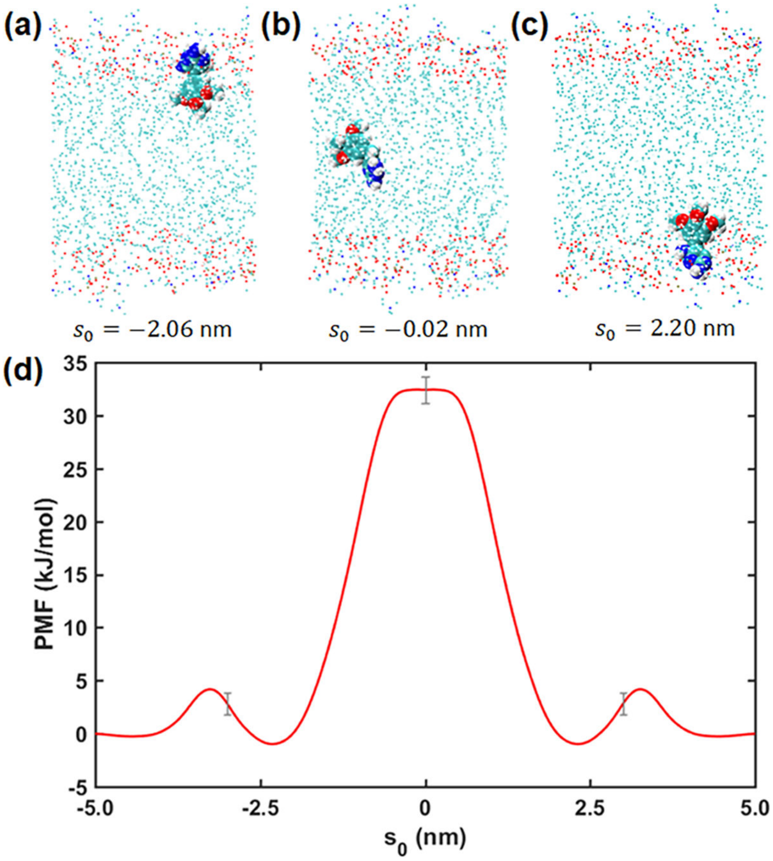


Figure 3.

(a–c) Typical snapshots of the system in the local minima and at the transition state with the corresponding values of s_0. (d) Average 1D PMF along the suboptimal CV s_0 for the permeation process of trimethoprim through the lipid bilayer. The free energy in bulk water was set to zero, and the error bars indicate the standard deviations for the arbitrarily selected points.

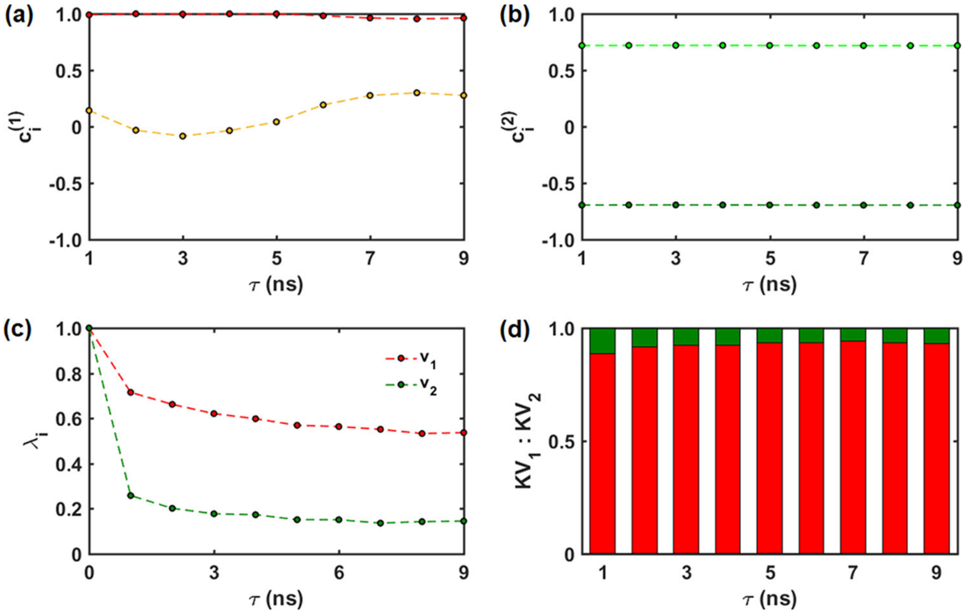


Figure 4.

tICA results from one of the WTMetaD trajectories we obtained. (a) First (red) and second (yellow) components ($c_1^{(1)}$, $c_2^{(1)}$) of the first eigenvector v_1, (b) first (green) and second (chartreuse) components ($c_1^{(2)}$, $c_2^{(2)}$) of the second eigenvector v_2, (c) first (red, $i = 1$) and second (green, $i = 2$) eigenvalues, and (d) the ratio of the kinetic variances KV$_i$ retained by the first (red, $i = 1$) and second (green, $i = 2$) eigenvectors as a function of the lag time τ.

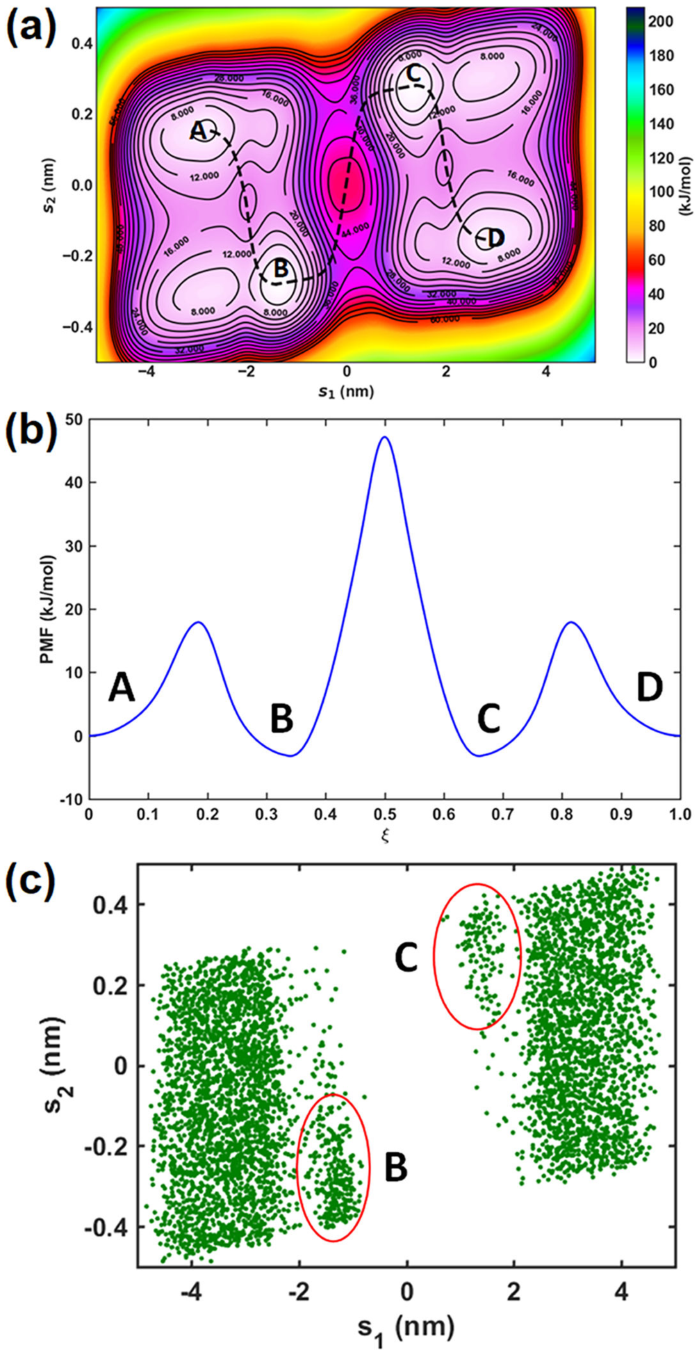


Figure 5.

(a) Average 2D PMF along the tICA CVs s_1 and s_2. The dashed black line indicates the MFEP found from the zero-temperature string method. (b) 1D PMF along MFEP ξ. The free energy in bulk water was set to zero. (c) Scatter plot showing all of the (s_1, s_2) pairs sampled by the long-timescale standard MD simulation. B and C represent the metastable states.

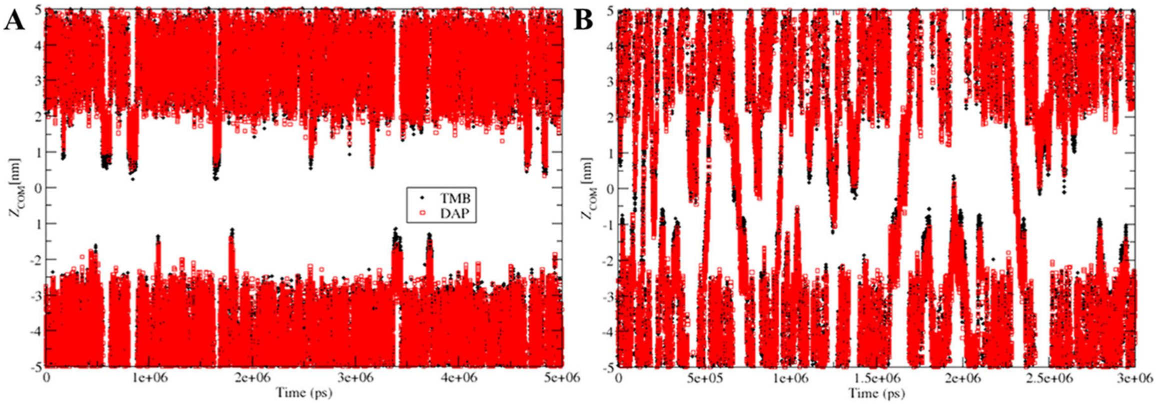


Figure 6.

Time evolution of the *z*-coordinates of the COMs of DAP (red) and TMB (black) in (a) unbiased and (b) biased simulations.

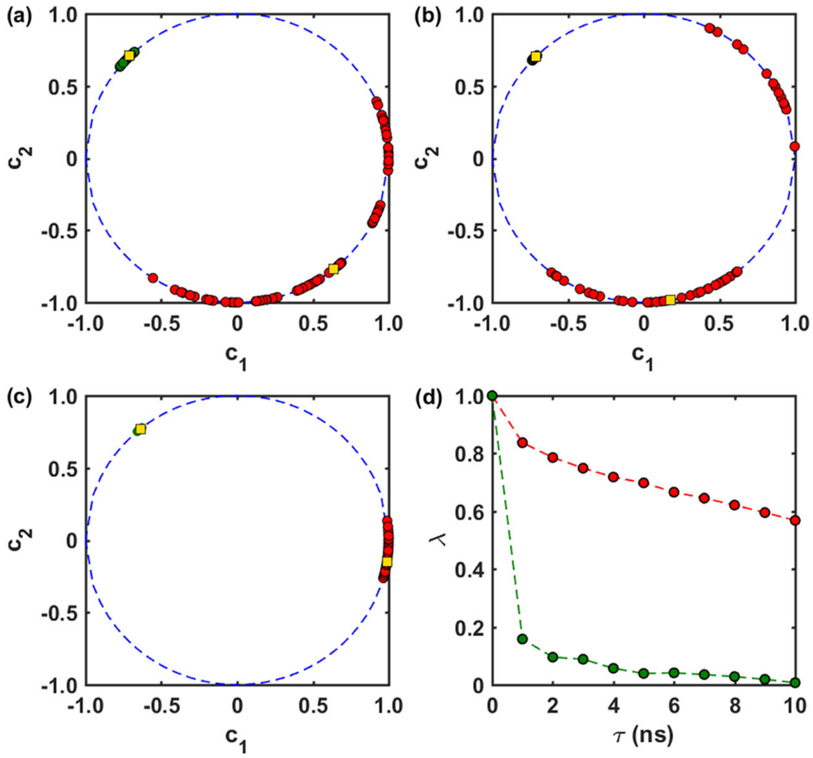


Figure 7.

First (red) and second (green) eigenvectors (c_1, c_2) obtained from tICA-MetaD with (a) the PBCs, (b) the upper and lower walls, and (c) the absolute values of the molecular features. The yellow squares represent the medoids. (d) First (red) and second (green) eigenvalues as a function of τ obtained from tICA on one of the full biased trajectories.

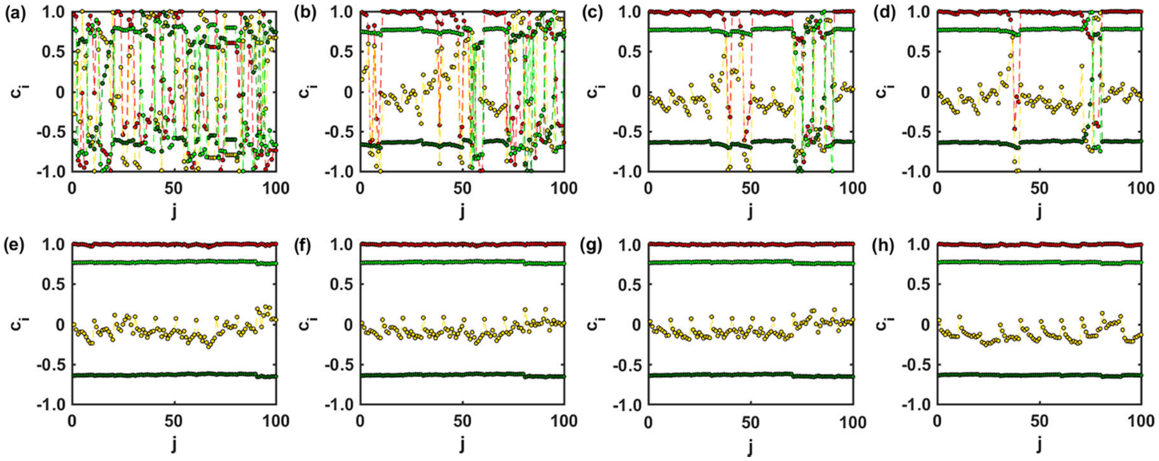


Figure 8.

Results from tICA on reweighted subtrajectories involving (a) one, (b) two, (c) three, (d) four, (e) five, (f) six, and (g) seven membrane crossings and (h) on the full biased trajectories after reweighting. The same color scheme was used as in Figure 4a,b to represent the components of the first and second eigenvectors.

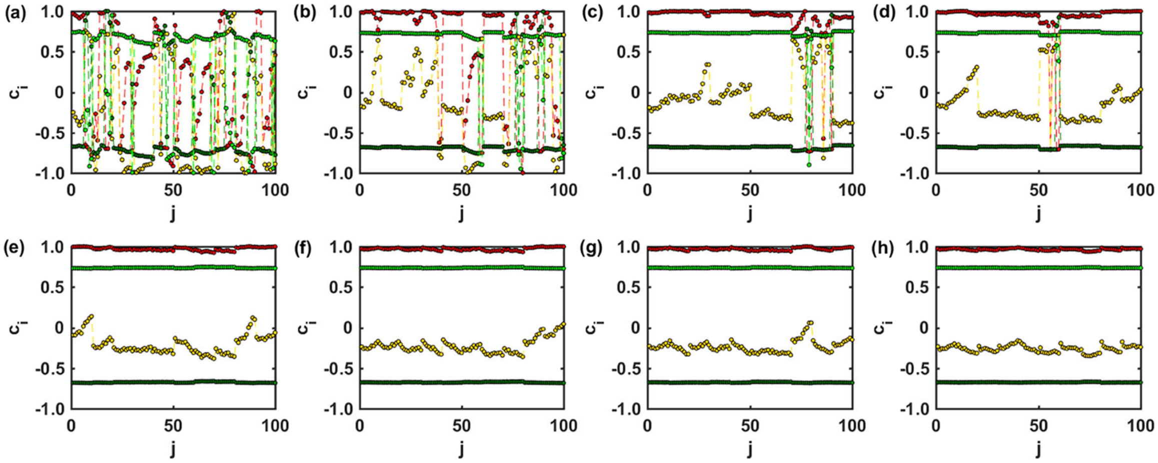


Figure 9.

Results from tICA on subtrajectories involving (a) one, (b) two, (c) three, (d) four, (e) five, (f) six, and (g) seven membrane crossings and (h) on the full biased trajectories without data reweighting. The same color scheme was used as in Figure 8 to represent the components of the first and second eigenvectors.

Table 1.

Summary of Simulations Showing the Type of Simulation, the Number of Replicas (N_{rep}), and the Length (t) of Each Replica and the CVs Biased[a]

simulations	N_{rep}	t (μs)	CVs
standard MD	1	5	
MD + WTMetaD	10	3	s_0
	10	3	s_1, s_2
	3	3	s_1', s_2'
MD + TTMetaD	1	1	rescaled z_1, z_2
MD + WTMetaD with walls	5	5	z_1, z_2

[a] z_1 and z_2 are molecular descriptors, s_0 is the suboptimal CV, s_1 and s_2 are the optimal CVs, and s_1' and s_2' denote the CVs obtained with $|z_1|$ and $|z_2|$.