THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Quantifying Energetic and Entropic Pathways in Molecular Systems

Published as part of The Journal of Physical Chemistry virtual special issue "Pablo G. Debenedetti Festschrift".

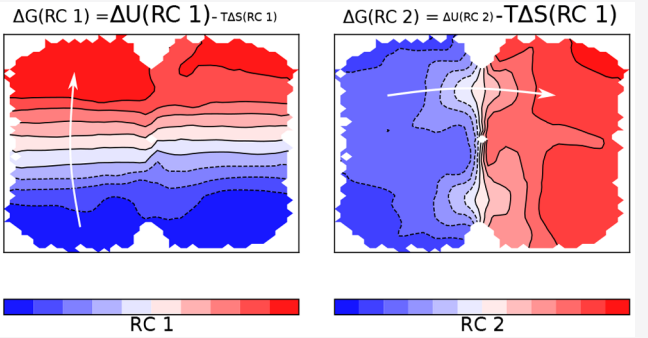
Eric R. Beyerle, Shams Mehdi, and Pratyush Tiwary*

Cite This: *J. Phys. Chem. B* 2022, 126, 3950−3960

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: When examining dynamics occurring at nonzero temperatures, both energy and entropy must be taken into account to describe activated barrier crossing events. Furthermore, good reaction coordinates need to be constructed to describe different metastable states and the transition mechanisms between them. Here we use a physics-based machine learning method called state predictive information bottleneck (SPIB) to find nonlinear reaction coordinates for three systems of varying complexity. SPIB is able to correctly predict an entropic bottleneck for an analytical flat-energy double-well system and identify the entropy- and energy-dominated pathways for an analytical four-well system. Finally, for a simulation of benzoic acid permeation through a lipid bilayer, SPIB is able to discover the entropic and energetic barriers to the permeation process. Given these results, we thus establish that SPIB is a reasonable and robust method for finding the important entropy, energy, and enthalpy barriers in physical systems, which can then be used to enhance the understanding and sampling of different activated mechanisms.



I. INTRODUCTION

The separation of driving forces for a generic chemical mechanism into their energetic and entropic components has been a topic of continued interest over the decades. This is relevant to diverse problems such as solute aggregation,[1,2] drug and ligand binding to proteins[3,4] and other biomolecules,[5,6] nucleation,[7,8] molecular permeation through membranes,;[9−11] and numerous phase transitions.[12−17] This separation provides fundamental insight into the natures of interactions that stabilize or destabilize a given material and can guide further design strategies.

For systems evolving at a nonzero temperature, the energy of the system alone does not determine the stabilities of different configurations. In addition to the change in energy, ΔU, or enthalpy, ΔH, that depends on whether the system is at constant volume or constant pressure conditions, an entropic contribution, $-T\Delta S$, driven by the system's temperature must also be taken into account. Restricting our attention for the sake of argument to systems at constant pressure that adhere to Kramers' rate law, the reaction rate at a temperature T is proportional to the exponential of the free energy barrier ΔG as follows:[18] $k \propto \exp\left[-\frac{\Delta G}{k_B T}\right] = \exp\left[\frac{-\Delta H}{k_B T}\right]\exp\left[\frac{\Delta S}{k_B}\right]$, where k_B is Boltzmann's constant. One can then disentangle entropic and enthalpic contributions by noting that the thus-defined rate can be viewed as the product of a temperature-independent term driven by entropy, with all enthalpic contributions restricted to the temperature-dependent term.

$$\ln\left(\frac{k}{k'}\right) = \frac{\Delta H}{k_B}\left(\frac{1}{T'} - \frac{1}{T}\right) \quad (1)$$

where k and k' denote rates at two temperatures T and T', respectively.

However, using eq 1 to calculate the enthalpic and entropic contributions to the free energy is burdensome, as it requires running experiments or simulations at multiple temperatures, which is likely difficult for myriad reasons. Furthermore, this approach would still not give insight into the molecular origins of the enthalpic and entropic parts of the rate constant. Thus, there is a clear need to develop theoretical and computational frameworks that can give direct molecular- and atomic-level information about entropic and enthalpic or energetic driving forces without the need to repeat expensive all-atom simulations at different temperatures.

In this work, we demonstrate a computational framework that allows the quantification of the energetic and entropic

Received: March 14, 2022
Revised: May 2, 2022
Published: May 23, 2022

contributions to a given chemical process of interest from unbiased or biased all-atom molecular dynamics (MD) simulations performed at a single temperature. Given the importance of this problem,[9,11,19−22] several other approaches have been proposed; however, we would argue that no single approach addresses both of the following key challenges under the same umbrella:

1. Chemical processes often involve multiple pathways[16,19,23−32] with differing energetic and entropic contributions. The approach should be able to quantify these contributions for the different pathways separately instead of just one overall trend and learn low-dimensional descriptors that correspond to them.
2. Many processes are effectively rare events when simulated in an all-atom femtosecond resolution. Simulating these thus requires specialized enhanced sampling methods, which work better if one has prior knowledge of approximate reaction coordinates for the different slow processes mentioned in the challenge above.[33−35]

Commonly used dimensionality reduction techniques such as time-lagged independent component analysis (TICA),[36] or Markov state models[37] can assist with the first challenge above, as they can ascertain the dominant slow modes in a given system. However, they need prior access to extensive sampling, thereby not offering a solution to the second challenge. We seek an approach that is able to both learn such slow multidimensional degrees of freedom that correspond to different pathways from preliminary biased or unbiased data and perform further biased sampling to enhance fluctuations along these pathways.[38]

Here, we utilize state predictive information bottleneck (SPIB),[39] which meets both of these challenges together. SPIB has the ability to both find relevant multidimensional reaction coordinates and perform enhanced sampling along them, even for simulations of rare events. We show here how it can be used to learn reaction coordinates for a set of systems known to possess distinct entropic and energetic barriers, clearly demarcating different pathways and their respective energetic and entropic components.

The SPIB protocol finds the relevant reaction coordinates by passing the input order parameters through a modified variational autoencoder[40] and enhances the sampling of the barrier regions by running metadynamics[41,42] along the optimized SPIB latent coordinate(s). This approach is in contrast to pure TICA, which does not intrinsically enhance the sampling along the discovered slow reaction coordinate(s), and umbrella sampling,[43] which enhances the sampling without discovering the reaction coordinate. Furthermore, SPIB, like its predecessor RAVE (reweighted autoencoded variational Bayes),[44,45] is used to iteratively update the discovered reaction coordinate to enable the sampling of rare events with a feasible amount of computational time. That is, rounds of SPIB can be performed to iteratively optimize the discovered reaction coordinates.[46]

We show that the SPIB approach was able to separate different pathways and distinguish the primarily entropic pathway from the primarily energetic pathway without much prior knowledge. We demonstrate this result for a pair of analytical potentials as well as in the description of benzoic acid permeation through a membrane. We compare the SPIB results to TICA on all problems to clearly demonstrate the advantage of using SPIB for the systems studied here. These results indicate that the SPIB approach is useful for finding the entropic and energetic reaction coordinates, even though there is no explicit accounting of either the energy or the entropy in the SPIB approach. Combined with SPIB's ability to discover and enhance the sampling along these entropic and energetic reaction coordinates when coupled with metadynamics, we propose that SPIB is a powerful protocol for sampling free energy barriers, no matter their thermodynamic origins.

II. METHODS

II.A. State Predictive Information Bottleneck (SPIB).

The formalism for SPIB was laid out in ref 39, and the method was an extension of the previously developed RAVE technique.[44,47,48] Briefly, SPIB takes as the input a set of coordinates from a time-ordered dynamical trajectory $\mathbf{X}(t)$ and finds a reduced representation of the dynamics $\mathbf{z}(t)$ that maximizes the following loss function $\mathcal{L}$, which can be seen as related to the information bottleneck loss function $\mathcal{L}_{\mathrm{IB}}$. Hence, the difference between the two mutual information terms is as follows:[39]

$$\mathcal{L}_{\mathrm{IB}} \equiv I(\mathbf{z}, \mathbf{y}) - \beta I(\mathbf{X}, \mathbf{z}) \geq \sum_{k=1}^{N} \log(p(\mathbf{y}(k+s)|\mathbf{z}(k))) - \beta \log\left(\frac{p(\mathbf{z}(k)|\mathbf{X}(k))}{p(\mathbf{z}_\theta)}\right) \tag{2}$$

Maximizing the loss function in eq 2 ensures that SPIB discovers a low-dimensional compressed representation of the input coordinates that is maximally predictive of the *state* of the system, $\mathbf{y}$, at a lagtime s in the future, $\mathbf{y}(k+s)$. The parameter $\beta \in (0,\infty)$ serves the same function as that in the deep variational information bottleneck approach,[40,49] with the traditional variational autoencoder[50] representing the special case $\beta = 1$. Tuning β governs the trade-off between how compressed the latent representation $\mathbf{z}$ is and how faithfully the latent representation is able to predict the future state of the system $\mathbf{y}(k+s)$. That is, the second term in eq 2 effectively serves as a regularization term[51] that penalizes a high-dimensional latent space.

The SPIB encoder, $p(\mathbf{z}(k)|\mathbf{X}(k))$, and decoder, $p(\mathbf{y}(k+s)|\mathbf{z}(k))$, are generated by a fully connected nonlinear neural network, as described in ref 39, with two encoding layers and two decoding layers in addition to the bottleneck layer. The parameter-informed prior $p(\mathbf{z}_\theta)$ is generated using a variation of the VampPrior proposed in ref 52 as follows:

$$p(\mathbf{z}_\theta) = \sum_{i=1}^{K} w_i p_\theta(\mathbf{z}|\mathbf{u}_i) \tag{3}$$

where K is the predicted number of states in the system and w_i is the weights of the representative inputs $\mathbf{u}_i$, each of which is, in practice, a single sample of $\mathbf{X}$ selected from each of the K states $\mathbf{y}$. The prior given in eq 3 is updated following each refinement of the network as the predicted state labels $\mathbf{y}$ are refined and updated. Furthermore, as SPIB is iterated, some of the initial state labels $\mathbf{y}(0)$ may merge depending on the lagtime s selected for the SPIB analysis. That is, as the lagtime s of the SPIB analysis increases, the anticipated number of states K will decrease. This is because faster motions, corresponding to small barrier crossings, are coarse-grained out due to the longer lagtimes. This effect of finding only the most metastable states of the system, corresponding to the highest barrier

https://doi.org/10.1021/acs.jpcb.2c01782
J. Phys. Chem. B 2022, 126, 3950−3960

crossing events, is analogous to what occurs in the Markov state modeling of dynamical systems[37] when a spectral clustering algorithm such as robust Perron cluster−cluster analysis[53] is applied to the obtained eigenvectors of the transition operator on the state space. Once the state labels $\mathbf{y}$ converge, the training process is terminated and the obtained $\mathbf{z}$ is analyzed.

II.B. Comparison of SPIB with Other Autoencoder Methods for Reaction Coordinate Discovery. At first glance, the SPIB method might appear architecturally similar to other variational autoencoder methods for conformational dynamics, such as the time-lagged autoencoder,[54] molecular enhanced sampling with autoencoders,[55] and RPNet,[56] among others; good surveys of autoencoder methods for finding low-dimensional latent representations of molecular dynamics are refs 48 and 57.

It is its ability to learn on-the-fly not only the reaction coordinate but also the number of metastable states and their locations that sets SPIB apart from these other methods. In contrast to the typical variational autoencoder framework, where the discovered latent space minimizes either the mean-squared or predicted error for the prediction of the input coordinates at some lagtime s in the future, $\mathbf{x}(t+s)$, SPIB aims to predict the *state* of the system at some lagtime s in the future, $p(\mathbf{y}(k+s)|\mathbf{z}(k))$; thus, in addition to finding optimal reaction coordinates, SPIB simultaneously discovers the slow transitions to the most metastable states of the system and the definitions of these states, similar to the combination of Markov state modeling and subsequent macroscopic clustering.[37,53,58] That is, as shown in ref 39, the SPIB decoder returns a "fuzzy" probability transition matrix, $p(\mathbf{y}(t+s)|\mathbf{z}(t))$, that can be used to estimate the slowest transition modes in the system identically to a Markov state model analysis.

II.C. Calculating Entropy and Other State Functions from SPIB. For analytical potentials or systems in a vacuum, the energy $U(t) = U(\mathbf{X}(t))$ of the system at any time t in the simulation can be directly calculated using the system's state $\mathbf{X}(t)$ at time t. For solvated systems, the situation is a bit more complicated, as both the solute intramolecular and solute−solvent intermolecular potentials must be taken into account. Following the procedure described in ref 59, the change in energy from a reference state, ΔU, is given by

$$\Delta U = \Delta U_{\text{intra}} + \Delta U_{\text{inter,SR}} \tag{4}$$

where ΔU_{intra} and $\Delta U_{\text{inter,SR}}$ denote the energy changes due to solute−solute intramolecular and short-range solute−solvent intermolecular interactions, respectively. In typical MD simulations, $\Delta U_{\text{inter,SR}}$ comes from Lennard-Jones and electrostatic interactions within the cutoff distance specified for the simulation, which is typically around 1 nm for solvated systems. For particle-mesh Ewald calculations of the electrostatics, only solute−solvent interactions from the short-ranged term are utilized to calculate ΔU. Furthermore, for systems simulated in the constant number, pressureb and temperature (NPT) ensemble, the change in the internal energy due to $P-V$ work performed by the barostat must be included. Here the solute enthalpy ΔH is calculated instead of ΔU as follows:

$$\Delta H = \Delta U_{\text{intra}} + \Delta U_{\text{inter,SR}} + P\Delta V \tag{5}$$

where $P\Delta V$ is the change in energy due to the work performed by the barostat; the reference volume used to calculate ΔV is the volume of the box at the beginning of the simulation.

In systems of practical interest with high free energy barriers, the sampling in the barrier regions can be noisy. Here, instead of using a regular histogram to calculate the free energy and the enthalpy along the reaction coordinate (RC), a kernel density estimate (KDE)[60] of the probability distribution along each reaction coordinate is used, with the choice of Gaussian kernel. In KDE with a Guassian kernel, the probability distribution along a RC is estimated at each value z of the RC using a sum of Gaussian basis functions as follows:[60]

$$p(z) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{\sqrt{2\pi h^2}}\exp[-(z - z(k))^2/2h^2] \tag{6}$$

where h is the selected bandwidth and N is the number of frames in the trajectory. Choosing a KDE over a regular histogram effectively amounts to binning with a Gaussian basis set in place of binning with a basis set of indicator functions; using the Gaussian basis set allows for a smoother estimate of the probability density in regions of the free energy surface where the sampling is noisy, which is the case for systems that describe rare events where the transition region is sampled infrequently.

To decompose the contribution of the entropic and energetic components along a generic RC z, we first define the general definition of geometric free energy along a reaction coordinate, $G(z)$, as follows:[61]

$$G(z) = -k_{\text{B}}T\ln\left(\int_{R^n} e^{-U(\mathbf{x})/k_{\text{B}}T}\times\delta(\Phi(\mathbf{x}) - z)\det(\tilde{G})^{1/2}\,d\mathbf{x}\right) \tag{7}$$

where $\Phi(\mathbf{x})$ is the desired level set of the reaction coordinate where we desire to calculate the geometric free energy, $\delta(x)$ is the Dirac δ-function, and $\tilde{G}$ is the Gram matrix of the transformation $\Phi(\mathbf{x}) = \mathbf{x} \rightarrow \mathbf{z}$. The Gram matrix can be decomposed into the product of the Jacobian matrix, $\nabla\mathbf{z}$, and its transpose[62]

$$\tilde{G} = (\nabla\mathbf{z}^T)\nabla\mathbf{z} \tag{8}$$

where the Jacobian matrix describes how the n-dimensional input space $\mathbf{x}$ is stretched or squeezed as it is transformed nonlinearly via the SPIB-defined neural network to generate the m-dimensional reaction coordinate $\mathbf{z}$.

The energy along each level set of $\mathbf{z}$, $U(\mathbf{z})$, can be calculated using the formula for averaging over each level set of the reaction coordinate as follows:[61]

$$\begin{aligned}\langle U(z)\rangle_{\Sigma(z)} :&= U(z)\\ &= \frac{1}{N_z}\int U(\mathbf{x})e^{-U(\mathbf{x})/k_{\text{B}}T}\times\delta(\Phi(\mathbf{x}) - z)\det(\tilde{G})^{1/2}\,d\mathbf{x}\end{aligned} \tag{9}$$

where $\Sigma(z)$ is the submanifold of the input coordinates constrained to the given value of the reaction coordinate z.[61] That is, $\Sigma(z)$ contains the set of all high-dimensional input coordinates in the trajectory mapped to $z \in (z-\epsilon, z+\epsilon)$ by the SPIB encoder for some small-enough value of $\epsilon > 0$. Finally, N_z is a normalization constant equal to the number of frames in the trajectory that are mapped to the submanifold $\Sigma(z)$ augmented by the Jacobian

$$N_z = \sum_{k=1}^{N} I_z(X(k))\det(\tilde{G})^{1/2} \tag{10}$$

https://doi.org/10.1021/acs.jpcb.2c01782
J. Phys. Chem. B 2022, 126, 3950−3960

where $I_z(X(k))$ is an indicator function over Σ_z, which is equal to 1 if $X(k)$ maps to Σ_z and equal to 0 otherwise. Finally, the entropy along z is calculated using the thermodynamic identity $\Delta G(z) = \Delta U(z) - T\Delta S(z)$, where $\Delta U(z) = U(z) - U(z_{ref})$ and $\Delta G(z) = G(z) - G(z_{ref})$. The reference value z_{ref} is defined as the value of z that minimizes $U(z)$, $z_{ref} = \arg\min_z U(z)$. Since $-T\Delta S(z)$ gives the entropic contribution to the free energy barrier along z, we will plot $-T\Delta S(z)$ when evaluating the entropic contribution to the free energy barrier along a given SPIB reaction coordinate.

$$\Rightarrow \Delta S(z) = \frac{1}{T}(\Delta U(z) - \Delta G(z))$$

II.D. Analytical Potentials. We start by considering two analytical potentials. The first system examined is the entropic double-well system described in ref 63, which possesses only an entropic "bottleneck" between two diffusive wells of equal area. This potential energy surface is shown as a contour plot in Figure 1a and is defined as follows:

$$U(x, y) = x^6 + y^6 + \exp[-y^2/\sigma_y^2](1 - \exp[x^2/\sigma_x^2]) \quad (11)$$

where $\sigma_x = \sigma_y = 0.1$ defines the width of the wells.

The second analytical model is a four-well system whose slowest dynamics changes from crossing a predominately entropic barrier to crossing a predominately energetic barrier, as described in ref 64. This potential energy surface is shown as a contour plot in Figure 1(b) and is defined as follows:

$$U(x, y) = h_x(x^2 - 1)^2 + (h_y + a((x, \delta))(y^2 - 1)^2 \quad (12)$$

where $h_x = 0.5$ and $h_y = 1.0$ describe the well width in the x- and y-directions, respectively; $\delta = 0.05$ describes how much the barrier-crossing pathway along the x-direction is squeezed relative to the pathway along the barrier in the y-direction; and $a(x, \delta) = \frac{1}{5}(1 - 5\exp[-(x - x_0)^2/\delta])$.

Simulation details regarding the trajectory length in integration timesteps, the temperature, and the friction coefficient γ for the entropic double-well and temperature-switch potentials are given in Table I. For both analytical potentials, the integration time step was 0.001 units, and the trajectories in each case were saved to file every $\Delta t = 10$ integration steps, giving a total of 1×10^6 and 3×10^6 frames for analysis. The SPIB parameters for both the analytical potentials are given in the Supporting Information. Since the potential was known a priori, the SPIB neural network was trained to find a one-dimensional reaction coordinate for the entropic double well-system and a two-dimensional reaction coordinate for the temperature-switch system.

II.E. Benzoic Acid Permeation through a DMPC Membrane Bilayer. As a more challenging problem, we consider small-molecule permeation through membranes, which is an important process for determining the efficacy of pharmaceuticals[65] as well as other biological processes.[66,67] However, it is known that for a variety of small molecules the membrane permeation cannot be adequately described using a single reaction coordinate.[66,67] Recently it was shown that one could use SPIB to find an adequate reaction coordinate for enhancing the sampling of benzoic acid (BA) crossing a 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) membrane bilayer.[46]

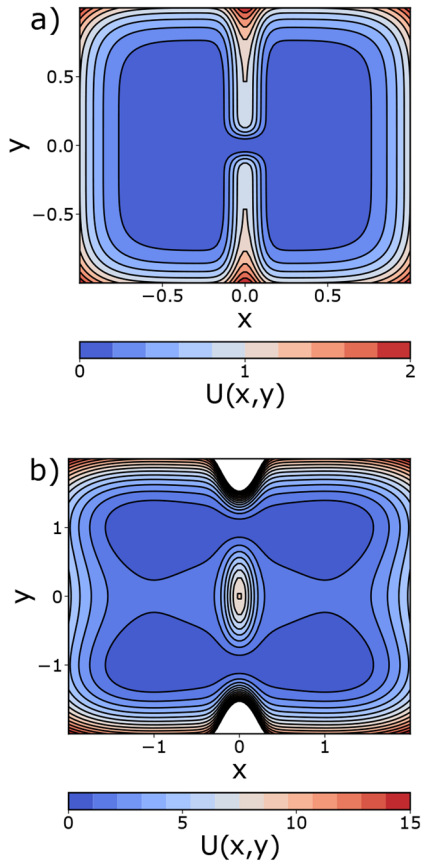


Figure 1. (a) Potential energy surface for the entropic double-well potential (eq 11). The entropic bottleneck between the two wells is centered at $(x, y) = (0.0, 0.0)$. (b) Potential energy surface for the temperature-switch potential (eq 12). Crossing the two channels between the left- and right-hand sides of the potential corresponds to surmounting a barrier that is primarily entropic, while crossing the two (broader) channels between the top and bottom wells corresponds to surmounting a barrier that is primarily energetic. The "masked" (white) regions of the landscape correspond to regions of the potential energy surface where $U(x, y) > 15$.

Table I. Simulation Details for the Entropic Double Well (EDW) and Temperature-Switch (TS) Potentials

parameter	EDW	TS
integration steps	1×10^7	3×10^7
$(k_BT)^{-1}$	10	1
γ	4.0	0.5

Building upon the analysis reported in ref 46, we use a set of generic 21 order parameters (OPs) as the input for the SPIB analysis. These are the distance vectors defined from the center of mass (COM) of the aromatic ring in BA to the COM of the lipid bilayer ($\vec{d}_1$), from the hydroxl oxygen of BA to the COM of the bilayer ($\vec{d}_2$), from the carbonyl oxygen of BA to the bilayer COM ($\vec{d}_3$), from the COM of the benzene ring to the hydroxyl oxygen ($\vec{d}_4$), and between the COMs of the two leaflets of the bilayer ($\vec{d}_5$). The x, y, and z-components of $\vec{d}_1$, $\vec{d}_2$, and $\vec{d}_3$ constitute the first nine OPs. Furthermore, the sines and cosines of the three angles made by $\vec{d}_4$ (θ_x, θ_y, θ_z) and $\vec{d}_5$ (ω_x, ω_y, ω_z) with the x-, y-, and z-axes of the simulation box are taken as the additional 12 OPs, for a total of 21 OPs. As described in ref 46, the input to the SPIB analysis is a pair of 25 ns long trajectories, where the benzoic acid is initially placed

https://doi.org/10.1021/acs.jpcb.2c01782
J. Phys. Chem. B 2022, 126, 3950−3960

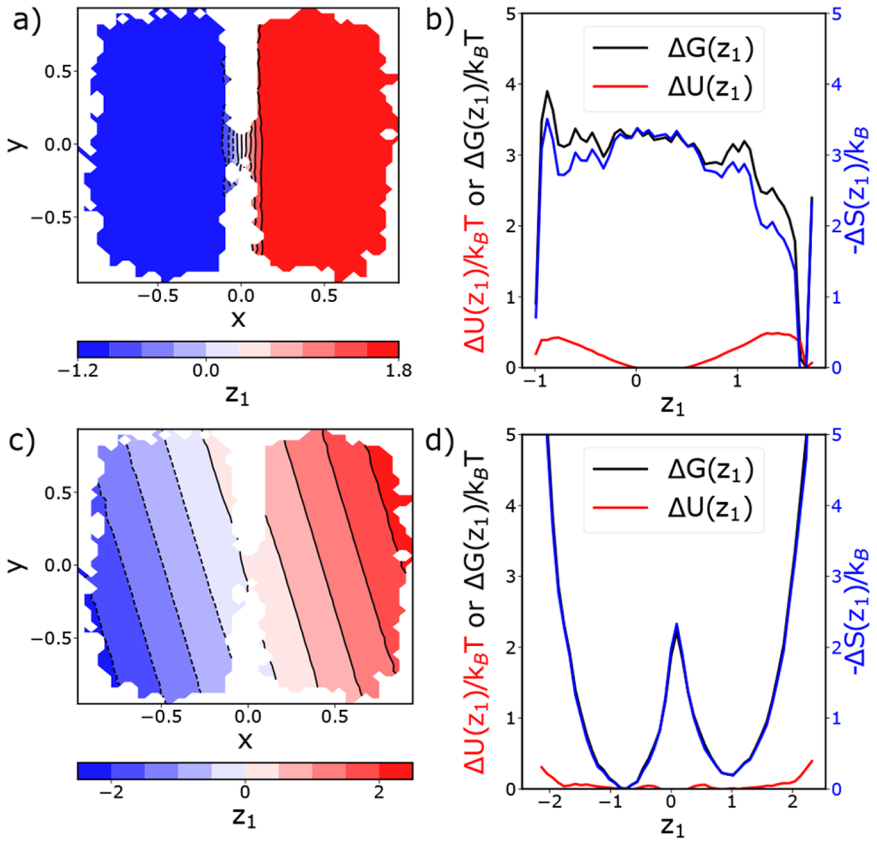


Figure 2. (a) Projection of the one-dimensional SPIB coordinate, z_1, onto the free energy surface for the entropic double-well. (b) Plot of $\Delta G(z_1)/k_BT$ (black), $\Delta U(z_1)/k_BT$ (red), and $-\Delta S(z_1)/k_B$ (blue) for the SPIB reaction coordinate. Since the reaction coordinate "expands" the bottleneck region in the latent space, the free energy barrier, which is narrow in (x, y) space, is extended in z_1 space. (c) The slowest TIC projected onto the entropic double-well potential. In contrast to SPIB, the gradient of the TIC is almost uniform throughout the free energy surface. (d) Thermodynamic decomposition of the free energy barrier along the slowest TICA coordinate. Comparing panel d with panel b shows that TICA predicts results qualitatively similar to those of SPIB, although the width of the barrier along the slowest TIC is much narrower than that in the SPIB latent space. This is again due to the uniform gradient of the slowest TIC.

on the positive side of the bilayer membrane in one trajectory and on the negative side in the other. A 500 ns biased simulation along an optimal SPIB one-dimensional reaction coordinate found using this initial pair of 25 ns unbiased molecular dynamics (MD) simulations was analyzed using SPIB. Full simulation details can be found in the SI or in ref 46. A snapshot from the simulation showing the physical setup of the system, with water molecules excluded, is given in Figure S1 in the SI.

To perform the enthalpy−entropy decomposition for this system, the enthalpy was calculated using eq 5. All energies, the pressure, and the volume were calculated using the gmx energy module in GROMACS 2020.2.[68] To calculate the free energy, enthalpy, and entropy profiles along each of the two SPIB RCs, the KDE method outlined in Section II.C was used, with bandwidths for the first and second RC of $h = 0.5$ and $h = 1.0$.

III. RESULTS

III.A. Entropic Double-Well System. *III.A.1. SPIB Discovers the Slow Entropy-Dominated Process.* The relevant results for the entropic double-well system are presented concisely in Figure 2, where Figure 2a shows the one-dimensional SPIB coordinate, z_1, for the two-dimensional entropic well system projected on the underlying (x, y) space. SPIB finds a RC that is constant inside the diffusive and energetically constant wells to the left and right of $x = 0.0$, while z_1 changes very rapidly inside the bottleneck between the wells. This sign change of z_1 inside the bottleneck indicates that z_1 describes transitions between the diffusive wells through the bottleneck as the relevant slow process that occurs in the system. The profiles of $\Delta G(z_1)$, $\Delta U(z_1)$, and $T\Delta S(z_1)$ along the SPIB coordinate z_1 for the entropic double-well system are plotted in Figure 2b. It is clear that the transition through the bottleneck is entropy-driven. However, since the real-space volume of the wells is large (i.e., a single level set of z_1 in Figure 1a covers the left- and right-hand wells), the largest values of z_1 are favorable entropically. We would like to point out that edge effects in $\Delta U(z_1)$ are due to thermal fluctuations pushing the system slightly up the sextic potential when sampling is performed outside the bottleneck.

III.A.2. Comparing SPIB with TICA. For comparison to a "reference" method used to extract the slow dynamics of a system, we ran TICA[36,69] on the same trajectory fed to the SPIB. The theory and applications of TICA to physical systems have been covered extensively over the past decade.[69−72]

As with the SPIB analysis of the entropic double-well trajectory, only a single TIC was output from TICA, which was run using a lagtime of $\tau = 10\Delta t$. The single TIC projected onto the entropic double-well potential is given in Figure 2c. At first glance, the result is similar to the SPIB result. The slowest TIC predicts transitions between the left and right wells through the bottleneck to be the slowest processes in the system. However,

https://doi.org/10.1021/acs.jpcb.2c01782
J. Phys. Chem. B 2022, 126, 3950−3960

the SPIB and TICA results differ in a crucial aspect. Due to its nonlinear nature, SPIB gives a much superior resolution in the entropic bottleneck region. On the other hand, the gradient of the slowest TIC is constant over the entire potential energy surface, as it is a linear method. Thus, the input space is neither stretched nor squeezed during TICA. In contrast, SPIB's nonlinearity allows for compression inside the bottleneck and expansion in the wells, where the dynamics are fast and mostly orthogonal to the slow barrier-crossing process. From the thermodynamic perspective, TICA is on par with the SPIB in predicting both that the barrier to transition between the wells is completely dominated by the entropy contribution. This agreement is demonstrated by comparing the results for SPIB in Figure 2(d) with the results from TICA in Figure 2(b).

III.A.3. Committors from SPIB and MSM. Finally, to quantitatively assess the quality of the SPIB-predicted reaction coordinate for this system, the committor function[37,73,74] was calculated using a Markov state model (MSM)[75] with a lagtime $\tau_{MSM} = 10\Delta t$. The MSM was constructed on the (x, y) state space discretized into 500 states using a regular space clustering algorithm.[76,77] To define the committor function on this discrete space, the "reactant state" or "state A" was defined as the set of all (x, y) mapped to $\hat{z}_1 = \frac{z_1 - \min(z_1)}{\max(z_1) - \min(z_1)} \leq 0.02$ by the SPIB neural network, and the "product state" or "state B" was defined as the set of all (x, y) mapped to $\hat{z}_1 > 0.98$. z_1 was transformed to its "min-max" version $\hat{z}_1$ because for two-state systems the committor function should map onto this rescaled version of the coordinate that describes the crossing between the wells.[78,79]

Figure 3 compares $\hat{z}_1$, i.e., a linearly scaled version of the SPIB coordinate, and the committor function calculated using

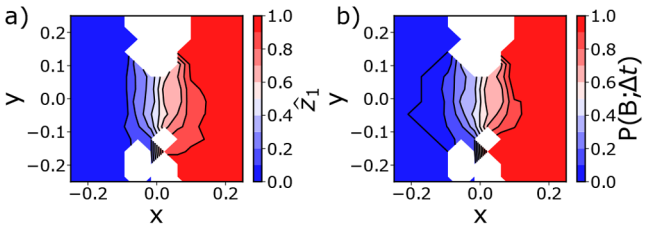


Figure 3. (a) Projection of $\hat{z}_1$ in the bottleneck region of the entropic double-well potential energy surface. (b) Projection of the committor function with 500 discrete states, predicted from the MSM, and lag $\tau_{MSM} = 10\Delta t$ onto the bottleneck region of the entropic double-well potential energy surface.

the MSM. Panels a and b in Figure 3 emphasize the comparison in the bottleneck or transition region of the free energy surface, where both $\hat{z}_1$ and the committor function change rapidly. With the exception of some "bowing" of the isocommittor surfaces at the entrance to the bottleneck, there is little quantitative difference between $\hat{z}_1$ and the committor function calculated using the MSM. This result indicates that SPIB is able to accurately and directly learn the committor function for this analytical system with an entropic barrier.

III.B. Temperature-Switch Potential. *III.B.1. SPIB Separates the Entropy- and Energy-Dominant Pathways.* A more challenging two-dimensional analytical system is given by the temperature-switch potential presented in ref 64. As shown in ref 64, an interesting feature of this system is that the slowest process changes from crossing the energetic barrier along the y-component at low temperatures (where the ΔU contribution to ΔA dominates) to crossing the entropic barrier at high temperatures (where the $-T\Delta S$ term in ΔA dominates), hence the moniker "temperature-switch" for this system. Here we are testing if SPIB can select the correct two-dimensional reaction coordinate that captures these two different energetic and entropic pathways at an arbitrary temperature of $(k_B T)^{-1} = 1.0$.

Panels a and b of Figure 4 show both the SPIB latent space coordinates projected individually onto the (x, y) coordinates. Figure 4a shows that z_1 describes transitions across the system's energetic barrier in the y-direction, while Figure 4b shows that z_2 describes transitions across the entropic barrier in the x-direction. As in the previous double well system, here SPIB demarcates the entropic bottleneck while also distinguishing the two possible pathways. Overall, on the basis of the SPIB decomposition presented in Figure 4, we conclude that SPIB is able to select two reaction coordinates that nearly completely separate the dynamics corresponding to crossing the energetic and entropic barriers in the system. This statement was further quantified by decomposing the free energy barrier into its energetic and entropic components. Panels c and d in Figure 4 show the analogous decomposition for both SPIB coordinates that was discovered for the temperature-switch potential. Along z_1 the change in free energy is almost entirely due to the potential energy, while along z_2 the change in free energy is almost entirely due to entropy, justifying their labeling as the energetic component and the entropic component, respectively. In addition, the nonlinear SPIB approach allows for greater resolution of the transition regions, where the slow processes occur, and decreased resolution in the free energy wells, where the "uninteresting" fast processes occur.

III.B.2. Comparing SPIB with TICA. As with the entropic double-well potential, TICA was performed on the trajectory for this system (see the SI for results). We find that while the barrier decomposition from TICA is qualitatively similar to the SPIB results, the TICA coordinates possess a linear gradient along the original (x, y) coordinates and thus gives a poorer resolution of the barrier regions compared to the SPIB reaction coordinates. Thus, we conclude that for this system as well it is important to have a nonlinear reaction coordinate to describe the slow dynamics in the vicinity of the transition state.

III.C. Benzoic Acid Membrane Permeation through the Phospholipid Bilayer. The third and final system studied here is the permeation of benzoic acid through a DMPC membrane (BA-DMPC). For this system, the reaction coordinate and the energetic and entropic components are not clear a priori. The problem is further complicated because the permeation of a small molecule through a lipid bilayer is a slow process, usually requiring enhanced sampling methods to calculate transport properties and permeation rates.[10,66,80] As such, the permeation of benzoic acid through a DMPC bilayer is an adequate example with which to test SPIB's ability to both find and accelerate the sampling along the system's reaction coordinates as well as separate the entropy and energy barriers to permeation.

III.C.1. Free Energy along Intuitive Physical Projections. For the BA-DMPC system, the number of input coordinates to SPIB is 21 (see the SI and ref 46), with a two-dimensional SPIB latent space. To seed the initial SPIB state labels, regular space clustering as implemented in PyEMMA2[77] was implemented in the two-dimensional space spanned by (1) the distance from the COM of the aromatic ring in benzoic acid to the COM of the membrane bilayer, $d_{1,z}$ and (2) the

https://doi.org/10.1021/acs.jpcb.2c01782
J. Phys. Chem. B 2022, 126, 3950−3960

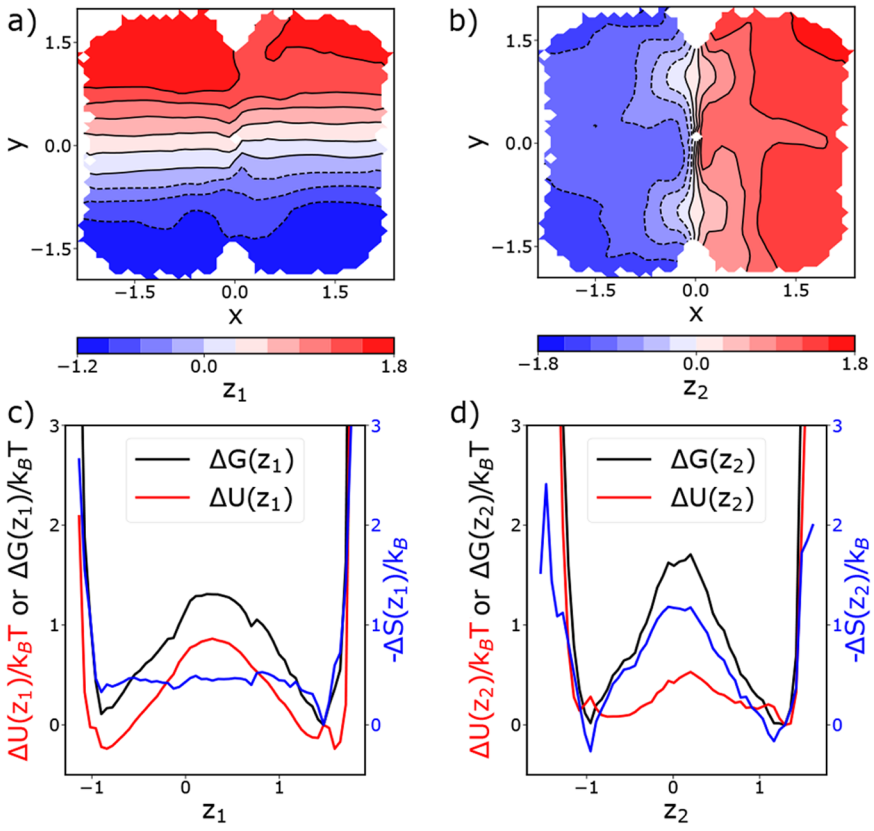


Figure 4. (a) Projection of z_1 onto the free energy surface of the temperature-switch system. At the given temperature, the first SPIB coordinate describes crossing of the energetic barrier in the *y*-direction. Panel b: projection of z_2 onto the free energy surface of the temperature switch system. At the given temperature, the second SPIB coordinate describes the crossing of the entropic barrier in the *x*-direction. (d) Decomposition of the free energy profile (black) along z_1 into its energetic (red) and entropic (blue) components. Based on this decomposition, it is clear that the main contribution to the barrier along z_1 is from the energetic component of the free energy, although there is some "baseline" contribution from the entropy barrier. (d) Decomposition of the free energy profile (black) along z_2 into its energetic (red) and entropic (blue) components. Based on this decomposition, it is clear that the main contribution to the barrier along z_2 is from the entropic component of the free energy, although there is a small contribution from the energy barrier.

angle between the bilayer normal and a vector pointing from the COM of the aromatic ring of benzoic to the COM of the hydroxyl oxygen in the carboxylic acid functional group, θ_z. All other parameters for the SPIB analysis are given in Table S1 of the SI.

Recent research has shown that a one-dimensional projection along the *z*-coordinate does not sufficiently describe the permeation process of small molecules through a lipid membrane.[67] For example, from the two-dimensional free energy surface given in Figure 5, to cross inside the membrane, it is clear that the angular values of θ_z, on average, must be restricted so that the −COOH group in benzoic acid points toward the phospholipid head groups, resulting in an energetically favorable dispersion interaction. However, the restriction of the conformational freedom of the ring after it enters the membrane compared with its orientational freedom outside the membrane represents an entropic barrier to membrane crossing. In contrast, the benzoic acid crossing the COM of the bilayer represents an energetic barrier, since the benzoic acid must lose its favorable dispersion interaction with the headgroup moieties to cross the bilayer. The free energy along these two physical OPs $d_{1,z}$ and θ_z is given in Figure 5. This free energy was obtained from a 500 ns biased well-tempered metadynamics simulation along the optimized one-dimensional SPIB reaction coordinate described in ref 46.

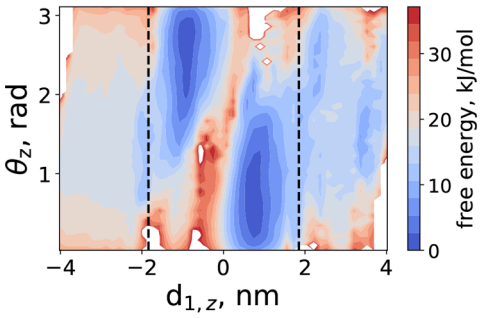


Figure 5. Free energy surface spanned by $d_{1,z}$ and θ_z for the BA-DMPC system. Vertical dashed magenta lines are drawn at the average locations of the centers of mass of the phosphorus atoms in the lipid headgroups for the top and bottom layers of the bilayer.

Since this is a biased trajectory, the contribution from each frame was reweighted appropriately[42,81] to construct the free-energy surface in Figure 5. We emphasize that SPIB was given a larger set of 21 OPs and discovered mechanistically relevant low-dimensional projections on its own.

III.C.2. SPIB Separates the Entropy and Enthalpy Barriers. Panels a and b in Figure 6 show the projection of the two latent space coordinates learned by an SPIB analysis of the biased 500 ns trajectory onto the two physical parameters

https://doi.org/10.1021/acs.jpcb.2c01782
J. Phys. Chem. B 2022, 126, 3950−3960

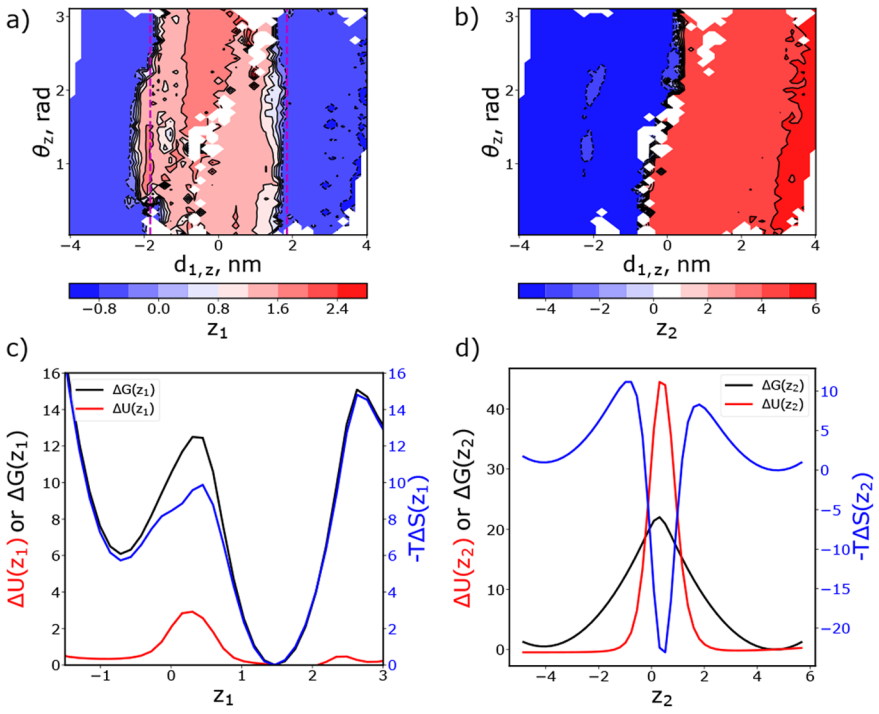


Figure 6. (a) Projection of z_1 from the SPIB analysis of the weighted BA-DMPC trajectory onto the surface spanned by the OPs $d_{1,z}$ and θ_z. (b) Projection of z_2 from the SPIB analysis of the weighted BA-DMPC trajectory onto the surface spanned by the OPs $d_{1,z}$ and θ_z. (c) Projection of the free energy (black), the enthalpy (red), and the entropy (blue) along z_1. The entropy profile nearly traces the free energy profile, indicating that the process described by z_1 is entropy-dominated. (d) Projection of the free energy (black), the enthalpy (red), and the entropy (blue) along z_2. Here, the free energy barrier is dwarfed by the enthalpic contribution, and there is an entropy−enthalpy compensation effect at the barrier due to the larger accessible volume in the middle of the membrane.[10,11] ΔG, ΔU, and $-T\Delta S$ are in units of kilojoules per mole (kJ/mol).

introduced in section III.C.1. Panels c and d in Figure 6 show the decomposition of the thermodynamic barrier along the two SPIB coordinates for the BA-DMPC system. Figure 6a shows that the first SPIB coordinate z_1 changes sign on both sides of the phospholipid bilayer, indicating this coordinate describes benzoic acid's entry into the bilayer. The decomposition of the free energy along z_1 into energy and entropy, as shown in Figure 6c, establishes that z_1 describes the entropic process of the permeation mechanism, comprising ligand diffusion into the membrane and subsequent reorientation. The small energetic barrier along z_2 likely corresponds to the unfavorable interactions between the nonpolar aromatic ring of benzoic acid as it passes through the polar phosolipid headgroups and into the interior of the bilayer.

Figure 6b shows that the second SPIB coordinate z_2 corresponds to benzoic acid passing through the center of the bilayer. Figure 6d establishes that along z_2 the free energy barrier is dwarfed by the enthalpic barrier due to an entropy−enthalpy compensation effect when benzoic acid reaches the center of the bilayer. This entropy−enthalpy compensation occurs because there is both a larger free volume for the benzoic acid to occupy[10,11] in the center of the membrane and no preferential oritentaion of the benzoic acid with respect to the membrane normal due to the loss of the favorable dispersion forces between the benzoic acid and the phospholipid headgroups. Thus, for this system SPIB was able to separate the majority of the enthalpic and entropic processes from each other.

III.C.3. Comparing SPIB with TICA. As with the other systems, the SPIB results are compared with the two slowest TICA coordinates for the same trajectory in the SM. We find that the slowest TICA coordinate essentially corresponds to dynamics along the $d_{1,z}$ coordinate and contains the relevant energetic and entropic barriers in the system. However, most strikingly (Figure S3 in the SI), the second TICA coordinate is a fairly nonphysical coordinate that describes dynamics almost strictly along θ_z, which does not surmount any relevant barriers in the $(d_{1,z}, \theta_z)$ coordinate space.

We postulate that this second TICA coordinate is "confused" by some slow, but likely irrelevant, process that occurs in one or more of the other 18 OPs. This effect is related to the point made previously by others that TICA is susceptible to catching slow correlated motions that happen to be irrelevant to the interesting dynamics.[82] That is, the second TIC must be capturing an irrelevant slow process that has already been projected out in the other 19 OPs and is noise when projected to $(d_{1,z}, \theta_z)$ space.

Overall, we see that TICA was not able to find separate coordinates to describe the entropy and energy barrier crossings in the system, and the second TICA coordinate was not readily interpretable in $(d_{1,z}, \theta_z)$. This outcome would have repercussions when TICA coordinates were used to perform additional rounds of metadynamics or other biased sampling calculations.

Finally, in the SI we give a comparison between the SPIB analysis and a MSM constructed on the two-dimensional $(d_{1,z}, \theta_z)$ space. We show that while the first MSM coordinate and the second SPIB coordinate are qualitatively similar, the MSM cannot find a single coordinate that describes permeation into the membrane. We also perform a committor analysis using the MSM (analogous to that done in Figure 3) and show that SPIB is once again able to find a latent coordinate that matches a

https://doi.org/10.1021/acs.jpcb.2c01782
J. Phys. Chem. B 2022, 126, 3950−3960

committor between suitably well-defined reactant and product states.

IV. DISCUSSION AND CONCLUSIONS

In this work we have shown that the state predictive information bottleneck (SPIB) method is able to extract reaction coordinates (RCs) for systems with energetic, entropic, or mixed barriers. The method was demonstrated to work with biased or unbiased simulations and quantified the precise enthalpic, energetic, or entropic contributions to a given activated pathway. Our results show the effectiveness of using a nonlinear method, here SPIB, to find RCs, as competing linear methods often do not do as satisfactory of a job. The separation of reaction coordinates into energetic and entropic components should be important for intelligently performing enhanced sampling for such systems; path-based methods such as forward-flux sampling[83] or milestoning[63] can be used to push the permeant over the entropic barrier, and adaptive biasing methods such as metadynamics[41] or umbrella sampling[43] can be used to push the permeant over the enthalpic barrier at the center of the bilayer.

An additional advantage of using the nonlinear RCs from SPIB is the "adaptive resolution" of the RC, with the higher resolution of the transition states and the lower resolution of the energetic wells compared to the linear TICA method; this effect was seen for all three systems examined here (viz. Figures 2a and c, 4a and b, 6a and b, S2a and b, and S3a and b). This increased resolution in the transition state is important because it allows a high-fidelity reproduction of the committor function (Figure 3) and should allow for better sampling of the transition state during subsequent enhanced sampling simulations biased along the nonlinear RC.

This increased resolution should allow the biased simulation to give better detail regarding the physical and chemical mechanisms occurring at the transition state because the resolution there is finer. That is, the linear method is suitable for finding the qualitatively correct RC, but a nonlinear RC should give better quantitative insight, especially regarding dynamics at the transition state. This effect is similar to what was seen in the Markov state modeling, where increasing the resolution of the indicator function basis set in the transition region greatly reduced the error in the approximation of the slow dynamics and hence resulted in a better model that displayed more Markovian dynamics.[76]

The BA-DMPC system is a good example of both the importance of using a nonlinear RC and SPIB's ability to separate the entropic and enthalpic contributions to a single reaction mechanism, in this case the permeation of a small molecule through a lipid bilayer. When the linear TICA method was used to find the RCs for this system, it found one useful RC that described the transition from one side of the bilayer to the other but failed to separate the entropy and enthalpy barriers along this reaction path. Instead, it lumped them into a single slow RC (Figure S3a). In contrast, SPIB was able to sift the permeation mechanism into the entropic process (entering and exiting the membrane bilayer) and the enthalpic process (benzoic acid moving from the underside of one leaflet to the underside of the other leaflet, coupled with a reorientation of benzoic acid's −COOH moiety).

Finally, the ability of the SPIB to separate the entropic and enthalpic barriers in a system such as BA-DMPC could still have been partly serendipitous, since these two types of barriers happen to separate cleanly for this system. For physical systems where the entropy and energy barriers are more entangled, it cannot be expected that the SPIB will perform so well at distinguishing the thermodynamic origin of the barrier a priori. This shortcoming of the method can be circumvented by adding an extra term to the loss function that explicitly forces one RC to traverse a pathway with the maximum entropy change and another the pathway with the maximum energy or enthalpy change. This exciting avenue for adding physics-based constraints to train SPIB will be explored in future work.

V. DATA AVAILABILITY

All codes and MD trajectories used to perform the analysis are available following a reasonable request to the authors. The code to perform the SPIB analysis is available on GitHub at https://github.com/tiwarylab/State-Predictive-Information-Bottleneck. Codes to calculate the Jacobian and make the plots in the main text for the entropic double-well and temperature-switch systems can also be found in the following GitHub repository: https://github.com/erb24/jacobian.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.2c01782.

Physical description of the BA-DMPC system, details of the SPIB parameters for each analysis demonstrated in the manuscript, comparison between SPIB and TICA for the temperature-switch and BA-DMPC systems, and MSM analysis for the BA-DMPC system (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Pratyush Tiwary − *Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, United States;* orcid.org/0000-0002-2412-6922; Email: ptiwary@umd.edu

Authors

Eric R. Beyerle − *Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20740, United States*

Shams Mehdi − *Biophysics Program and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, United States;* orcid.org/0000-0002-4078-7501

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.2c01782

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was entirely supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, CPIMS Program, under Award DE-SC0021009. Computational resources were provided by Deepthought2, MARCC, and XSEDE[84] (projects CHE180007P and CHE180027P). The authors thank Dedi Wang for useful discussions and feedback regarding the SPIB analysis as well as a critical reading of the manuscript. The authors also thank Sun-Ting Tsai for

supplying the codes for performing the Langevin simulations for the entropic double-well and temperature switch-systems as well as a critical reading of the manuscript.

REFERENCES

(1) Mondal, J.; Yethiraj, A. Driving Force for the Association of Amphiphilic Molecules. *J. Phys. Chem. Lett.* **2011**, *2*, 2391−2395.
(2) Choudhury, N.; Pettitt, B. M. Enthalpy-entropy contributions to the potential of mean force of nanoscopic hydrophobic solutes. *J. Phys. Chem. B* **2006**, *110*, 8459−8463.
(3) Freire, E. Do enthalpy and entropy distinguish first in class from best in class? *Drug discovery today* **2008**, *13*, 869−874. PMID: 18703160.
(4) Ladbury, J. E.; Klebe, G.; Freire, E. Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat. Rev. Drug Discovery* **2010**, *9*, 23−27.
(5) Jen-Jacobson, L.; Engler, L. E.; Jacobson, L. A. Structural and thermodynamic strategies for site-specific dna binding proteins. *Structure* **2000**, *8*, 1015−1023.
(6) Starikov, E.; Nordén, B. Entropy-enthalpy compensation as a fundamental concept and analysis tool for systematical experimental data. *Chem. Phys. Lett.* **2012**, *538*, 118−120.
(7) Black, S. Simulating nucleation of molecular solids. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2007**, *463*, 2799−2811.
(8) Radhakrishnan, R.; Trout, B. L. Nucleation of hexagonal ice (ih) in liquid water. *J. Am. Chem. Soc.* **2003**, *125*, 7743−7747.
(9) Jo, S.; Rui, H.; Lim, J. B.; Klauda, J. B.; Im, W. Cholesterol flip-flop: Insights from free energy simulation studies. *J. Phys. Chem. B* **2010**, *114*, 13342−13348. PMID: 20923227
(10) Marrink, S.-J.; Berendsen, H. J. Simulation of water transport through a lipid membrane. *J. Phys. Chem.* **1994**, *98*, 4155−4168.
(11) MacCallum, J. L.; Tieleman, D. P. Computer simulation of the distribution of hexane in a lipid bilayer: spatially resolved free energy, entropy, and enthalpy profiles. *J. Am. Chem. Soc.* **2006**, *128*, 125−130.
(12) Ashbaugh, H. S.; Truskett, T. M.; Debenedetti, P. G. A simple molecular thermodynamic theory of hydrophobic hydration. *J. Chem. Phys.* **2002**, *116*, 2907−2921.
(13) Frenkel, D. Order through disorder: Entropy-driven phase transitions. In *Complex Fluids, Proceedings of the XII Sitges Conference*, Barcelona, Spain, June 1−5, 1992; Garrido, L., Ed.; Lecture Notes in Physics, Vol. *415*; Springer, 1993; pp 137−148.
(14) Sharma, S.; Debenedetti, P. G. Free energy barriers to evaporation of water in hydrophobic confinement. *J. Phys. Chem. B* **2012**, *116*, 13282−13289.
(15) Frenkel, D. Order through entropy. *Nature materials* **2015**, *14*, 9−12.
(16) Lee, S.; Teich, E. G.; Engel, M.; Glotzer, S. C. Entropic colloidal crystallization pathways via fluid−fluid transitions and multidimensional prenucleation motifs. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 14843−14851.
(17) Vo, T.; Glotzer, S. C. A theory of entropic bonding. *Proc. Natl. Acad. Sci.* **2022**, *119*, No. e2116414119, DOI: 10.1073/pnas.2116414119.
(18) Hänggi, P.; Talkner, P.; Borkovec, M. Reaction-rate theory: fifty years after kramers. *Rev. Mod. Phys.* **1990**, *62*, 251−341.
(19) Guarnera, E.; Vanden-Eijnden, E. Optimized markov state models for metastable systems. *J. Chem. Phys.* **2016**, *145*, 024102.
(20) Gimondi, I.; Tribello, G. A.; Salvalaglio, M. Building maps in collective variable space. *J. Chem. Phys.* **2018**, *149*, 104104.
(21) Tsai, S.-T.; Smith, Z.; Tiwary, P. Reaction coordinates and rate constants for liquid droplet nucleation: Quantifying the interplay between driving force and memory. *J. Chem. Phys.* **2019**, *151*, 154106.
(22) Katiyar, A.; Thompson, W. H. Temperature dependence of peptide conformational equilibria from simulations at a single temperature. *J. Phys. Chem. A* **2021**, *125*, 2374−2384. PMID: 33720712
(23) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Struct., Funct., Bioinf.* **1995**, *21*, 167−195.
(24) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.
(25) Zhuravlev, P.; Papoian, G. Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. *Q. Rev. Biophys.* **2010**, *43*, 295−332.
(26) Leoni, F.; Russo, J. Nonclassical nucleation pathways in stacking-disordered crystals. *Phys. Rev. X* **2021**, *11*, 031006.
(27) Jiang, H.; Haji-Akbari, A.; Debenedetti, P. G.; Panagiotopoulos, A. Z. Forward flux sampling calculation of homogeneous nucleation rates from aqueous nacl solutions. *J. Chem. Phys.* **2018**, *148*, 044505.
(28) Jiang, H.; Debenedetti, P. G.; Panagiotopoulos, A. Z. Nucleation in aqueous nacl solutions shifts from 1-step to 2-step mechanism on crossing the spinodal. *J. Chem. Phys.* **2019**, *150*, 124502.
(29) Schwantes, C. R.; Pande, V. S. Improvements in markov state model construction reveal many non-native interactions in the folding of ntl9. *J. Chem. Theory Comput.* **2013**, *9*, 2000−2009. PMID: 23750122
(30) Beauchamp, K. A.; McGibbon, R.; Lin, Y.-S.; Pande, V. S. Simple few-state models reveal hidden complexity in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17807−17813.
(31) Tsai, S.-T.; Smith, Z.; Tiwary, P. Sgoop-d: Estimating kinetic distances and reaction coordinate dimensionality for rare event systems from biased/unbiased simulations. *J. Chem. Theory Comput.* **2021**, *17*, 6757−6765.
(32) Finney, A. R.; Salvalaglio, M. Multiple pathways in NaCl homogeneous crystal nucleation. *Faraday Discuss.* **2021**, DOI: 10.1039/D1FD00089F.
(33) Tiwary, P.; van de Walle, A. A review of enhanced sampling approaches for accelerated molecular dynamics. In *Multiscale Materials Modeling for Nanomechanics*; Weinberger, C. R., Tucker, G. J., Eds.; Springer Series in Materials Science, Vol. *245*; Springer International Publishing: Cham, The Netherlands, 2016; pp 195−221.
(34) Tiwary, P.; van de Walle, A. Accelerated molecular dynamics through stochastic iterations and collective variable based basin identification. *Phys. Rev. B* **2013**, *87*, 094304.
(35) Bussi, G.; Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nature Reviews Physics* **2020**, *2*, 200−212.
(36) Hyvärinen, A.; Karhunen, J.; Oja, E. *Independent Component Analysis*; Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control; Wiley, 2001.
(37) *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*; Bowman, G., Pande, V., Noé, F., Eds.; Advances in Experimental Medicine and Biology, Vol. *797*; Springer: Cham, The Netherlands, 2014.
(38) Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint. *Annu. Rev. Phys. Chem.* **2016**, *67*, 159−184. PMID: 26980304
(39) Wang, D.; Tiwary, P. State predictive information bottleneck. *J. Chem. Phys.* **2021**, *154*, 134111.
(40) Alemi, A. A.; Fischer, I.; Dillon, J. V.; Murphy, K. Deep variational information bottleneck. International Conference on Learning Representations 2017, Toulon, France, April 24−26, 2017. *arXiv (Computer Science.Machine Learning)*, 1612.00410. .
(41) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(42) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(43) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

(44) Ribeiro, J. M. L.; Bravo, P.; Wang, Y.; Tiwary, P. Reweighted autoencoded variational bayes for enhanced sampling (rave). *J. Chem. Phys.* **2018**, *149*, 072301.
(45) Pant, S.; Smith, Z.; Wang, Y.; Tajkhorshid, E.; Tiwary, P. Confronting pitfalls of ai-augmented molecular dynamics using statistical physics. *J. Chem. Phys.* **2020**, *153*, 234118.
(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*, 3231−3238.
(47) Wang, Y.; Tiwary, P. Understanding the role of predictive time delay and biased propagator in rave. *J. Chem. Phys.* **2020**, *152*, 144102.
(48) Wang, Y.; Lamim Ribeiro, J. M.; Tiwary, P. Machine learning approaches for analyzing and enhancing molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2020**, *61*, 139−145.
(49) Tishby, N.; Pereira, F. C.; Bialek, W. The information bottleneck method. *arXiv (Physics.Data Analysis, Statistics and Probability)*, April 24, 2000, 0004057, ver. 1. .
(50) Kingma, D. P.; Welling, M. Auto-Encoding Variational Bayes. International Conference on Learning Representations 2014, Banff, Canada, April 14−16, 2014. *arXiv (Statistics.Machine Learning)*, December 20, 2013, 1312.6114. .
(51) Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; Adaptive Computation and Machine Learning; MIT Press: Cambridge, MA, 2016.
(52) Tomczak, J. M.; Welling, M. VAE with a VampPrior. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (AISTATS) 2018*, April 9−11, 2018, Lanzarote, Canary Islands; Storkey, A., Perez-Cruz, F., Eds.; Proceedings of Machine Learning Research, Vol. *84*; PMLR, 2018; pp 1214−1223 (2018), arXiv:1705.07120.
(53) Deuflhard, P.; Weber, M. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra and Its Applications* **2005**, *398*, 161−184.
(54) Wehmeyer, C.; Noé, F. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *J. Chem. Phys.* **2018**, *148*, 241703.
(55) Chen, W.; Ferguson, A. L. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *J. Comput. Chem.* **2018**, *39*, 2079−2102.
(56) Gu, H.; Wang, W.; Cao, S.; Unarta, I. C.; Yao, Y.; Sheong, F. K.; Huang, X. Rpnet: a reverse-projection-based neural network for coarse-graining metastable conformational states for protein dynamics. *Phys. Chem. Chem. Phys.* **2022**, *24*, 1462−1474.
(57) Sidky, H.; Chen, W.; Ferguson, A. L. Machine learning for collective variable discovery and enhanced sampling in biomolecular simulation. *Mol. Phys.* **2020**, *118*, No. e1737742.
(58) Röblitz, S.; Weber, M. Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Advances in Data Analysis and Classification* **2013**, *7*, 147−179.
(59) Kollias, L.; Cantu, D. C.; Glezakou, V.-A.; Rousseau, R.; Salvalaglio, M. On the role of enthalpic and entropic contributions to the conformational free energy landscape of mil-101(cr) secondary building units. *Advanced Theory and Simulations* **2020**, *3*, 2000092.
(60) Silverman, B. *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability; Chapman & Hall, 1986.
(61) Hartmann, C.; Latorre, J. C.; Ciccotti, G. On two possible definitions of the free energy for collective variables. *Eur. Phys. J. Spec. Top.* **2011**, *200*, 73−89.
(62) Lelièvre, T.; Rousset, M.; Stoltz, G. *Free Energy Computations: A Mathematical Perspective*; World Scientific: Singapore, 2010.
(63) Faradjian, A. K.; Elber, R. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.* **2004**, *120*, 10880−10889.
(64) Banisch, R.; Trstanova, Z.; Bittracher, A.; Klus, S.; Koltai, P. Diffusion maps tailored to arbitrary non-degenerate Itô processes. *Applied and Computational Harmonic Analysis* **2020**, *48*, 242−265.
(65) Shinoda, W. Permeability across lipid membranes. *Biochimica et Biophysica Acta - Biomembranes* **2016**, *1858*, 2254−2265.
(66) Lee, C. T.; Comer, J.; Herndon, C.; Leung, N.; Pavlova, A.; Swift, R. V.; Tung, C.; Rowley, C. N.; Amaro, R. E.; Chipot, C.; et al. Simulation-based approaches for determining membrane permeability of small compounds. *J. Chem. Inf. Model.* **2016**, *56*, 721−733.
(67) Fathizadeh, A.; Elber, R. Ion Permeation through a Phospholipid Membrane: Transition State, Path Splitting, and Calculation of Permeability. *J. Chem. Theory Comput.* **2019**, *15*, 720−730.
(68) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(69) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **2013**, *139*, 015102.
(70) Husic, B. E.; McGibbon, R. T.; Sultan, M. M.; Pande, V. S. Optimized parameter selection reveals trends in markov state models for protein folding. *J. Chem. Phys.* **2016**, *145*, 194103.
(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*, 065101.
(72) Beyerle, E. R.; Guenza, M. G. Identifying the leading dynamics of ubiquitin: a comparison between the tica and the le4pd slow fluctuations in amino acids' position. *J. Chem. Phys.* **2021**, *155*, 244108.
(73) E, W.; Vanden-Eijnden, E. Towards a theory of transition paths. *J. Stat. Phys.* **2006**, *123*, 503−523.
(74) Noe, F.; Schutte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 19011−19016.
(75) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *J. Chem. Phys.* **2007**, *126*, 155102.
(76) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, 174105.
(77) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J. H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525−5542.
(78) Berezhkovskii, A.; Szabo, A. Ensemble of transition states for two-state protein folding from the eigenvectors of rate matrices. *J. Chem. Phys.* **2004**, *121*, 9186−9187.
(79) Buchete, N. V.; Hummer, G. Coarse master equations for peptide folding dynamics. *J. Phys. Chem. B* **2008**, *112*, 6057−6069.
(80) Cardenas, A. E.; Elber, R. Computational study of peptide permeation through membrane: searching for hidden slow variables. *Mol. Phys.* **2013**, *111*, 3565−3578. PMID: 26203198
(81) Bussi, G.; Tribello, G. A. Analyzing and biasing simulations with PLUMED. In *Biomolecular Simulations*; Bonomi, M., Camilloni, C., Eds.; Springer, 2019; pp 529−578.
(82) Sittel, F.; Stock, G. Perspective: Identification of collective variables and metastable states of protein dynamics. *J. Chem. Phys.* **2018**, *149*, 150901.
(83) Allen, R. J.; Valeriani, C.; Ten Wolde, P. R. Forward flux sampling for rare event simulations. *J. Phys.: Condens. Matter* **2009**, *21*, 463102.
(84) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; et al. Xsede: Accelerating scientific discovery. *Computing in Science Engineering* **2014**, *16*, 62−74.

https://doi.org/10.1021/acs.jpcb.2c01782
J. Phys. Chem. B 2022, 126, 3950−3960