

# Molecular transport through membranes: Accurate permeability coefficients from multidimensional potentials of mean force and local diffusion constants

Rui Sun, Yining Han, Jessica M. J. Swanson, Jeffrey S. Tan, John P. Rose, and Gregory A. Voth

Citation: *The Journal of Chemical Physics* **149**, 072310 (2018); doi: 10.1063/1.5027004

View online: <https://doi.org/10.1063/1.5027004>

View Table of Contents: <http://aip.scitation.org/toc/jcp/149/7>

Published by the [American Institute of Physics](#)

---

---

**PHYSICS TODAY**

WHITEPAPERS

## ADVANCED LIGHT CURE ADHESIVES

Take a closer look at what these environmentally friendly adhesive systems can do

READ NOW

PRESENTED BY  
 **MASTERBOND**  
ADHESIVES | SEALANTS | COATINGS

# Molecular transport through membranes: Accurate permeability coefficients from multidimensional potentials of mean force and local diffusion constants

Rui Sun,<sup>1,2</sup> Yining Han,<sup>1</sup> Jessica M. J. Swanson,<sup>1</sup> Jeffrey S. Tan,<sup>3</sup> John P. Rose,<sup>3</sup> and Gregory A. Voth<sup>1,a)</sup>

<sup>1</sup>Department of Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637, USA

<sup>2</sup>Department of Chemistry, The University of Hawai'i, Manoa, Honolulu, Hawaii 96822, USA

<sup>3</sup>Small Molecule Design and Development, Lilly Corporate Center, Eli Lilly and Company, Indianapolis, Indiana 46285, USA

(Received 26 February 2018; accepted 30 April 2018; published online 21 May 2018)

Estimating the permeability coefficient of small molecules through lipid bilayer membranes plays an important role in the development of effective drug candidates. *In silico* simulations can produce acceptable relative permeability coefficients for a series of small molecules; however, the absolute permeability coefficients from simulations are usually off by orders of magnitude. In addition to differences between the lipid bilayers used *in vitro* and *in silico*, the poor convergence of permeation free energy profiles and over-simplified diffusion models have contributed to these discrepancies. In this paper, we present a multidimensional inhomogeneous solubility-diffusion model to study the permeability of a small molecule drug (trimethoprim) passing through a POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) lipid bilayer. Our approach improves the permeation model in three ways: First, the free energy profile (potential of mean force, PMF) is two-dimensional in two key coordinates rather than simply one-dimensional along the direction normal to the bilayer. Second, the 2-D PMF calculation has improved convergence due to application of the recently developed transition-tempered metadynamics with randomly initialized replicas, while third, the local diffusivity coefficient was calculated along the direction of the minimum free energy path on the two-dimensional PMF. The permeability is then calculated as a line integral along the minimum free energy path of the PMF. With this approach, we report a considerably more accurate permeability coefficient (only 2–5 times larger than the experimental value). We also compare our approach with the common practice of computing permeability coefficients based only on the translation of the center of mass of the drug molecule. Our paper concludes with a discussion of approaches for minimizing the computational cost for the purpose of more rapidly screening a large number of drug candidate molecules. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5027004>

## I. INTRODUCTION

Small molecules (<900 Da, generated by chemical synthesis), biomolecules (>10 000 Da, generated by bioengineering), and peptides (~300-10 000 Da, generated by chemical synthesis and bioengineering) are usually selected as compounds for pharmaceutical drugs.<sup>1</sup> Small molecules are most commonly used for various reasons, including high chemical stability<sup>2</sup> and well-studied physicochemical properties.<sup>3</sup> These advantages enable them to be taken orally, whereas peptides and biomolecules generally require parenteral administration (e.g., by injection) due to their degradation in the human digestion system. However, oral administration requires small molecules with high rates of absorption, especially for those absorbed through the intestines. As a result, pharmaceutical companies typically screen the absorption potential early in the discovery and development phase, when the investment

in a compound is low, to remove poor performers and to identify candidates that need to be modified.<sup>4</sup> Lipid bilayer membranes, in various forms, are the main physical barrier for the absorption of drug molecules, whether they are traversed in the transcellular pathway (moving through cells) or in the less common paracellular pathway (passing in between cells).<sup>5</sup> Passive permeation,<sup>6</sup> facilitated permeation,<sup>7</sup> and vesicular transport<sup>8</sup> are three main mechanisms for the transcellular process, among which passive permeation is dominant, especially for small molecules.<sup>9</sup>

Due to its importance, great effort has been devoted to understanding passive permeation of small molecules through bilayers for over 100 years.<sup>9,10</sup> In the first lipid bilayer permeation model proposed by Overton,<sup>11</sup> the permeability coefficient of a solute is solely correlated to its oil/water partition coefficient such that the lipid bilayer is represented by a homogeneous oil slab. Later, planar lipid bilayers were employed as one of the main experimental models to investigate the permeability coefficient of small molecules, in which two chambers with different concentrations of permeant are separated by

<sup>a)</sup>Author to whom correspondence should be addressed: [gavoth@uchicago.edu](mailto:gavoth@uchicago.edu)

planar lipid bilayers formed on small holes ( $\sim 1$  mm<sup>2</sup>) in a polyethylene or Teflon platform.<sup>12,13</sup> In recent years, scientists have developed various lipid assays *in vitro* that are closely related to the observed physiological drug absorption data. Caco-2 is a cell-based assay that mimics the passage of drugs through the intestinal mucosa, where a monolayer of human colorectal carcinoma cells (characterized by morphological and functional similarities to the small intestinal epithelium cells) is grown on a filter separating two stacked microwell plates.<sup>14,15</sup> Madin–Darby Canine Kidney (MDCK) is another cell-based assay that is widely used with a similar experimental apparatus.<sup>16</sup> Besides the cell-based assay, Parallel Artificial Membrane Permeation Assay (PAMPA) has been developed to utilize a filter coated with a solution of lipid material in inert organic solvent acting as the artificial membrane.<sup>17,18</sup> Since their introduction, the aforementioned *in vitro* methods have been greatly adapted by both academia and pharmaceutical industries in drug development, as tools of measuring permeability coefficients. However, despite their success, there are some shortcomings.

For example, the preparation of stable monolayers for the cell-based assay experiments (Caco-2 and MDCK) requires up to three weeks in a contamination-free environment, and the assay needs to be examined for tight-junction formation prior to use (to eliminate any potential paracellular pathways).<sup>19</sup> The cell-based assay also contains endogenous transporter and efflux systems, which usually work against the passive permeation process, therefore complicating the data interpretation and reducing the reproducibility of the experiments.<sup>20</sup> The PAMPA methods focus on simplicity and flexibility, eliminating the slow growth period and possible contributions from transporters, while enabling the use of different lipids to construct the bilayer.<sup>21</sup> Although the results from the two methods are generally similar, the PAMPA sometimes shows worse linear correlation with oral absorption than the cell-based assays do, which could be arguably allotted to the unknown phase formed by lipids in the PAMPA membranes.<sup>22–24</sup> Regardless of the experimental method chosen, the relative permeability coefficients of a series of small molecules are typically better reproduced than the absolute permeability coefficients of individual small molecules due to difficulties in measurement calibration and experimental perturbations.<sup>9,25</sup>

In addition to the value of the permeability coefficient itself, it is important to understand the passive permeation process at an atomistic resolution, which is generally not possible with *in vitro* experiments alone, in order to provide a fundamental insight that can facilitate drug development both specifically and broadly. In this light, molecular dynamics (MD) simulations, which have revealed atomistic information for numerous physical and chemical processes in biological systems, offer a promising tool. Developed from the homogeneous model,<sup>26</sup> the inhomogeneous solubility-diffusion model is a more accurate *in silico* model that takes account the inhomogeneity of the interface between the bulk water and the lipid bilayer.<sup>27,28</sup> In this model, the resistivity coefficient ( $R$ ), which is the inverse of the permeability coefficient ( $P_m$ ), is computed as

$$R = \frac{1}{P_m} = \int_{r_2}^{r_1} \frac{\exp[\beta W(r)]}{D(r)} dr. \quad (1)$$

In this equation,  $r$  is a collective variable (CV) describing the permeation process, e.g., often the  $z$  component distance ( $z$  is perpendicular to the membrane) between the center of mass of the permeating drug and the center of mass of the lipid bilayer.  $W(r)$  is the free energy profile (potential of mean force or PMF) of the permeation with respect to  $r$  and  $D(r)$  is the local diffusivity of the permeating drug along the direction of  $r$ . Lee *et al.*<sup>29</sup> have recently computed the permeability coefficient of urea, benzoic acid, and codeine using MD simulations with different enhanced sampling techniques and concluded that their permeability coefficients agree within 1.5 log units ( $\sim 30$ -fold difference) of the experimental values. In a similar study in 2017, Bennion *et al.*<sup>30</sup> employed umbrella sampling<sup>31</sup> MD simulations to comprehensively assess the passive permeability profile of a range of compounds through a lipid bilayer. Even though their *in silico* results differ from the *in vitro* results by about five log units, the relative permeability coefficients of a series of custom synthesized and closely related compounds are qualitatively predicted. Therefore, it remains very challenging for computational simulations to estimate a permeability coefficient that is within an order of magnitude of the experimental result.

There are several sources that contribute to the disagreement between *in vitro* and *in silico* permeability coefficients. The first one is the differences between the membrane used in the experiments and the simplified lipid bilayers used in the simulations. This difference is more significant in the cell-based assay experiment, where the cell membrane is made of various kinds of lipids, cholesterol, and transporters; however, even for the PAMPA experiment, a direct comparison between simulation and experiment is difficult to make (i.e., egg yolk was widely used to make the artificial membrane<sup>17</sup>) unless the artificial membrane is specifically made of a couple kinds of synthetic phospholipids.<sup>21</sup> Simulating a physiologically relevant and multicomponent lipid bilayer that is consistent with the experiment remains mostly infeasible given the limited computation resources.<sup>32</sup>

Besides the difference in the lipid bilayer, the accuracy of the *in silico* permeability coefficient depends largely upon the accuracy of the PMF,  $W(r)$  [because of the exponential factor in Eq. (1)], and the local diffusivity coefficient,  $D(r)$ . Permeation is usually a rare event corresponding to high free energy barrier; thus, enhanced sampling methods are usually employed to assist the calculation of the PMF. Lee *et al.*<sup>29</sup> have computed the permeation PMF using umbrella sampling,<sup>31</sup> replica exchange umbrella sampling,<sup>33</sup> adaptive biasing force,<sup>34</sup> and multiple-walker adaptive biasing force<sup>35</sup> and reported no definitive advantage for any of these methods in their ability to predict the drug permeability coefficient. However, none of the above enhanced sampling methods utilized metadynamics (MetaD)<sup>36</sup> that is able to converge the estimate of PMF with proper tempering.<sup>37,38</sup> The recent development of transition tempered MetaD<sup>39</sup> (TTMetaD) has been proven to provide a reliable estimate of the PMF at the early stage of the simulation, especially for systems in which a rough knowledge of the PMF landscape is available prior to the simulation (i.e., an estimate of the positions of the minima and barrier between them on the PMF).<sup>40</sup> This behavior of the free energy sampling

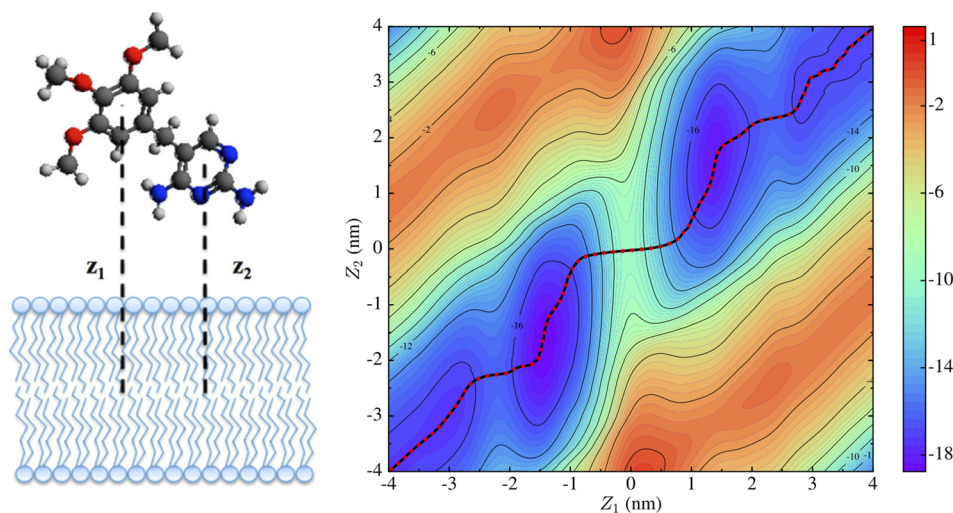


FIG. 1. Left panel: Definition of the collective variables (CVs) used in the TTMetaD simulation. Right panel: 2D PMF (in kcal/mol) of trimethoprim permeation through a POPC lipid bilayer after 2.5  $\mu$ s of simulation (0.5  $\mu$ s on 5 replicas).

is essential for drug candidates considered by pharmaceutical companies since the number of drugs needing to be screened is usually large. The good performance of an enhanced sampling method with limited sampling is thus highly desirable to reduce the overall computational cost and hence time to solution. This has motivated us to utilize TTMetaD to achieve a more accurate permeation PMF and to do so more efficiently.<sup>40</sup> Moreover, TTMetaD has proven to be so accurate and rapidly convergent that it becomes possible to calculate a *two-dimensional PMF* for the drug permeation process, i.e., a PMF involving the coordinate of the permeating molecule normal to the membrane (usually the  $z$ -coordinate) plus a second key coordinate such as an orientation of the molecule relative to the lipid bilayer plane. These 2-D PMFs reveal a great deal more about the molecular conformations and orientations that are the most pertinent to the permeation process, and they can in principle provide valuable information to the drug designer to enhance the permeation.

A third challenge is the diffusion model that utilizes  $D(r)$  in Eq. (1), which is generally chosen as simply the diffusion along the  $z$  component (perpendicular to the lipid bilayer) of the distance between the center of mass of the permeating drug and the center of mass of the lipid bilayer.<sup>29,30</sup> Though the center of mass translation of a small molecule through a lipid bilayer is the most important component of a permeation process, there are other important degrees of freedom such as orientation of the drug with respect to the membrane, especially for drug candidate molecules that have both hydrophobic region(s) and hydrophilic region(s). Besides being an over simplified model to represent local diffusion, if this orientational degree of freedom is slow enough, using only the center of mass translation CV also slows down the convergence of the enhanced sampling calculation, sometimes leading to hysteresis in the PMF estimates.

As an example, the permeation of trimethoprim through a POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) lipid bilayer has clearly shown its preference in the orientation of the permeant: the difference in lipophilicity of the trimethoxybenzyl and pyrimidine group leads to tilted metastable configurations of trimethoprim interacting with the membrane (see Fig. 1).<sup>40</sup> As a result, the minimum free energy

path (MFEP), which represents the average pathway over a large ensemble of the permeation processes, is zigzag instead of a straight diagonal line. An ideal representation of  $r$  should contain the information of the position of trimethoprim (which can be realized in the traditional usage of the center of mass distance) as well as the orientation of trimethoprim, and the local diffusivity coefficient,  $D(r)$ , should represent the local diffusion along the direction of the MFEP. To our knowledge, there has not been such a permeability coefficient calculation that utilizes the MFEP from a multidimensional PMF, taking into account slow degrees of freedom other than the simple center of mass translation.

The purpose of this paper is therefore to demonstrate, through the example of the permeation of trimethoprim through POPC bilayer, that a combination of TTMetaD to improve the quality of  $W(r)$  and to increase its dimensionality, along with a computation of  $D(r)$  along the MFEP and a calculation of the permeability along that path from Eq. (1), can provide a systematic estimate of the permeability coefficient that is within one order of magnitude of the experimental result. As a comparison, we have also followed the conventional protocol of permeability coefficient calculation with only one CV (the center of mass translation). This paper concludes with discussions of the computational cost of our method and presents a test case that uses significantly less simulations but provides a reasonably accurate result. We anticipate this method can become a new standard of *in silico* permeability coefficient calculation. Future work will also focus on utilizing this approach to screen the relative permeability coefficient of a series of small molecule drugs.

## II. METHODS

### A. Transition-tempered metadynamics

In MetaD simulations, the process of interest (permeation in this case) is sampled by introducing a bias force on a small number of preselected CVs,  $s$ , that are expected to transition the system across the free energy barriers and to delineate the reactant and product minima. A history dependent bias potential ( $V_G$ ), which usually has the form of a summation

of Gaussian functions centered on previously visited points in CV space, is systematically added to the Hamiltonian.<sup>36</sup> As a result, the system is discouraged from revisiting configurations that have already been sampled and is encouraged to explore the high-energy regions of CV space. After a long simulation, the bias potential offsets the resistance from the system and makes the sampling in CV space ergodic and diffusive; therefore, the negative of the MetaD bias potential can be used to estimate the underlying PMF.<sup>38,41</sup>

In its original implementation, the MetaD bias potential ( $V_G$ ) took the form<sup>36</sup>

$$V_G(s(R), t) = \int_0^t dt' w \exp \left[ - \sum_{i=1}^d \frac{(s_i(R) - s_i(R(t')))^2}{2\sigma_i^2} \right], \quad (2)$$

where  $w$  is the bias energy addition rate, computed as the height of individual Gaussian functions added,  $w_0$ , divided by a deposition stride  $\tau$ . The index  $i$  represents the  $i$ th CV, which is a function of the system coordinates,  $R$ , and  $s_i(R(t'))$  is the value of the CV calculated at time  $t'$ .  $\sigma$  is the width of the Gaussian function. Since the bias energy addition rate ( $w$ ) is constant along the simulation, this approach is usually referred to as non-tempered MetaD. It is obvious that non-tempered MetaD yields a bias potential that does not converge asymptotically,<sup>38</sup> in contrast to the fact that the PMF is a time independent property. A suggested solution to this problem is to calculate the time average of the bias potential after the motion of the system on the CV space becomes diffusive.<sup>41</sup> Practically, however, we have found this approach can show false convergence or instability in the late stages of simulations, likely due to too much bias energy being added.<sup>40,42</sup>

Well-Tempered Metadynamics (WTMetaD)<sup>37</sup> was developed to address this convergence issue by tempering the height of the Gaussian functions exponentially with respect to the local bias potential,  $V_G(s, t)$ . A time-dependent bias energy addition rate,  $w(t)$ , replaces the time-independent energy addition rate,  $w$ , in Eq. (2),

$$w(t) = w_0 \exp \left[ - \frac{V_G(s(R), t)}{k_B \Delta T} \right], \quad (3)$$

where  $\Delta T$  is a parameter that tunes the tempering of the Gaussian height and needs to be chosen prior to the simulation. In spite of the success of WTMetaD in many simulations, it bears an obvious trade-off between exploration of CV space in the early stages of a simulation and the asymptotic convergence rate.<sup>39,40,42</sup> On one hand, a large value of  $\Delta T$  builds up a fast-growing but noisy bias potential that hinders the convergence; on the other hand, a small value of  $\Delta T$  generates a smooth but slow-growing bias potential that might decrease the efficiency of escaping a minimum. The original WTMetaD paper suggests that  $\Delta T$  should be chosen according to the free energy barrier to make  $(\Delta T + T)$  the same order of magnitude as the barrier height.<sup>37</sup> However, barrier heights are typically the quantity that is sought, not known *a priori*, which complicates the implementation of WTMetaD and leads to many trial simulations that either remain stuck in minima or place too much bias force into the system, forcing it into clearly unphysical regions of phase space.

Inspired by WTMetaD, transition-tempered MetaD (TTMetaD)<sup>39</sup> was designed to converge rapidly and asymptotically without sacrificing the exploration in the early stages of simulations.<sup>40,42</sup> TTMetaD tempers the Gaussian height with a similar exponential rule as WTMetaD, but the amount of tempering is determined according to the overall progress of the MetaD sampling. The time-independent energy addition rate  $w$  in Eq. (2) is replaced by

$$w(t) = w_0 \exp \left[ - \frac{V^*(s(\lambda), t)}{k_B \Delta T} \right], \quad (4)$$

in which  $V^*$  is the minimal bias on the maximally biased path among all the continuous paths  $s(\lambda)$  that connect all of the pre-selected basin points in the CV space. One needs to roughly define positions of the basins on the PMF prior to the simulation, and TTMetaD tracks the region that the bias energy has expanded to. The tempering does not kick in until all the basins are fully covered, and it tempers according to the least biased point. Therefore, TTMetaD is able to explore the CV space with full Gaussians and use more aggressively tempered Gaussians to converge the simulation once the basins have been filled.

Metadynamics with a tempering rule (e.g., TTMetaD) has been proven to converge the bias energy asymptotically to a linearly scaled inverse of the underlying PMF.<sup>38</sup> Therefore, the PMFs of the permeation of trimethoprim were estimated as the inverse of the bias energy throughout the simulation.

## B. Diffusion along the minimum free energy path (MFEP)

According to Eq. (1), the permeability coefficient computed from the solubility-diffusion model depends on the local diffusivity,  $D(r)$ , where  $r$  is the permeation collective variable. Herein, we employed restrained MD simulations (umbrella sampling) to calculate the local diffusivity coefficient of trimethoprim at different positions along the MFEP. A harmonic potential restrained the trimethoprim to oscillate at specified points along the permeation pathway while the surrounding water and lipid molecules effectively served as the frictional bath for the solute. This method assumes that the harmonic potential overwhelms the underlying PMF that governs the interaction between the small molecule drug and its surrounding. Lee *et al.*<sup>29</sup> discussed the magnitude of the harmonic potential and concluded that any value above the conventional umbrella sampling spring [10.0 kcal/(mol/Å<sup>2</sup>)] constant are appropriate.

The umbrella sampling simulations provided a time series of motion along  $r$ , which was used to calculate the local diffusivity. One way to do this is to use a numerical Laplace transform of the velocity autocorrelation functions for several values of the transform parameter  $\tau$  and extrapolate it to the limit of  $\tau = 0$ .<sup>43</sup> Improving upon this, Hummer<sup>44</sup> derived a method wherein the local diffusion coefficients could be estimated directly from the autocorrelation function of  $r$  in a harmonically restrained simulation,

$$D(r) = \frac{[var(r)]^2}{\int_0^\infty C_r(t) dt}. \quad (5)$$

In the above equation,  $\text{var}(r)$  is the variance of  $r$  and  $C_r(t)$  is the autocorrelation function of  $r$  at time  $t$ , computed as

$$C_r(t) = \frac{\Delta t}{t_s - t} \sum_{i=0}^{(t_s-t)/\Delta t} (r(i) - \bar{r})(r(i+t) - \bar{r}), \quad (6)$$

in which  $r(i)$  is the time series of  $r$  from the restrained simulations of length  $t_s$ , with  $\Delta t$  as the MD integration time step, and  $\bar{r}$  is the average value of  $r$ .

As discussed in the Introduction, the  $z$  component of the distance between the center of mass of the permeating drug and the center of mass of the lipid bilayer has typically been used as  $r$  to study permeation processes.<sup>29,30</sup> As a result, the local diffusivity coefficient represents the diffusion of the center of mass of the small molecule drug perpendicular to the lipid bilayer regardless of its orientation or any other relevant slow degree of freedom. However, the multi-dimensional PMF of trimethoprim permeation clearly shows that the orientation of the molecule is significant (Fig. 1) and, thus, that the diffusion should be evaluated along the MFEP. Therefore, we employed two-dimensional umbrella sampling simulations to restrain both CVs of the trimethoprim molecule along the MFEP, and the time series of the sampled points from the simulations were projected onto the tangent of the MFEP. A restraining force constant of 47.8 kcal/(mol·Å<sup>2</sup>) [20 000 kJ/(mol·nm<sup>2</sup>)] was chosen to make the sampled region of each umbrella window so small that the tangent of the MFEP is a good approximation of the MFEP in the sampling vicinity. Those projected points correspond to a time series of the reaction coordinate,  $r$ , with which we are able to calculate the local diffusivity coefficient using Eq. (5). This process is illustrated in Fig. 2. There are 41 umbrella windows covering the reaction coordinate from 0 to 0.5, corresponding to the black curve from (−4 nm, −4 nm) to (0 nm, 0 nm) on the 2D PMF (Fig. 3). Each window was equilibrated for 50 ns and then underwent another 50 ns simulation as the production run.

The convergence of a MetaD free energy profile requires filling the underlying minima in the free energy landscape. Thus, it is generally believed that it takes a longer simulation time to reach convergence when sampling more CVs, corresponding to more CV space. However, the underlying physics of the systems are important, and this intuitive conjecture is

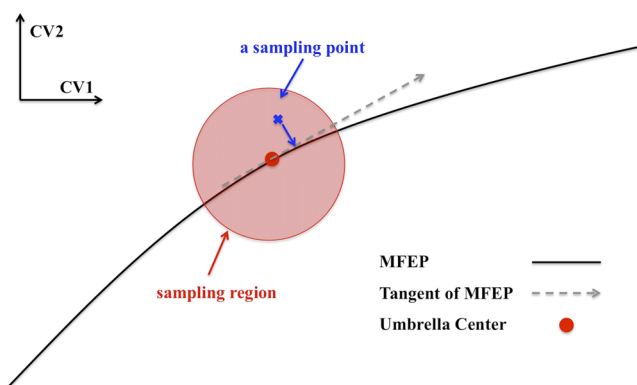


FIG. 2. A diagram showing the method of generating a time series of diffusion of  $r$  (reaction coordinate) along the MFEP.

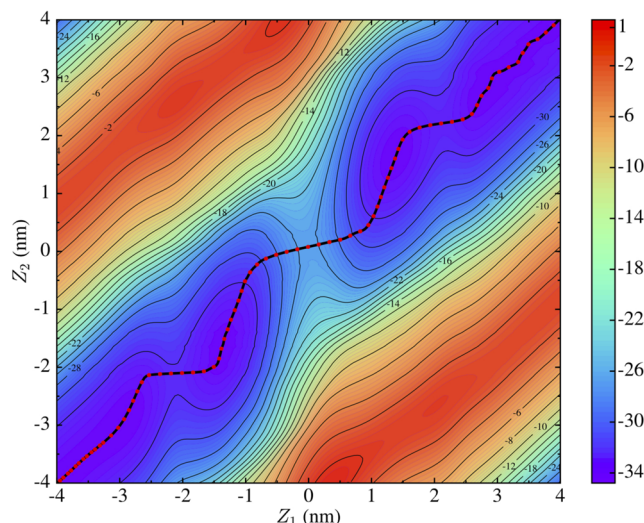


FIG. 3. 2D potential of mean force for trimethoprim permeation through the POPC lipid bilayer using longer sampling times than in our previous work (Ref. 40) and without symmetrization. The black curve traces the MFEP. The energy units are kcal/mol.

not necessarily true. In cases where this is a coupled, orthogonal slow CV, which corresponds to a non-negligible barrier in the free energy surface, overlooking that CV introduces hysteresis in the bias energy and, in some cases, even prevents the simulation from converging. Specifically, in the case of trimethoprim, there is an orientational preference when the drug interacts with the lipid head groups corresponding to a free energy difference that is larger than the thermal fluctuations. Therefore, it is important to include explicitly sampling of the orientation CV in the MetaD simulations; in fact, the MetaD simulations with both CVs (center of mass distance and orientation) converge substantially faster than that with only one CV (center of mass distance).

### C. Calculation of a 1D PMF from a 2D PMF

$W(r)$  in Eq. (1) corresponds to the free energy of a system with respect to  $r$ . The variable  $r$  is often the  $z$  component distance ( $z$  is perpendicular to the membrane) between the center of mass of the permeating drug and the center of mass of the lipid bilayer. All the TTMetaD calculations performed herein involved two collective variables; therefore, the permeation PMF was two-dimensional. In order to use Eq. (1), we must first calculate  $W(r)$  from the 2D PMF. This problem can be generalized as the following: Consider a 2D free energy sampling using collective variables  $x$  and  $y$  that yields a PMF given by  $F(x,y)$ . The MFEP on this 2D PMF can be described by a normalized coordinate,  $\xi(x,y)$ , that is a function of  $x$  and  $y$ . For any point on the MFEP,  $\xi_0$ , the tangent vector along the MFEP,  $\vec{\xi}_0$ , and the normal vector,  $\vec{\eta}_0$ , can form a local 2D coordinate system, which can be understood as a result of a transformation from the previous coordinates  $x$  and  $y$ . Applying this transformation to all the points on the MFEP,  $\xi(x,y)$ , leads to a collection of the local  $\vec{\eta}$  axis orthogonal to  $\xi$ . As a result, the points,  $F(x,y)$ , on the original PMF of coordinates  $x$  and  $y$  can be transformed to on a new set of coordinates  $F(\xi,\eta)$ .

As a result of the above argument, the free energy profile of the MFEP,  $F(\xi_0)$ , can be formulated in the new coordinate system as

$$F(\xi_0) = -\frac{1}{\beta} \ln \int \exp(-\beta F(x, y)) \cdot \delta(\xi(x, y) - \xi_0) dx dy. \quad (7)$$

Changing the integration variables from  $dx dy$  to  $d\xi d\eta$  gives

$$F(\xi_0) = -\frac{1}{\beta} \ln \int \exp(-\beta F(x, y)) \cdot \delta(\xi(x, y) - \xi_0) \times \left| \frac{\partial(x, y)}{\partial(\xi, \eta)} \right| d\xi d\eta, \quad (8)$$

where  $|\partial(x, y)/\partial(\xi, \eta)|$  is the Jacobian. As discussed above, the free energy,  $F(x, y)$ , can be expressed in the coordinates  $\xi$  and  $\eta$  as

$$F(\xi_0) = -\frac{1}{\beta} \ln \int \exp(-\beta F(\xi(x, y), \eta(x, y))) \cdot \delta(\xi(x, y) - \xi_0) \times \left| \frac{\partial(x, y)}{\partial(\xi, \eta)} \right| d\xi d\eta. \quad (9)$$

Integrating over  $\xi$  then yields

$$F(\xi_0) = -\frac{1}{\beta} \ln \int \exp(-\beta F(\xi_0, \eta)) \cdot \left| \frac{\partial(x, y)}{\partial(\xi, \eta)} \right|_{\xi_0} d\eta. \quad (10)$$

Note that by definition, the local coordinate system  $(\xi, \eta)$  is just a rotation of the original coordinate system  $(x, y)$ . The Jacobian for any rotational transformation equals unity, i.e.,

$$\left| \frac{\partial(x, y)}{\partial(\xi, \eta)} \right|_{\xi_0} \equiv 1. \quad (11)$$

The equation that we then adopt in this work to construct the free energy profile along the MFEP is given by

$$F(\xi_0) = -\frac{1}{\beta} \ln \int \exp(-\beta F(\xi_0, \eta)) d\eta. \quad (12)$$

We note that  $(x, y)$  and  $(\xi, \eta)$  in the above derivation can be dummy variables and can be any combination of collective variables in a PMF. As we will show later, this expression will also be used to compute the permeation free energy with respect to the center of mass distance ( $r$ ) between the drug molecule and the lipid bilayers from two-dimensional ( $r$  and an angle CV) TTMetaD calculations.

#### D. Calculation of $W(r')$ on the minimum free energy path

The MFEP of a multidimensional free energy landscape is, within the limits of the sampled degrees of freedom and those faster than them, the physically relevant pathway for, e.g., lipid bilayer permeation. For example, and as discussed in the Introduction, a line integral of Eq. (1) along the MFEP from a multidimensional PMF can better account for the hydrophilic-hydrophobic interactions during drug permeation. We denote the reaction coordinate along the MFEP to be  $r'$  to distinguish it from the commonly used center of mass distance between the drug and lipid bilayer,  $r$ .

Therefore, the free energy value along the MFEP is taken as  $W(r')$  in Eq. (1) and is computed by integrating the 2D PMF over the direction orthogonal to the MFEP. The detailed procedure of computing  $W(r')$  is explained as follows: Suppose

$r'_i(z_1, z_2)$  is a random point on the MFEP (e.g., Fig. 3). A line crossing  $r'_i$  that is orthogonal to the tangent of the MFEP at  $r'_i$  can be sketched, which we define as  $l_i$ . The coordinates of the points on this line  $l_i$  are denoted as  $(z_1^i, z_2^i)$ . Since the free energy of the 2D landscape has already been computed (see Fig. 3),  $F(z_1, z_2)$  is known for all the values of  $z_1$  and  $z_2$  and, consequently, the free energy of every point on the line  $l_i$ , i.e.,  $F(z_1^i, z_2^i)$  is known. For every  $l_i$  that is orthogonal to the tangent of MFEP at  $r'_i$ , the points  $(z_1^i, z_2^i)$  sitting on this line can be projected to  $r'_i$ ; and based on their relative position on the line  $l_i$ , we map those points to another degree of freedom,  $l$ , that is orthogonal to the MFEP. Through this approach, we obtain a free energy  $F(r', l)$  from  $F(z_1, z_2)$ ; then, Eq. (12) was applied to integrate over  $l$  to get the one-dimensional permeation free energy profile with respect to  $r'$ ,  $W(r')$  for Eq. (1).

From a more intuitive perspective, we summarize the procedure for calculating the free energy profile of the permeation as follows: After the 2D PMF with respect to the two predefined collective variables is sampled (Sec. II A), we then extract the free energy profile along the minimum free energy path (MFEP) along the 2D PMF. This was done in two separate steps, as described above and again here. First, at each point on the MFEP, we construct a local orthogonal coordinate system with one axis along the tangent vector at this point and the other axis aligning with the normal vector. Based on this new coordinate frame, the previous 2D PMF is transformed to a 2D PMF with respect to the two new collective variables in which the first represents the MFEP coordinate and the second denotes the orthogonal degrees of freedom. Then, the free energy profile along the MFEP is obtained by integrating over the orthogonal collective variable with Eq. (12) in the above discussion.

#### E. Molecular dynamics simulations

There were  $16 \times 2$  POPC molecules in the lipid bilayers spanning a  $3.2 \text{ nm} \times 3.2 \text{ nm}$  region (approximately), a size that has been proven to be appropriate for the permeation of a small molecule.<sup>45</sup> The bilayer was solvated with a hydration ratio of 65 and with periodic boundary conditions enforced along all the directions of Cartesian space. CHARMM36,<sup>46</sup> CHARMM general force fields (CGenFF),<sup>47</sup> and TIP3P<sup>48</sup> were used to model the POPC lipid, trimethoprim, and water, respectively. The lipid and the water plus small molecules were individually coupled to two 323 K heat baths using velocity rescaling with a stochastic term.<sup>49</sup> This temperature is higher than the physiological temperature (310 K) and was chosen to be consistent with our previous study on the permeation of trimethoprim.<sup>40</sup> Pressure control was also applied by a Berendsen barostat<sup>50</sup> in which the box was rescaled every 5 ps. The cutoff distance for the short-range neighbor list was 12 Å and the neighbor list was updated every 40 steps. Fast smooth Particle-Mesh Ewald (SPME)<sup>51</sup> was used to compute the long-range electrostatic interactions. All hydrogen bonds were constrained by linear constraint solver (LINCS)<sup>52</sup> and the MD integration time step was 2 fs. The MD simulations were consistent with the simulations of trimethoprim in our previous paper<sup>40</sup> and were carried out using GROMACS-5.1.4<sup>53</sup> patched with

PLUMED2<sup>54</sup> modified by the Voth research group to perform TTMetaD.<sup>39,55</sup> The total simulation time analyzed in this paper is about 30  $\mu\text{s}$ , of which 7.5  $\mu\text{s}$  was from our previous work.<sup>40</sup>

### III. RESULTS

#### A. Potential of the mean force

The central advantage of TTMetaD is a reliable estimate of the PMF early in the simulations without prior knowledge of the free energy barrier height. Therefore, our previous simulations of trimethoprim permeation<sup>40</sup> were intentionally stopped after 500 ns when they showed aggressive tempering and ample sampling of CV space. To test the robustness of TTMetaD to the choice of CV(s), the previous simulations were carried out with three different sets of CVs (2 CVs in each set). Although, as demonstrated,<sup>40</sup> these CV choices produced similar final 1D PMFs, the third set was not effective at capturing the orientational dependence of permeation, which was best captured by the second CV set. Thus, the second CV set, composed of (1) the  $z$  component distance between the center of mass of the trimethoxybenzyl group and the membrane,  $z_1$  and (2) the  $z$  component distance between the center of mass of the pyrimidine group and the membrane,  $z_2$  (see the left panel of Fig. 1), was selected for the 2D permeability coefficient calculation presented herein.

Given that tempered MetaD simulations are only guaranteed to reach asymptotic convergence with infinite sampling,<sup>38</sup> achieving diffusive motion in CV space and agreement between replicas are commonly accepted criteria convergence in practice.<sup>40,42,56,57</sup> In this work, we extended our previous simulations to refine the estimate of PMF of the permeation of trimethoprim through a POPC lipid bilayer. This is an effort to show that with a lengthy TTMetaD calculation, one is able to get an almost exact PMF that benefits the calculation of the permeability coefficient. We simulated 16 replicas with randomized velocities to avoid false convergence and decreased the required wall and central processing unit (CPU) times.<sup>40,42</sup> After 500 ns of additional sampling for each replica, the final 2D PMF was estimated (see Fig. 3), which was then symmetrized with respect to the center of the membrane (figure not shown) to generate as much sampling as possible to calculate the 1D PMF. The string method at zero temperature<sup>58</sup> was employed to search for the MFEP on the symmetrized 2D PMF, and the 1D free energy profile of the permeation (red curve in Fig. 4) is computed from Eq. (12) to integrate over the direction orthogonal to the MFEP. To evaluate the change in the free energy profile due to the extended sampling conducted herein, we compared the new MFEP with the MFEP from the previous simulations,<sup>40</sup> the difference is  $\sim 0.76$  kcal/mol. The total TTMetaD simulation time for calculating this 2D permeation free energy profile was 10.5  $\mu\text{s}$ .

The advantage of the aforementioned method of computing a permeability coefficient can be seen from its comparison to the commonly used 1D permeability coefficient model, where only the distance ( $r$ ) between center of masses of the drug molecule and membrane is used. For clarity, we name our model (MFEP from a 2D PMF) the 2D model and the model that uses only center of mass distance the 1D model. The free

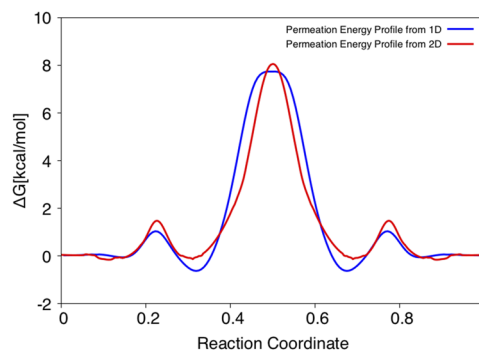


FIG. 4. The permeation energy profile from MFEP (2D) and the center of mass translation (1D) calculated from TTMetaD. The reaction coordinate represents the permeation process from one side of the lipid bilayer to the other: 0 and 1 correspond to  $(-4 \text{ nm}, -4 \text{ nm})$  and  $(4 \text{ nm}, 4 \text{ nm})$  in the 2D case. The same values correspond to  $(-4 \text{ nm})$  and  $(4 \text{ nm})$  in the 1D case.

energy profile for the permeation in the 1D model,  $W(r)$  in Eq. (1), was computed as follows: To avoid unnecessary computation, we reused two sets of TTMetaD simulations from our previous work<sup>40</sup> in which the CVs were the center of mass distance and an angle (the details of the CV definitions can be found in Fig. 2 of Ref. 40). As noted earlier, these two sets of simulations were originally designed to test the robustness of TTMetaD to the choice of CVs. Using Eq. (12), we integrated over the angle CV in each 2D PMF to obtain the 1D free energy profile for the permeation,  $W(r)$ . Again, for a more accurate estimate of the permeation PMF, each of the five replicas in these two sets of simulations was extended to 1100 ns from our previous simulations,<sup>40</sup> making the total simulation time (11  $\mu\text{s}$ ) for the 1D model comparable to that for the 2D model (10.5  $\mu\text{s}$ ). The permeation energy profile for the 1D model is depicted as the blue curve in Fig. 4.

The approach described above is quite different from a conventional 1D free energy calculation, e.g., a TTMetaD simulation with the center of mass distance as the only CV. Our permeation free energy profile for the 1D model was mapped from two TTMetaD PMFs, each with two collective variables. In addition to avoiding unnecessary new simulations, we consider our approach to be superior to a conventional 1D free energy calculation because biasing the extra CV enhanced the sampling of the center of mass distance and eliminated the hysteresis from the orientation of the drug molecule (see Sec. II). The permeation energy profiles from 1D and 2D models agree reasonably well, bearing  $\sim 0.2$  kcal/mol differences near the center and head group of the lipid bilayer. The evidence for the convergence of these calculations has been well-discussed in our previous work.<sup>40</sup> This energy profile (blue curve in Fig. 4) was then used to compute the permeability coefficient with the local diffusion coefficient computed from 1D umbrella sampling (restraining only  $z$ ), as described below.

#### B. Local diffusivity coefficient

For the 2D model, 2D umbrella sampling was employed to restrain both CVs of the system to collect a time series of the reaction coordinate (i.e., the motion of trimethoprim) along the MFEP. The autocorrelation function of the reaction coordinate

is depicted in Fig. 5 for trimethoprim in the bulk water (top panel), at the water-bilayer interface (middle panel), and inside the lipid bilayer (bottom panel).

The autocorrelation function of the reaction coordinate,  $C_r(t)$  in Eq. (5), needs to decay to zero in order to make the integral converge asymptotically to a finite value (so that the local diffusion coefficient is a non-zero value). This is not particularly a problem when the trimethoprim is solvated in bulk water—the autocorrelation function (blue curve in Fig. 5) decays to zero (the gray line in Fig. 5) and oscillates around it. It has been shown that such oscillations in the autocorrelation function can happen even at long times with extensive sampling.<sup>44</sup> However, a heterogeneous environment can delay the convergence of the autocorrelation function, which is exactly the case when the trimethoprim molecule is near the interface and inside the lipid bilayer (middle and bottom panels of Fig. 5). This behavior is consistent with the findings of Lee *et al.*,<sup>29</sup> where the autocorrelation function of the urea restrained at 10 Å and 0 Å away from the center of the membrane only decayed to 8% and 13% of its initial values in 5 ps, respectively. Other work<sup>59</sup> has shown that the convergence of an autocorrelation function for a molecule restrained in lipid bilayers can require microsecond to millisecond of simulations. Through biased and unbiased simulations, Chipot and Comer<sup>60</sup> showed that one-dimensional translocation of methanol across a pure lipid membrane is sub-diffusive instead of a typical diffusive motion. In a follow-up paper,<sup>61</sup> Chipot and Comer used the fractional version of the solubility-diffusion model to take into account the sub-diffusion motion. Their study of the permeation of short-chain alcohols shows that the relative permeability coefficients are

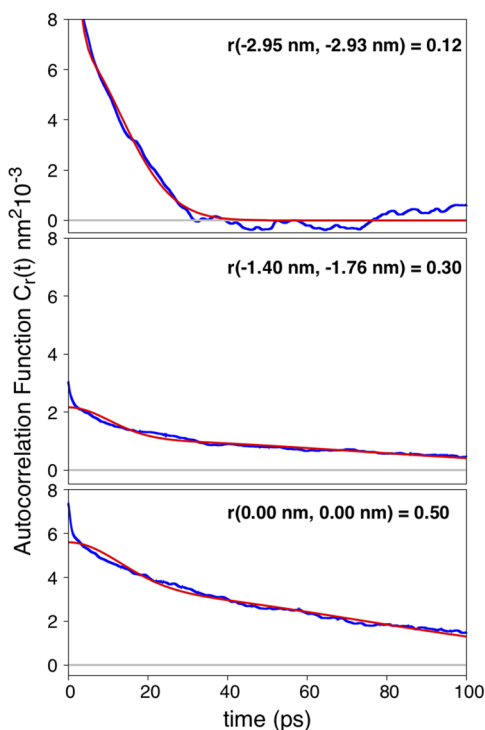


FIG. 5. The autocorrelation function of the diffusion of trimethoprim along the MFEP in Fig. 3. The blue and red curves represent the data from 2D umbrella simulations and the fitted result from Eq. (13), respectively.

highly correlated with existing experimental results, albeit their values are off by three to four orders of magnitude, worse than their previous work with the common solubility-diffusion model.<sup>29</sup> All of the aforementioned research suggests that the slow diffusion of the lipid tails and the inhomogeneity of the bilayer interface slow down the convergence of autocorrelation calculations and therefore present a very challenging problem in computing the local diffusivity coefficient. However, since the benefit from the sub-diffusion model in permeability coefficient calculation is not obvious,<sup>61</sup> we decided to follow the common solubility-diffusion model in this research.

There have also been methods proposed to improve the autocorrelation convergence in heterogeneous environments. For example, using a cut off distance for the integral, which is usually chosen as a value when the autocorrelation function drops below a threshold value of 5% of the variance, has been suggested to limit the contribution of noise.<sup>62</sup> On the other hand, the goal of this work is not to understand the sub-diffusive nature of small molecules inside the lipid bilayer nor to compute the most accurate local diffusivity; instead, we are interested in computing the local diffusivity with reasonable accuracy and a computational cost so that it can be applied to small molecule drug screening. Therefore, we have chosen to fit the autocorrelation function to a function converging asymptotically. To increase the quality of the fitting, the autocorrelation function calculation was gathered from a 100 ns restrained simulation (50 ns equilibration and 50 ns production) and was evaluated to 100 ps, much longer than the common practice of 20 ns of sampling and 5 ps correlation time.<sup>29</sup> Many functional forms have been tested to minimize the root mean square deviation (RMSD) between the fitted function and the autocorrelation data. The summation of two Gaussian functions [Eq. (13)] gives the best overall fitting for trimethoprim at different positions along the MFEP,

$$C_r(t) = a_1 \exp\left[-\left(t^2/b_1\right)\right] + a_2 \exp\left[-\left(t^2/b_2\right)\right], \quad (13)$$

where  $a_1$ ,  $a_2$ ,  $b_1$ , and  $b_2$  are the fit parameters. Good agreement between the fitted curve (red) and the autocorrelation data is demonstrated in Fig. 5. The local diffusivity coefficient along the MFEP is shown in the red curve in Fig. 6. We have also computed the local diffusivity coefficients of the center of mass translation along the  $z$  direction for the 1D case with the same approach and the data are also shown in the blue curve in Fig. 6.

The local diffusivity coefficients of trimethoprim from the 2D and the 1D umbrella sampling show a very similar trend: the drug diffuses an order of magnitude slower in the lipid tail region than it does in the bulk water. This result qualitatively agrees well with the diffusivity coefficient calculation of different drugs reported by Lee *et al.*<sup>29</sup> The inset in Fig. 6 illustrates the local diffusivity when trimethoprim is in the tail region of the lipid, demonstrating trimethoprim diffuses on average ~40% slower along the MFEP (i.e., the molecule is held at a certain configuration while diffusing) than it does in the  $z$  direction (center of mass translation diffusion). This difference plays a role in the permeability coefficient

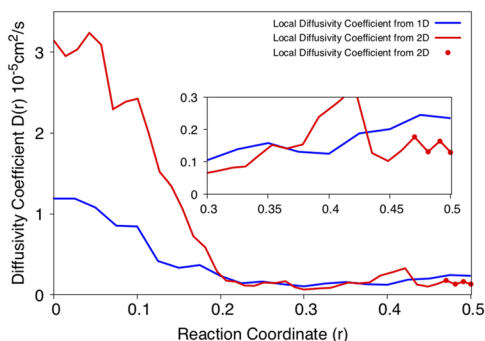


FIG. 6. The local diffusivity coefficient of the diffusion of trimethoprim from 2D umbrella simulations (along the MFEP in Fig. 3) and 1D umbrella simulations (along the blue curve in Fig. 4). The reaction coordinate represents the whole permeation process from one side of the membrane to the other: 0 and 1 correspond to  $(-4 \text{ nm}, -4 \text{ nm})$  and  $(4 \text{ nm}, 4 \text{ nm})$  in the 2D case. The same values correspond to  $-4 \text{ nm}$  and  $4 \text{ nm}$  in the 1D case. 0.5 is the center of the membrane. Only half of the permeation system is shown in this figure. The inset highlights the local diffusivity when trimethoprim is in the tail region of the lipid. Dots denote the essential points used in the permeability calculation below.

calculation because the configurations of trimethoprim inside the lipid correspond to high free energy [Eq. (1)].

### C. Comparison with experimental permeability coefficient

The permeability coefficient from 2D PMF was computed from Eq. (1) by integrating from one side of the membrane [ $r = 0$ , corresponding to  $(-4 \text{ nm}, -4 \text{ nm})$  on Fig. 3] to the other side [ $r = 1$ , corresponding to  $(4 \text{ nm}, 4 \text{ nm})$  on Fig. 3]. The trapezoidal rule was used for the integration with  $\Delta r = 0.0125$ . The permeability coefficient from the center of mass translation (1D) was also computed from Eq. (1) by integrating it from  $r = 0$  ( $z = -4 \text{ nm}$ ) to  $r = 1$  ( $z = 4 \text{ nm}$ ). The results of the permeability coefficient from those calculations are summarized in Table I.

It is of interest to compare the permeability coefficient from simulations with those from experiments. Some degree of disagreement is expected due to the differences between the simulation conducted with homogeneous lipid bilayers and the experimental lipid bilayers made up of different kinds of lipids

TABLE I. The permeability coefficient ( $P_m$ ) from different methods.

Method	Lipid bilayer/membrane	$P_m$ (nm/s)
<i>In vitro</i>	MDCK <sup>a,b</sup>	374
<i>In vitro</i>	PAMPA (egg lecithin) <sup>c,d</sup>	72
<i>In vitro</i>	PAMPA (skin) <sup>e,d</sup>	170
<i>In silico</i> (1D) <sup>f</sup>	POPC	1732
<i>In silico</i> (2D) <sup>g</sup>	POPC	1885
<i>In silico</i> (2D) <sup>h</sup>	POPC	587.5

<sup>a</sup>Experimental data are from Ref. 63.

<sup>b</sup>The permeability is scaled from 310 K to 323 K according to Ref. 64.

<sup>c</sup>Experimental data are from Ref. 65.

<sup>d</sup>The permeability is scaled from 310 K to 323 K according to Ref. 66.

<sup>e</sup>Experimental data are from Ref. 67.

<sup>f</sup>Total TTMetaD simulation time is 11  $\mu\text{s}$ : 10 replicas with 1.1  $\mu\text{s}$ /replica.

<sup>g</sup>Total TTMetaD simulation time is 2.5  $\mu\text{s}$ : 5 replicas with 0.5  $\mu\text{s}$ /replica.

<sup>h</sup>Extended from g (16 replicas with 0.5  $\mu\text{s}$ /replica), total TTMetaD simulation time is 10.5  $\mu\text{s}$ .

and sometimes proteins (e.g., the MDCK assay). Another difference between our simulations and experiments are the temperatures. To take advantage of the extensive simulation data from our previous work, the permeation of trimethoprim was simulated at 323 K, i.e., higher than the physiological temperature at which the drug permeability experiments are usually conducted. However, various efforts have focused on determining the temperature dependence of permeability coefficients measured by experiment. Sinkó and co-workers<sup>66</sup> measured the effective permeability ( $\log P_e$ ) of seven drugs representing diverse structures and different acid–base properties on three different PAMPA models (GIT, BBB, and Skin) with the temperature range of 288 K–328 K. Their result shows that the slope of the  $\log P_e = aT + b$  regression equation is a good measure of the temperature effect on permeability and suggests the values of parameters  $a$  and  $b$  for small molecule drugs. Kulkarni *et al.*<sup>64</sup> studied the influence of the experimental temperature on the permeability coefficients with cell-based assays. Their research concludes that the permeability coefficient and temperature are related by an exponential relationship that conformed to the Arrhenius equation, similar to the PAMPA experiments.<sup>66</sup> As a result, we are able to estimate the permeability coefficient of trimethoprim at 323 K from the cell-based assay (MDCK) experiment<sup>63</sup> and the artificial membrane assay (PAMPA) experiment<sup>65,67</sup> conducted at 310 K. The results of the permeability coefficient from those experiments are summarized in Table I. As discussed before, even though there are some disagreements between the experimental permeability coefficients due to differences in the measurements and perturbations, all the experiments yield a permeability coefficient value between  $\sim 75$  and  $\sim 400$  nm/s.

## IV. DISCUSSION

### A. Permeation along the MFEP

The results in Table I demonstrate improved accuracy when permeation is modeled along the MFEP compared to those using the  $z$ -direction center of mass translation. The permeability coefficient predicted from the 2D permeation PMF (last row in Table I) is only  $\sim 2$ – $5$  times larger than the experiment. This is a small difference compared to recent work reporting errors of one to two orders of magnitude.<sup>29</sup> This result also supports our expectation that small molecules with groups of different hydrophobicity (e.g., trimethoprim) have a more complicated permeation pathway that optimizes hydrophilic–hydrophobic interactions. For these cases, the local diffusivity is best represented by the diffusive motion along the MFEP, as opposed to the oversimplified, and most commonly used model (i.e., diffusion of the center of mass along the direction perpendicular to the lipid bilayers).

It is important to appreciate, however, that the quality of the permeation free energy profile,  $W(r)$  in Eq. (1), governs the accuracy of the permeability coefficient more significantly than the local diffusivity coefficient,  $D(r)$ . This partially explains why the *in silico* permeability coefficients reported in the early 2000's were usually off by a few orders of magnitude, when the sampling capacity, and thus accuracy of  $W(r)$ ,

was more limited.<sup>68–70</sup> Herein, the 1D permeation free energy from TTMetaD, even applied with the common practice of local diffusivity (the  $z$ -direction-only; see the third to last row in Table I), is able to predict a permeability coefficient within one order of magnitude of the experimental value, highlighting the ability to calculate a reliable PMF within reasonable simulation times.

## B. Computational cost of screening drug candidates

As reviewed earlier in this paper, computational simulation of small molecule drug permeation provides mechanistic information at atomistic resolution that can be quite valuable in the pharmaceutical industry. Predicting the permeability coefficients of small molecule drugs before synthesis is environmentally friendly and has the potential to decrease cost during pharmaceutical design and development. To accomplish this goal, the simulation methodology must provide reliable permeability estimates with low to moderate computational effort and within a reasonable time frame.

Moreover, instead of the absolute permeability coefficient of each molecule, the relative permeability coefficients of a series of molecular candidates are often sought for screening purposes. We have shown herein that calculating the diffusion along the MFEP of an extensively sampled 2D PMF is more accurate than simply using the  $z$ -direction. But how does this method work when sampling times are limited? To answer this, we revisited our previous simulations<sup>40</sup> and used the 5 TTMetaD replicas (500 ns each) to compute an average PMF (Fig. 1). The MFEP is computed on this PMF to estimate  $W(r)$  following Eq. (12). The local diffusivity coefficient was approximated as the same  $D(z)$  from the converged PMF (Fig. 6). The resulting permeability coefficient is shown in the second to last row in Table I. Even though the total simulation time used to calculate this 2D PMF is only a quarter of the simulation time used to obtain the 1D simulation, the results are comparable. More importantly, the permeability coefficient computed from this very limited sampling is within one order of magnitude of the experimental result. Comparing this result with the fully converged 2D TTMetaD calculation (last row in Table I), it is fair to conclude that the accuracy of the permeation free energy profile is less with less sampling, but a significant amount of computational time is saved (2.5  $\mu$ s vs. 10.5  $\mu$ s) and the result is still reasonable.<sup>29</sup>

The permeability coefficient is dominated by the permeation free energy [Eq. (1)] because of the exponential dependence. Accordingly, the local diffusion along the highest free energy regions of the PMF will play the most significant role in the integration of Eq. (1). Therefore, we also explored the minimum number of local diffusivity coefficients that are needed for a reasonable estimate of the permeability coefficient. For this purpose, the integral [Eq. (1)] was evaluated from the center of the lipid bilayer [ $r = 0.5$ , corresponding to (0 nm, 0 nm) in Fig. 3] to the outside of the lipid bilayer. The predicted permeability coefficient vs. the number of restrained calculations (i.e., 100 ns of 2D umbrella samplings per calculation) is shown in Fig. 7. We conclude that only four of the points on the MFEP corresponding to the highest free energy are needed to estimate the permeability coefficient within 10% of the error of the converged (most accurate) value. These four points

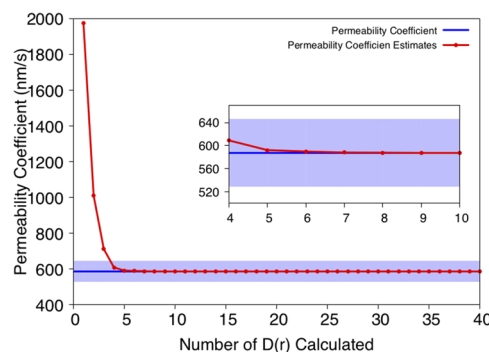


FIG. 7. The estimate of permeability coefficient vs. number of local diffusivities calculated from 2D umbrella sampling.

correspond to the trimethoprim in the center of the membrane and are highlighted with red dots in Fig. 6.

From the above tests, we conclude that 2.9  $\mu$ s of simulation are needed for the present method to predict a permeability coefficient that agrees to within one order of magnitude of the experimental result (500 ns of TTMetaD with 5 replicas and 100 ns of 2D umbrella sampling on 4 points). These simulations produced  $\sim$ 250 ns/day on a graphics processing unit (GPU) node with four NVIDIA GeForce GTX TITAN X cores; therefore, the method took less than 60 h of wall time to screen a small molecule drug candidate with five such GPU nodes. With these and continued improved efficiencies, simulation-based permeability coefficients may be an increasingly attractive option in the pharmaceutical industry and academia.

It should be emphasized that assessing the permeability of small molecule drugs through a lipid bilayer remains very challenging both *in vitro* and *in silico*. For example, the inhomogeneous solubility-diffusion [Eq. (1)] requires a very accurate free energy profile of the permeation; a difference of  $k_B T$  in  $W(r)$  introduces approximately 2.72-fold difference in the permeability coefficient. Given this level of sensitivity to  $W(r)$ , combined with the errors from the force field parameters and the limitations in the complexity of the lipid bilayer models, we do not anticipate this level of agreement (within one order of magnitude) between *in vitro* and *in silico* studies to be universally obtained—especially considering the diversity of drug molecules and lipid bilayers. Nevertheless, our model improves the physical accuracy of permeation coefficient calculation and contributes to the collective progress toward accurate *in silico* studies of drug permeation. Currently, we are applying this methodology to a series of drugs to verify its advantages in predicting relative permeability coefficients. We believe its advantages will be further demonstrated in the comparison of relative permeability where the errors from the force field and lipids may cancel out.

As a final note, the enhanced sampling method described in this paper, TTMetaD, is available in the most recent version of PLUMED (v.2.4). TTMetaD is easy to understand and apply, and, compared to a conventional non-tempered MetaD or WTMetaD simulation, the extra input parameters that TTMetaD requires are straightforward to incorporate. A detailed user manual for this method is also available upon

request, and the new features require no specialized or separate mathematical library.

## ACKNOWLEDGMENTS

This research was supported in part by the Eli Lilly and Company and in part by the National Science Foundation (NSF Grant No. CHE-1465248). Simulations were performed using resources provided by the University of Chicago Research Computing Center (RCC) and the San Diego Supercomputing Center (Comet) through the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation Grant No. ACI-1053575.

- <sup>1</sup>C. R. Ganellin, R. Jefferis, and S. M. Roberts, *Introduction to Biological and Small Molecule Drug Research and Development: Theory and Case Studies* (Elsevier Science, 2013).
- <sup>2</sup>J. Samanen, *Introduction to Biological and Small Molecule Drug Research and Development* (Elsevier, Oxford, 2013), pp. 161.
- <sup>3</sup>M. Stocks, *Introduction to Biological and Small Molecule Drug Research and Development* (Elsevier, Oxford, 2013), pp. 81.
- <sup>4</sup>A. Avdeef, *Absorption and Drug Development* (John Wiley & Sons, Inc., 2003), pp. xxiv.
- <sup>5</sup>A. Pagliara, M. Reist, S. Geinoz, P.-A. Carrupt, and B. Testa, *J. Pharm. Pharmacol.* **51**, 1339 (1999).
- <sup>6</sup>J. T. Penniston, L. Beckett, D. L. Bentley, and C. Hansch, *Mol. Pharmacol.* **5**, 333 (1969); available at <http://molpharm.aspetjournals.org/content/5/4/333.long>.
- <sup>7</sup>B. Freeman, Y. Yampolskii, and I. Pinnau, *Materials Science of Membranes for Gas and Vapor Separation* (Wiley, 2006).
- <sup>8</sup>G. M. Cooper and D. Ganem, *Nat. Med.* **3**, 1042 (1997).
- <sup>9</sup>M. Orsi and J. W. Essex, *Molecular Simulations and Biomembranes: From Biophysics to Function* (The Royal Society of Chemistry, 2010), pp. 76.
- <sup>10</sup>T.-X. Xiang and B. D. Anderson, *Adv. Drug Delivery Rev.* **58**, 1357 (2006).
- <sup>11</sup>D. J. Benos, D. W. Deamer, A. Kleinzeller, and D. M. Fambrough, *Membrane Permeability: 100 Years Since Ernest Overton* (Elsevier Science, 1999).
- <sup>12</sup>A. Walter and J. Gutknecht, *J. Membr. Biol.* **77**, 255 (1984).
- <sup>13</sup>T. X. Xiang, X. Chen, and B. D. Anderson, *Biophys. J.* **63**, 78 (1992).
- <sup>14</sup>J. Karlsson and P. Artursson, *Int. J. Pharm.* **71**, 55 (1991).
- <sup>15</sup>A. R. Hilgers, R. A. Conradi, and P. S. Burton, *Pharm. Res.* **7**, 902 (1990).
- <sup>16</sup>J. D. Irvine, L. Takahashi, K. Lockhart, J. Cheong, J. W. Tolan, H. E. Selick, and J. R. Grove, *J. Pharm. Sci.* **88**, 28 (1999).
- <sup>17</sup>M. Karsy, F. Senner, and K. Gubernator, *J. Med. Chem.* **41**, 1007 (1998).
- <sup>18</sup>A. Avdeef, M. Strafford, E. Block, M. P. Balogh, W. Chambliss, and I. Khan, *Eur. J. Pharm. Sci.* **14**, 271 (2001).
- <sup>19</sup>A. Avdeef, P. Artursson, S. Neuhoff, L. Lazorova, J. Gråsjö, and S. Tavelin, *Eur. J. Pharm. Sci.* **24**, 333 (2005).
- <sup>20</sup>A. L. Hopkins, G. M. Keseru, P. D. Leeson, D. C. Rees, and C. H. Reynolds, *Nat. Rev. Drug Discovery* **13**, 105 (2014).
- <sup>21</sup>K. Sugano, H. Hamada, M. Machida, and H. Ushio, *J. Biomol. Screening* **6**, 189 (2001).
- <sup>22</sup>X. Jin, T.-L. Luong, N. Reese, H. Gaona, V. Collazo-Velez, C. Vuong, B. Potter, J. C. Sousa, R. Olmeda, Q. Li, L. Xie, J. Zhang, P. Zhang, G. Reichard, V. Melendez, S. R. Marcisin, and B. S. Pybus, *J. Pharmacol. Toxicol. Methods* **70**, 188 (2014).
- <sup>23</sup>D. A. Volpe, *Future Med. Chem.* **3**, 2063 (2011).
- <sup>24</sup>D. A. Volpe, *J. Pharm. Sci.* **97**, 712 (2008).
- <sup>25</sup>D. Huster, A. J. Jin, K. Arnold, and K. Gawrisch, *Biophys. J.* **73**, 855 (1997).
- <sup>26</sup>A. Finkelstein and A. Cass, *J. Gen. Physiol.* **52**, 145 (1968).
- <sup>27</sup>J. M. Diamond and Y. Katz, *J. Membr. Biol.* **17**, 121 (1974).
- <sup>28</sup>S.-J. Marrink and H. J. C. Berendsen, *J. Phys. Chem.* **98**, 4155 (1994).
- <sup>29</sup>C. T. Lee, J. Comer, C. Herndon, N. Leung, A. Pavlova, R. V. Swift, C. Tung, C. N. Rowley, R. E. Amaro, C. Chipot, Y. Wang, and J. C. Gumbart, *J. Chem. Inf. Model.* **56**, 721 (2016).
- <sup>30</sup>B. J. Bennion, N. A. Be, M. W. McNerney, V. Lao, E. M. Carlson, C. A. Valdez, M. A. Malfatti, H. A. Enright, T. H. Nguyen, F. C. Lightstone, and T. S. Carpenter, *J. Phys. Chem. B* **121**, 5228 (2017).
- <sup>31</sup>G. M. Torrie and J. P. Valleau, *J. Comput. Phys.* **23**, 187 (1977).
- <sup>32</sup>A. J. Sodt, R. W. Pastor, and E. Lyman, *Biophys. J.* **109**, 948 (2015).
- <sup>33</sup>E. Darve, *J. Chem. Phys.* **115**, 9169 (2001).
- <sup>34</sup>D. R.-G. Eric Darve, A. Pohorille, *J. Chem. Phys.* **128**, 144120 (2008).
- <sup>35</sup>R. H. Swendsen and J.-S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986).
- <sup>36</sup>A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12562 (2002).
- <sup>37</sup>A. Barducci, G. Bussi, and M. Parrinello, *Phys. Rev. Lett.* **100**, 020603 (2008).
- <sup>38</sup>J. F. Dama, M. Parrinello, and G. A. Voth, *Phys. Rev. Lett.* **112**, 240602 (2014).
- <sup>39</sup>J. F. Dama, G. Rotskoff, M. Parrinello, and G. A. Voth, *J. Chem. Theory Comput.* **10**, 3626 (2014).
- <sup>40</sup>R. Sun, J. F. Dama, J. S. Tan, J. P. Rose, and G. A. Voth, *J. Chem. Theory Comput.* **12**, 5157 (2016).
- <sup>41</sup>A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello, *J. Phys. Chem. B* **109**, 6714 (2005).
- <sup>42</sup>R. Sun, O. Sode, J. F. Dama, and G. A. Voth, *J. Chem. Theory Comput.* **13**, 2332 (2017).
- <sup>43</sup>B. J. Berne, M. Borkovec, and J. E. Straub, *J. Phys. Chem.* **92**, 3711 (1988).
- <sup>44</sup>G. Hummer, *New J. Phys.* **7**, 34 (2005).
- <sup>45</sup>Z. Ghaemi, M. Minozzi, P. Carloni, and A. Laio, *J. Phys. Chem. B* **116**, 8714 (2012).
- <sup>46</sup>J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, and R. W. Pastor, *J. Phys. Chem. B* **114**, 7830 (2010).
- <sup>47</sup>K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell, *J. Comput. Chem.* **31**, 671 (2010).
- <sup>48</sup>W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, *J. Chem. Phys.* **79**, 926 (1983).
- <sup>49</sup>G. Bussi, D. Donadio, and M. Parrinello, *J. Chem. Phys.* **126**, 014101 (2007).
- <sup>50</sup>H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *J. Chem. Phys.* **81**, 3684 (1984).
- <sup>51</sup>U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).
- <sup>52</sup>B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, *J. Comput. Chem.* **18**, 1463 (1997).
- <sup>53</sup>M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, *SoftwareX* **1–2**, 19 (2015).
- <sup>54</sup>G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, *Comput. Phys. Commun.* **185**, 604 (2014).
- <sup>55</sup>J. F. Dama, G. M. Hocky, R. Sun, and G. A. Voth, *J. Chem. Theory Comput.* **11**, 5638 (2015).
- <sup>56</sup>A. Laio and F. L. Gervasio, *Rep. Prog. Phys.* **71**, 126601 (2008).
- <sup>57</sup>A. Barducci, M. Bonomi, and M. Parrinello, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **1**, 826 (2011).
- <sup>58</sup>E. Weinan, W. Ren, and E. Vanden-Eijnden, *J. Chem. Phys.* **126**, 164103 (2007).
- <sup>59</sup>C. Neale, W. F. D. Bennett, D. P. Tieleman, and R. Pomès, *J. Chem. Theory Comput.* **7**, 4175 (2011).
- <sup>60</sup>C. Chipot and J. Comer, *Sci. Rep.* **6**, 35913 (2016).
- <sup>61</sup>J. Comer, K. Schulten, and C. Chipot, *J. Chem. Theory Comput.* **13**, 2523 (2017).
- <sup>62</sup>J. S. Hub, B. L. de Groot, and D. van der Spoel, *J. Chem. Theory Comput.* **6**, 3713 (2010).
- <sup>63</sup>K. M. M. Doan, J. E. Humphreys, L. O. Webster, S. A. Wring, L. J. Shampine, C. J. Serabjit-Singh, K. K. Adkison, and J. W. Polli, *J. Pharmacol. Exp. Ther.* **303**, 1029 (2002).
- <sup>64</sup>U. D. Kulkarni, R. Mahalingam, X. Li, I. Pather, and B. Jasti, *AAPS PharmSciTech* **12**, 579 (2011).
- <sup>65</sup>C. Zhu, L. Jiang, T.-M. Chen, and K.-K. Hwang, *Eur. J. Med. Chem.* **37**, 399 (2002).
- <sup>66</sup>G. Vizserálek, T. Balogh, K. Takács-Novák, and B. Sinkó, *Eur. J. Pharm. Sci.* **53**, 45 (2014).
- <sup>67</sup>T. V. Chagas, L. S. Teixeira, and K. R. Rezende, <http://cifarp.com.br/historico/2013/cd2011/abstracts/PP%20042%20-%20568.pdf>.
- <sup>68</sup>D. J. V. A. dos Santos and L. A. Eriksson, *Biophys. J.* **91**, 2464 (2006).
- <sup>69</sup>D. Bemporad, C. Luttmann, and J. W. Essex, *Biochim. Biophys. Acta, Biomembr.* **1718**, 1 (2005).
- <sup>70</sup>J. Ulander and A. D. J. Haymet, *Biophys. J.* **85**, 3475 (2003).