Position-Dependent Diffusion from Biased Simulations and Markov State Model Analysis

François Sicard$^{1,2,*}$, Vladimir Koskin$^{1,*}$, Alessia Annibale$^{3}$, and Edina Rosta$^{1,2,†}$

$^1$Department of Chemistry, King’s College London, SE1 1DB London, UK
$^2$Department of Physics and Astronomy, University College London, WC1E 6BT London, UK
$^3$Department of Mathematics, King’s College London, SE11 6NJ London, UK

A variety of enhanced statistical and numerical methods are now routinely used to extract important thermodynamic and kinetic information from the vast amount of complex, high-dimensional data obtained from molecular simulations. For the characterization of kinetic properties, Markov state models, in which the long-time statistical dynamics of a system is approximated by a Markov chain on a discrete partition of configuration space, have seen widespread use in recent years. However, obtaining kinetic properties for molecular systems with high energy barriers remains challenging as often enhanced sampling techniques are required with biased simulations to observe the relevant rare events. Particularly, the calculation of diffusion coefficients remains elusive from biased molecular simulation data. Here, we propose a novel method that can calculate position-dependent diffusion coefficients equally from either biased or unbiased simulations using the same formalism. Our method builds on Markov state model analysis and the Kramers-Moyal expansion. We demonstrate the validity of our formalism using one- and two-dimensional analytic potentials and also apply it to data from explicit solvent molecular dynamics simulations, including the water-mediated conformations of alanine dipeptide and umbrella sampling simulations of drug transport across a lipid bilayer. Importantly, the developed algorithm presents significant improvement compared to standard methods when the transport of solute across three-dimensional heterogeneous porous media is studied, for example, the prediction of membrane permeation of drug molecules.

I. Introduction

For more than a century, mathematical methods have been developed to describe the stochastic evolution of complex dynamical systems, with current applications in physics, chemistry, biology, engineering and finance [1–10]. Of particular interest in toxicology and pharmacology is the prediction of kinetic quantities such as transition rates, permeation coefficients, and mean first-passage times of solutes, including drugs and small peptides, which provide fundamental understanding of numerous biochemical transport processes [9–11]. As originally described by Hendrik A. Kramers in his seminal work [12], such quantities can be calculated by employing models of diffusive motion. In the simplest diffusive model, one can ignore inertia and memory effects and regard the diffusive motion as the random walk of a particle under a position-dependent potential [13, 14]. The dynamics of the reaction coordinate, $X(t)$, is determined by two functions: the potential of mean force, $V(X)$, and the diffusivity, $D(X)$, along this coordinate. In general, $V(X)$ and $D(X)$ are both position-dependent and are likely to vary substantially in heterogeneous systems.

A variety of numerical methods for calculating $V(X)$ based on enhanced sampling techniques are now well-established. These employ biasing potentials along a few selected degrees of freedom, i.e., reaction coordinates (RC), to overcome the challenges of Boltzmann sampling in the presence of rare events in which energy barriers and transition states may be poorly sampled or not sampled at all [14–23]. The calculation of $D(X)$, on the other hand, has received less attention. The Einstein-Smoluchowski relation, which relates the diffusivity to the mean square deviation of the position of the solute in the long-time limit, can be used to calculate the diffusion coefficient of a solute in a homogeneous solution by analysing molecular dynamics (MD) trajectories [4, 24–26]. This relationship, though, might offer a less accurate approximation of the true diffusivity in inhomogeneous systems such as a bilayer [10]. In these systems, the variation of the solute diffusivity is large because the frictional environment varies dramatically as the solute moves from bulk water through the interface, and into the membrane interior, potentially encountering free energy barriers with heights greater than $k_B T$. For similar reasons, estimates based on a Green-Kubo relation of the velocity are also expected to be less accurate [27, 28].

To circumvent this limitation, Marrink and Berendsen calculated the diffusivity profile for the permeation of water using the force autocorrelation function [29, 30]. This method requires the solute to be constrained to a point along the RC through the modification of the equation of motion of the MD integration. In practice, it is more convenient to perform simulations where the solute is simply restrained to remain near a given position with a biasing potential. Among the latter, two strategies have been commonly employed for calculating $D(X)$ using biased MD simulations. The first is based on the generalized
Langevin equation for a harmonic oscillator [9, 10]. The second employs Bayesian inferences on the likelihood of the observed dynamics of the solute [8, 9, 31].

The generalized Langevin equation provides a useful method to calculate position-dependent diffusion coefficients from restrained MD simulations, where the dynamics of the system is described as a strongly restrained harmonic oscillator undergoing Langevin dynamics. As discussed in [10], the spring constants of the restraining potential that are sufficiently large to justify this assumption could be too large to unbiase the same umbrella sampling (US) simulations. Once a time series of the \( X(t) \) position of the quantity of interest is collected, the diffusion coefficient can be calculated from the position or velocity autocorrelation functions (PACF and VACF, respectively). These methods were first introduced by Berne and co-workers [32] and further developed by Woolf and Roux to calculate position-dependent diffusion coefficients [33]. Hummer proposed a simpler method to calculate diffusion coefficients from harmonically restrained simulations which avoid the need for multiple numerical Laplace transforms of the VACF [4]. In practical applications, convergence of PACFs can be difficult to achieve, particularly for heterogeneous bilayer environments where the PACF does not decay to values near zero at long time scales due to the lack of ergodicity in sampling [9]. Furthermore, these algorithms are only applicable to restrained MD simulations, and unbiased simulations cannot be analysed with PACF or VACF.

A fundamentally different approach for the determination of position-dependent diffusivities employs Bayesian inferences. In this method originally developed by Hummer [4] and further employed by Türkcan et al. [34] and Comer et al. [8], no assumptions are made regarding the form of the free energy landscape and the diffusion coefficient and free energies are estimated self-consistently. Using Bayes’ formula to maximise the likelihood function of observations, the posterior density of the unknown parameters are determined, which provides values for the force \( F(X,t) \) and the diffusion coefficient \( D(X) \) within the diffusive model considered. However, it can be numerically difficult for the Bayesian scheme to find unique solutions in a large parameter space, and greater sampling may be needed [4, 8]. Additionally, a crucial component of this scheme is the Brownian integrator time step, which needs to be larger than any correlation time of the system [9].

To assess multi-dimensional position-dependent drift and diffusion coefficients, equally valid in both biased and unbiased simulations, we propose a novel general method built on Markov state model analysis and the Kramers-Moyal (KM) expansion [6, 35–39]. We employ the dynamic histogram analysis method (DHAM) [40] to construct global Markov state models (MSMs) from the time series generated by stochastic processes. We use the resulting transition probabilities between discretized values of the RCs to calculate position-dependent drift and diffusion coefficients, using their relations with the KM coefficients. DHAM has been successfully employed to compute stationary quantities and long-time kinetics of molecular systems with significantly high energy barriers, for which transition states can be poorly sampled or not sampled at all with unbiased MD simulations [11]. In this context, MSMs are extremely popular because they can be used to compute stationary quantities and long-time kinetics from ensembles of short simulations [40–44].

Here we use MSMs to analyze simulation trajectories generated from biased and unbiased simulations. We apply the DHAM method to determine the unbiased transition probabilities and use the KM expansion to assess the original drift and diffusion coefficients. Our method only requires the numerical determination of the KM coefficients and it otherwise neither requires prior assumptions regarding the form of the free energy landscape nor additional numerical integration schemes. While Bayesian methods can in principle also handle both biased and unbiased simulations, our method offers a numerically more tractable approach, as we have significantly fewer parameters that require fitting and no need for a numerical integration of the dynamics of the system. We therefore present significant improvements in determining diffusion coefficients compared to using standard methods both in terms of the simplicity and generality of the analysis.

We apply the formalism to one- and two-dimensional analytic potentials and data from explicit solvent MD simulations, including the water-mediated conformations of alanine dipeptide and the permeation of the Domperidone drug molecule across a lipid membrane. We demonstrate that our algorithm presents significant improvement compared to standard methods (for example PACF) even when long timescale fluctuations are present in complex simulation systems, enabling us to determine e.g., the transport of solutes across three-dimensional heterogeneous porous media.

II. Methods

A. Definition of the Diffusion coefficient

A wide range of dynamical systems can be described with a stochastic differential equation, the non-linear Langevin equation [6, 36, 38, 45–47]. Considering a one-dimensional stochastic trajectory \( X(t) \) in time \( t \), the time derivative of the system’s trajectory \( dX/dt \) can be expressed as the sum of two complementary contributions, one being purely deterministic and another one being stochastic. For a stationary stochastic process, the deterministic term is defined by a function \( D^{(1)}(X) \) and the stochastic contribution is given by another function, \( D^{(2)}(X) \), which do not explicitly depend on time, yielding the evolution equation of \( X \),

\[
\frac{dX}{dt} = D^{(1)}(X) + \sqrt{D^{(2)}(X)} \Gamma(t),
\]

(1)
where $\Gamma(t)$ is a zero-average Gaussian white noise, i.e. $\langle \Gamma(t) \rangle = 0$ and $\langle \Gamma(t) \Gamma(t') \rangle = 2\delta(t-t')$, with $\delta$ the Dirac function. According to the Ito’s prescription, this is equivalent to the Fokker-Planck equation [35]

$$\frac{\partial P(X,t)}{\partial t} = \left[ -\frac{\partial}{\partial x} D^{(1)}(X) + \frac{\partial^2}{\partial x^2} D^{(2)}(X) \right] P(x,t), \quad (2)$$

with stationary solution

$$P(X) \propto \frac{1}{D^{(2)}(X)} \exp \left( \int_X dx \frac{D^{(1)}(x)}{D^{(2)}(x)} \right). \quad (3)$$

Different integration schemes (other than the Ito’s prescription) can be envisaged to integrate Langevin equations with a position-dependent diffusion coefficient, as in Eq. 1, e.g. the Stratonovich convention. Such schemes lead to different drift coefficients (often referred to as ‘anomalous’ drifts) in the Fokker-Planck equation, resulting in different stationary distributions [45, 47–50].

We can evaluate the drift and diffusion coefficients of a stationary Markovian stochastic process by retrieving the Kramers-Moyal coefficients from the time series $X(t)$ [51–53]. The KM coefficients arise from the Taylor expansion of the master equation describing the Markov process, as

$$D^{(n)}(X) = \lim_{\tau \to 0} \frac{1}{n! \tau} c^{(n)}(X,\tau), \quad (4)$$

which can be seen as the derivative with respect to $\tau$ of the $n$-th moment of the stationary conditional probability density function $p(X', t + \tau \mid X, t) = p(X', \tau \mid X)$ [54]

$$c^{(n)}(X, \tau) = \int dX' [X' - X]^n \ p(X', \tau \mid X). \quad (5)$$

Mathematically the drift and diffusion coefficients are defined as the first two KM moments, i.e. $n = 1$ and $n = 2$ in Eq. 4, respectively. Under the ergodic hypothesis, the average over the microstates defined in Eq. 5 can be equivalently replaced by the average over time of the trajectory $X(t)$, defined as

$$c^{(n)}(X, \tau) = \left\langle (X(t+\tau) - X(t))^n \right\rangle_{X(t)=X}. \quad (6)$$

Additionally, the deviation from the stochastic Langevin description given in Eq. 1, i.e. the deviation of the driving noise $\Gamma(t)$ from a Gaussian distribution, can be tested with the Pawula theorem [35]. To do so, one can compute the fourth-order coefficient in the KM expansion, $D^{(4)}(X) = \lim_{\tau \to 0} \frac{1}{\tau^4} c^{(4)}(X, \tau)$ and compare it to the diffusion coefficient, expecting $D^{(4)}(X) \ll (D^{(2)}(X))^2$ at all $X$. In this work, we use MSMS to model stationary time series resulting from trajectories generated with a position-dependent drift $D^{(1)}(X)$ and diffusion coefficient $D^{(2)}(X)$.

### B. Kramers-Moyal and Diffusion coefficients from MSM

Typically, in molecular simulations we assume that the state space of a system evolving stochastically in time can be discretised, and this discretised data leads to a Markovian process. This usually involves the assumption that we have a low-dimensional RC, $X$, along which the time evolution of the system is approximately Markovian. In constructing MSMSs, we discretise the RC, $X$, in $N_{\text{bin}}$ bins, $\{x_i, \ldots, x_{N_{\text{bin}}}, \}$, defining the set of states (also called microstates), which the system can occupy. Each MD trajectory can then be analysed as a series of microstate assignments rather than as a series of conformations. The dynamics of the system is regarded as a memoryless process such that the next state of the system depends only on its present state. The number of transitions between each pair of state $i$ and state $j$ in the lagtime $\tau$ can then be counted and stored as a transition count matrix, $C^{(\tau)(ij)}$, which can be normalised to provide a numerical estimate of the transition probability matrix $M^{(\tau)(ij)}$, at lagtime $\tau$ [40, 55–57]. A spectral decomposition can be performed to write the Markov matrix $M^{(\tau)}$ in terms of its eigenvalues and eigenvectors which provide information about the dynamics of the system, [58]

$$M^{(\tau)}_{ji} = \sum_n \psi^R_n(j) \psi^L_n(i) \lambda_n, \quad (7)$$

where $\psi^R_n$ and $\psi^L_n$ are the right and left eigenvectors, respectively, corresponding to eigenvalues $\lambda_n$. The latter are ordered such that $1 = |\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_N|$. The second largest eigenvalue (in magnitude) of the MSMSs describes the slowest relaxation process in the system. In practice, the slowest relaxation time, $\tau_{\text{relax}} = -\tau / \ln \lambda_2$, determined from MSMSs will have a functional dependence on the lagtime at which the model is constructed, as illustrated in Figs. 1(c), (d), and (e). We can assess such non-Markovian effects by carrying out the Chapman-Kolmogorov test. This is based on the observation that in a truly Markovian process, multiplying the lagtime times $n$ equivalent to raising the transition matrix to power $n$ [59]. The relaxation time of a Markovian system must then be invariant under lagtime changes. The smallest lagtime $\tau$ at which this condition is sufficiently fulfilled is called the Markov timescale, $\tau_M$ [6, 54].

Given the one-dimensional time series of microstates $i \in \{1, \ldots, N_{\text{bin}}\}$, one can rewrite the $n$-th conditional moment in its discrete form

$$c^{(n)}(x_i, \tau) = \sum_{j=1}^{N_{\text{bin}}} (x_j - x_i)^n \ M^{(\tau)}_{ji}, \quad (8)$$

where $x_i$ is the (discretized) value of the RC in the center of bin $i$. In many cases, for a given $X$, $c^{(n)}(X, \tau)$ depends linearly on $\tau$ in the range of $\tau$ for which the diffusive
FIG. 1. (a) Analytical potentials $V_{\text{low}}$ and (b) $V_{\text{high}}$ considered in the 1D Langevin simulations expressed in $k_B T$ units. The lagtime dependence of the slowest relaxation time, $\tau_{\text{relax}}$, in the overdamped Langevin dynamics is shown for (a) the unbiased simulations in $V_{\text{low}}$, (b) the biased simulations in $V_{\text{low}}$, and (c) the biased simulations in $V_{\text{high}}$ for different numbers of bins, $N_{\text{bin}}$. The limiting relaxation times, $\mu_{\text{relax}}$, obtained from Eq. 18 are shown with dashed lines for comparison. We measured $\mu_{\text{relax}} = 5.8 \times 10^3$, $5.1 \times 10^3$, and $1.7 \times 10^{10}$ in panels (c), (d), and (e), respectively. Uncertainties, as represented in shaded area in panels (c), (d), and (e), were estimated from 10 independent runs, determining the profiles independently, and calculating the standard error.

regime is satisfied. Consequently the drift and diffusion coefficients are estimated solely by the quotient between the corresponding conditional moment and the lagtime $\tau$ in this range. Integrating Eq. 1 within the Taylor-Ito framework yields the stochastic Euler equation [45, 47]

$$X(t + \tau) = X(t) + D^{(1)}(X, \tau) + \sqrt{D^{(2)}(X, \tau)} \eta(t),$$

where $\eta(t)$ is a Gaussian white noise with the same average and correlation as $\Gamma(t)$. Inserting Eq. 9 into Eq. 6 yields the relations between the conditional moments and the drift and diffusion coefficients [38]

$$c^{(1)}(X, \tau) \approx D^{(1)}(X, \tau),$$

$$c^{(2)}(X, \tau) \approx 2D^{(2)}(X, \tau) + (D^{(1)}(X, \tau))^2.$$  \hspace{1cm} (11)

The latter can be subsequently assessed from the slope of a weighted polynomial regression of Eqs. 10 and 11.

Similarly to the one-dimensional case discussed above, the two-dimensional case comprehends two stochastic variables, $X(t)$ and $Y(t)$, governed by the stochastic differential equation [6, 38]

$$\frac{d}{dt} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} D^{(1)}_1(X, Y) \\ D^{(1)}_2(X, Y) \end{bmatrix} + \begin{bmatrix} g_{11}(X, Y) & g_{12}(X, Y) \\ g_{21}(X, Y) & g_{22}(X, Y) \end{bmatrix} \begin{bmatrix} \Gamma_1(t) \\ \Gamma_2(t) \end{bmatrix},$$

where the drift coefficient $D^{(1)} = (D^{(1)}_1, D^{(1)}_2)^T$ is a two-dimensional vector and the diffusion coefficient is a $2 \times 2$ matrix given by $D^{(2)} = gg^T$, with $T$ denoting the transpose algebraic operator. Similar to the one-dimensional case the integration of Eq. 12 follows from a simple Euler scheme leading to

$$\begin{bmatrix} X(t + \tau) \\ Y(t + \tau) \end{bmatrix} = \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} + \tau \begin{bmatrix} D^{(1)}_1(X, Y) \\ D^{(1)}_2(X, Y) \end{bmatrix} \begin{bmatrix} \Gamma_1(t) \\ \Gamma_2(t) \end{bmatrix} + \sqrt{\tau} \begin{bmatrix} g_{11}(X, Y) & g_{12}(X, Y) \\ g_{21}(X, Y) & g_{22}(X, Y) \end{bmatrix} \begin{bmatrix} \eta_1(t) \\ \eta_2(t) \end{bmatrix}.$$  \hspace{1cm} (13)

The trajectory data can then be analyzed with 2D MSMs [40] to construct the two-dimensional Markov transition probability matrix, $M(\tau)$, at lagtime $\tau$ and
the respective conditional moments \( c^{(n)} = c^{(l,m)} \) with \( n = l + m \) the order of the KM coefficient [35],

\[
c^{(l,m)}(x_i, y_i, \tau) = \sum_{j=1}^{N_{\text{bin}}} (x_j - x_i)^l (y_j - y_i)^m M_{ji}^{(\tau)}
\]

yielding the expression for the drift (\( n=1 \)) and diffusion (\( n=2 \)) coefficients, which now depend on the initial bin \( i \) through the values of the two discretised RCs, \( x_i \) and \( y_i \). Extracting the drift and diffusion coefficients from the KM coefficients of the MSMs depends on two parameters. The first parameter is the number of bins, \( N_{\text{bin}} \), dividing the RC, at which \( D^{(1)} \) and \( D^{(2)} \) are estimated. This integer should not be too large that each bin no longer includes sufficient statistics of the transition counts and also not too small that each bin no longer includes sufficient statistics and \( y \) bin \( i \) \((n=2)\) coefficients, which now depend on the initial yielding the expression for the drift \((n=1)\) and diffusion \((n=2)\) values in Eq. 4. The conditional moments in Eq. 8 are computed for each bin and for each lagtime. For each bin, a linear fit is computed for all lagtimes in \( L_\tau \), used to build the Markov transition matrix \( M^{(\tau)} \) and calculate the KM coefficients for different \( \tau \) values in Eq. 4. The conditional moments in Eq. 8 are computed for each bin and for each lagtime. For each bin, a linear fit is computed for all lagtimes in \( L_\tau \). In practice, we have to carefully consider the range of lagtimes, as too short lagtimes lead to non-Markovian effects, where the approximations for MSMs can break down. To address the choice of lagtimes, we can also consider a more extensive phase space in higher dimensions, with more finely discretized MSMs, where non-Markovian effects are reduced.

The construction of MSMs from biased simulation data has not been traditionally possible. Biased simulations modify the potential energy function of the system of interest such that the system is, for example, harmonically restrained to a given region of the energy landscape. As this approach allows sampling of regions which might otherwise not be adequately visited during the simulation time, the kinetic behavior observed is no longer representative of the true system. Several recent numerical methods have been proposed in the literature to estimate unbiased MSMs from biased (e.g., umbrella sampling or replica exchange) MD simulations [40, 41, 60–62]. We used here the dynamic histogram analysis method (DHAM) [40], which uses a maximum likelihood estimate of the MSM transition probabilities \( M^{(\tau)}_{ji} \) given the observed transition counts during each biased trajectory. The unbiased estimate \( M^{(\tau)}_{ji} \) can then be inserted into Eq. 8 within the KM framework, which yields the estimation of the drift and diffusion coefficients.

III. Results

In the following we give several illustrative examples of the approach presented in the Methods section. We used analytical models both in one and two dimensions within the Brownian overdamped or full inertial Langevin equations. We also used explicit solvent simulations, including the water-mediated conformations of alanine dipeptide and the permeation of the Domperidone drug molecule across a lipid membrane.

A. 1D Brownian overdamped Langevin equation

As a first example we integrated the 1D Brownian overdamped Langevin equation

\[
\frac{dx(t)}{dt} = \frac{F(x(t))}{\gamma(x(t))} + \sqrt{\frac{k_B T}{\gamma(x(t))}} \eta(t),
\]

with \( k_B \) the Boltzmann constant and \( T = 300 \) K the temperature of the system. In Eq. 15, the length, time, and energy of the system are made dimensionless with the mass, characteristic length and \( k_B T \) conveniently set to unity. \( F(x) = -\nabla V(x) \) is the deterministic force derived from the 1D potential energy \( V(x) = V_{\text{rel}}(x) + V_{\text{bias}}(x) \). In the following, \( V_{\text{rel}} = \sum_{n=1}^{6} \alpha(n)x^n \) is defined as a polynomial of degree 6. We considered two different choices of the coefficients \( \alpha \), as detailed in the supporting information (SI), leading to the high \( (V_{\text{high}}) \) and low \( (V_{\text{low}}) \) barrier potentials plotted in Fig. 1(a) and (b), respectively. The biased potential is defined as \( V_{\text{bias}} = \frac{1}{2}K(x - x^{(k)})^2 \) with \( K \) the biasing spring constant and \( x^{(k)} \) the center of the harmonic bias in simulation \( k \). The parameter \( \gamma(x) \) represents the position-dependent friction coefficient with \( D(x) = k_B T/\gamma(x) \) the natural generalization of Einstein’s relation defining the position-dependent diffusion coefficient [50]. We used the Itô convention to obtain the first order integrator of the overdamped Langevin equation [45, 47]. Comparing Eq. 1 with
Eq. 15 offers a relationship to determine the potential of mean force from $D^{(1)}(X)$ and $D^{(2)}(X)$. This gives us the opportunity to self-validate our $D^{(1)}(X)$ and $D^{(2)}(X)$ values by comparing the obtained potential of mean force with the equilibrium populations obtained directly from the MSM. The total number of timesteps $N_{\text{step}}$ in the unbiased or biased simulation runs was chosen such that the simulation time $\Delta t \times N_{\text{step}} = 5 \times 10^4$ was kept constant. The simulation timestep was set to $\Delta t = 10^{-5}$, which is at least an order of magnitude smaller than the slow characteristic time scale for the diffusion in the system, $\tau_{\text{diff}} = 1/\max_x (\gamma(x))$.

Unbiased simulation. We studied the evolution of the system under Eq. 15 with the low barrier analytical potential ($V_{\text{low}}$) shown in Fig. 1(a) and $V_{\text{bias}} = 0$. We considered either a quadratic or step-like position-dependent diffusion profile with $\gamma(x)$ a parabolic (P) or a $Z$-shaped membership (Z) function [63],

$$\gamma^P(x) = \gamma_0^P \left(1 - \frac{1}{3}(x - x_P)^2\right), \quad \gamma^Z(x) = \gamma_0^Z \left(2 + \text{zmf}(x, a, b)\right),$$

where $\gamma_0^P = 3000$, $x_P = 0.8$, $\gamma_0^Z = 1700$, $a = 0.5$, $b = 1.1$, and $\text{zmf}(x, a, b)$ the sigmoidal membership function (see details in the SI). We first assessed the Markov timescale $\tau_M$ associated with the trajectories generated via Eq. 15 for different binning $N_{\text{bin}} = 100, 200, 400$, via the Chapman-Kolmogorov test. Then, we inferred the relaxation time of the system, in the limit of very long lag-times, using the fitting procedure introduced in previous work [58], which describes the relaxation time, $\tau_{\text{relax}}$, as

$$\tau_{\text{relax}} = \frac{\tau \times \mu_{\text{relax}}}{\tau + \epsilon \mu_{\text{relax}}}. \quad (18)$$

In Eq. 18, $\tau$ is the lagtime, and $\mu_{\text{relax}}$ and $\epsilon$ are two free parameters which represent the true (limiting) relaxation time.
timescale and the initial rate of change of the effective relaxation time, respectively. As shown in Fig. 1(c), increasing moderately the number of bins from 100 to 400 increases the rate of convergence towards the true relaxation time, \( \mu_{\text{relax}} = 5.8 \times 10^5 \), obtained from Eq. 18 with \( N_{\text{bin}} = 400 \). We obtained similar values for the true relaxation times measured with \( N_{\text{bin}} = 200 \) and 100 (data not shown). This yields a smaller Markov timescale, \( \tau_M \), which, in turn, gives a more accurate measure of the drift and diffusion coefficients, in line with the variational principle satisfied by the MSMs [64]. Increasing further \( N_{\text{bin}} \), however, each bin would eventually no longer include the requisite sufficient statistics for the MSM analysis. Considering \( N_{\text{bin}} = 400 \) in Fig. 1(c), the relaxation time can be seen to level off in the region of lagtimes greater than 0.2. In the analysis that follows, we chose to define \( \tau_M \approx 0.4 \), as it is sufficiently large to be insensitive to the precise choice of the lagtime. Subsequently, the first and second conditional moments were evaluated using Eq. 4 over the range of lagtime \( \tau \geq \tau_M \).

In Fig. 2(a) and (b), the representative evolution of \( \epsilon^{(2)}(X, \tau) \) is shown for different positions along the RC and different number of bins, respectively. At relatively long lagtime, for \( \tau \geq \tau_M \), \( \epsilon^{(2)} \) follows the linear trend expected from Eq. 11. Close analysis of the evolution of \( \epsilon^{(2)} \) shows that the linear trend is satisfied above a lagtime threshold, \( \tau_L \), significantly lower than \( \tau_M \). At sufficiently short lagtime, on the other hand, for \( \tau < \tau_L \), \( \epsilon^{(2)} \) deviates from the linear trend, which stems from non-Markovian effects coming from the system discretization. As shown in Fig. 2(b), increasing the number of bins from \( N_{\text{bin}} = 400 \) to 1200 decreases the value of the lagtime threshold, \( \tau_L \), above which the linear trend obtained in Eq. 11 is satisfied.

Drift and diffusion coefficients were subsequently assessed from the slope of a weighted polynomial regression following Eqs. 10 and 11. We also calculated the fourth-order coefficient and evaluated the ratio \( D^{(4)}(X)/(D^{(2)}(X))^2 < 5 \times 10^{-3} \), indicating the validity of the condition of the Pawula theorem. As shown in Fig. 3(a)-(d), we observed excellent agreement between the theoretical profiles and the numerical results for both position-dependent drift and diffusion coefficients, with higher variability around \( X \approx 0.5 \) and \( \approx 1.2 \), where the derivative of the low barrier potential \( V_{\text{low}} \), shown in Fig. 1(a), is maximal.

**Biased simulation.** We extended the previous analysis to the biased evolution of the system in the low (\( V_{\text{low}} \)) and high (\( V_{\text{high}} \)) barrier potentials shown in Figs. 1(a) and (b), respectively, within the US framework. We ran standard US simulations with the biasing potential \( V_{\text{bias}} \) defined above in each umbrella window. We used 50 uniformly distributed umbrella windows in the range \([0.25, 1.35] \) along the RC. This number was sufficiently high to obtain accurate sampling, given the different biasing spring constants considered in this work.

We first studied the effect of the bias on the reconstruction of the diffusion coefficient when the system evolves in \( V_{\text{low}} \). We set the biasing spring constant to \( K = 800 \, k_B T \), which was sufficiently strong to allow good sampling in both low (\( V_{\text{low}} \)) and high (\( V_{\text{high}} \)) barrier potentials. The lagtime dependence of \( \tau_{\text{relax}} \) is shown in Fig. 1(d). We observed a similar behaviour to the one obtained in the unbiased simulations (Fig. 1(c)) with the decline of the non-Markovian effects on the dynamics when the number of bins, \( N_{\text{bin}} \), increases from 100 to 400. Most noticeably, the use of the biasing spring constant decreases significantly the variability of the relaxation timescale at longer lagtimes. As shown in Figs. 3(e) and (f), excellent agreement is observed between the theoretical profiles and the numerical results. We also confirmed that \( D^{(4)}(X)/(D^{(2)}(X))^2 < 5 \times 10^{-3} \). As expected, the use of the biasing spring constant yields better sampling across the RC.

We complemented this analysis with the reconstruction of the diffusion coefficient of the system using \( V_{\text{high}} \) and a biasing spring constant \( K = 800 \, k_B T \). The relaxation timescale is shown in Fig. 1(e). We observed a similar behavior to the unbiased simulations with the decline of the non-Markovian effects on the dynamics when the number of bins, \( N_{\text{bin}} \), increases from 100 to 400. The relaxation time of the system calculated for \( N_{\text{bin}} = 400 \) converges towards a limiting value, \( \mu_{\text{relax}} = 1.7 \times 10^{10} \), significantly higher than the one measured for \( V_{\text{low}} \), due to the longer equilibration time needed to cross the high energy barrier. As shown in Figs. 4(a) and (b), we observed good agreement between the theoretical profiles and the numerical results for the drift and diffusion coefficients. We also verified that \( D^{(4)}(X)/(D^{(2)}(X))^2 < 5 \times 10^{-3} \). We noticed, however, higher variability around \( X \approx 0.5 \) and \( \approx 1.2 \), where the derivative of the high barrier potential \( V_{\text{high}} \) (Fig. 1(b)) is maximal. As shown in the SI, the increase of the biasing spring constant from 800 \( k_B T \) to
3000 \( k_B T \) yields better sampling around the transition states with lower variability on the reconstruction of the diffusion profile, as already observed in the low barrier case.

B. 1D full inertial Langevin equation

To take into account the role played by inertia in the reconstruction of the drift and diffusion coefficients, we extended Eq. 15 to the full inertial Langevin equation

\[
\frac{d^2 x(t)}{dt^2} = F(x(t)) - \gamma(x(t)) \frac{dx(t)}{dt} + \sqrt{k_B T \gamma(x(t)) \eta(t)}.
\]

In Eq. 19, the length, time, and energy of the system are made dimensionless with the mass, characteristic length and \( k_B T \) conveniently set to unity. From this relation, we used the Vanden-Eijnden and Ciccotti algorithm [65] that generalises the Velocity Verlet integrator to Langevin dynamics, along with the Itô convention. This scheme takes the inertial term into account and is accurate to order \( \Delta^2 \).

We used the same simulation timestep, \( \Delta t = 10^{-5} \), and biasing spring constants, \( K = 800 \ k_B T \), as in the Brownian overdamped Langevin dynamics. The lagtime dependence of the relaxation time \( \tau_{\text{relax}} \), as shown in Fig. 5(a) and (b) for \( V_{\text{low}} \) and \( V_{\text{high}} \), respectively, is similar to the one measured in the Brownian overdamped Langevin simulation, with the decline of the non-Markovian effects when the number of bins, \( N_{\text{bin}} \), increases from 100 to 400, and the convergence towards the limiting relaxation times \( \mu_{\text{relax}} = 4.4 \times 10^9 \) and \( 1.3 \times 10^{10} \) for \( V_{\text{low}} \) and \( V_{\text{high}} \), respectively, calculated for \( N_{\text{bin}} = 400 \).

We limited the rest of the analysis to the biased evolution of the system in \( V_{\text{high}} \). The associated drift and diffusion profiles are shown in Fig. 5(c) and (d) (see details in the SI). We observed good agreement between theoretical and numerical profiles, with \( D^{(4)}(X)/D^{(2)}(X) < 4 \times 10^{-3} \). As already noted in the overdamped Langevin dynamics, we observed higher variability around \( X \approx 0.5 \) and \( \approx 1.2 \), where the derivative of the potential energy \( V_{\text{high}} \) is maximal. Increasing the biasing spring constant from \( K = 800 \ k_B T \) to 3000 \( k_B T \) improved the accuracy of the measure of the diffusion coefficient (Fig. S2).

C. 2D Brownian overdamped Langevin equation

We extended the 1D analysis above to the estimation of the 2D diffusion tensor of the analytical system evolving under the 2D Brownian overdamped equation [66]

\[
\gamma(t) \frac{dX}{dt} = F(X) + \sqrt{k_B T} b(t) \eta(t)
\]

where \( X \equiv (X, Y) \) is a 2D vector and \( F(X) = -\nabla V(X) \) is the deterministic force derived from the 2D potential energy \( V(X) \). In Eq. 20, the length, time, and energy of the system are made dimensionless with the mass, characteristic lengths and \( k_B T \) conveniently set to unity. The force \( b(t) \) is related to the friction tensor \( \gamma(t) \) as follows

\[
\gamma(t) = b(t)b^T(t),
\]

and \( \eta(t) = (\eta_1(t), \eta_2(t)) \) is a 2D vector of independent, identical Gaussian white noise sources, i.e. \( \langle \eta_i(t) \rangle = 0 \) and \( \langle \eta_i(t) \eta_j(t') \rangle = 2 \delta_{ij} \delta(t - t') \). From this relation, we used the Itô convention to obtain the first order integrator of Eq. 20. For economy of computational resources, we focused our analysis on a diagonal friction tensor

\[
\gamma(t) = \begin{bmatrix} \gamma_{11}(X) & 0 \\ 0 & \gamma_{22}(X) \end{bmatrix}
\]

with constant friction \( \gamma_{11}(X) = 300 \) and \( \gamma_{22}(X) = 30 \) and the 2D analytical potential \( V(X) = -3X^2 + X^4 - 3XY + Y^4 \) shown in Fig. 6(a). We modeled standard US simulations, evenly positioned along the RC, with a 1D bias potential \( V_{\text{bias}}(X) = V^{(k)}(X) = \frac{1}{2} K (X - X^{(k)})^2 \) in each umbrella window \( k \), with \( K \) the biasing spring
obtained from the analysis of the KM coefficients, we constructed a discretized 2D grid to determine the theoretical (th) values measured the deviation between the observed (ob) and the accuracy of the reconstructed diffusion coefficient, we assessed the condition of the Pawula theorem. To assess the reliability of the PACF method of Hummer (grey) [4], we used a Langevin thermostat to enforce the temperature $T = 300 \, K$. We used a Langevin thermostat to enforce the temperature $T = 300 \, K$. The relatively low energy barriers allow the system to be sampled with both unbiased and biased simulations, which yields a more accurate measure for the center of the harmonic bias in window $k$. We used 50 uniformly distributed umbrella windows in the range $[0.25, 1.35]$ along the $x$-axis. We then constructed a discretized 2D grid to determine the MSMs along the $x$- and $y$-axes.

The diagonal elements of the diffusion tensor were obtained from the analysis of the KM coefficients defined in Eq. 14 (see details in the SI). We also calculated the fourth-order coefficients and verify that the ratio $D_{11}^{(4)}(X)/(D_{11}^{(2)}(X))^2 < 5 \times 10^{-3}$ and $D_{22}^{(4)}(X)/(D_{22}^{(2)}(X))^2 < 6 \times 10^{-4}$, indicating the validity of the condition of the Pawula theorem. To assess the accuracy of the reconstructed diffusion coefficient, we measured the deviation between the observed (ob) and theoretical (th) values

$$\Delta D_{ii}(X) = \frac{|D_{ii}^{(ob)}(X) - D_{ii}^{(th)}(X)|}{D_{ii}^{(th)}(X)}, \quad (23)$$

with $D_{ii}^{(th)}(X) = k_B T / \gamma_{ii}(X)$. In Fig. 6(b) and (c), it is shown the numerical measures for $\Delta D_{11}(X)$ and $\Delta D_{22}(X)$, respectively. The observations are in good agreement with the theoretical single values $D_{11}^{th}(X) = 2 \times 10^{-3}$ and $D_{22}^{th}(X) = 2 \times 10^{-2}$. Most noticeably, the lower value of the friction parameter $\gamma_{22}(X)$ allows better sampling, which yields a more accurate measure for $D_{22}(X)$. Finally, to quantify the accuracy of the reconstruction, we measured the average diffusion coefficients

$$D_{ii}^{eff} = \frac{1}{N_{bin}} \sum_X \sum_Y D_{ii}(X, Y), \quad (24)$$

with $N_{bin}$ the number of bins used in the 2D MSM with $X = (X, Y)$. We measured $D_{11}^{th} = (1.9 \pm 0.4) \times 10^{-3}$ and $D_{22}^{th} = (2.0 \pm 0.7) \times 10^{-2}$, in good agreement with the single values $D_{11}^{th}$ and $D_{22}^{th}$, respectively.

D. Water-mediated conformations of Alanine Dipeptide in 1D

The conformational transition between the different conformers of the solvated alanine dipeptide (Ala2) has been extensively used as a case study for several theoretical and computational investigations [4, 5, 14, 25, 67–72]. We studied the transition between the metastable states $\alpha$ and $\beta$ of Ala2 shown in Fig. 7(a), which can be differentiated by the values of the backbone dihedral angle $\Psi$ and are separated by an activation free energy barrier of $\approx 2 \, kcal/mol$ at the temperature $T = 300 \, K$. We used a Langevin thermostat to enforce the temperature $T = 300 \, K$. We used a Langevin thermostat to enforce the temperature $T = 300 \, K$. The transitions are all separated by an activation free energy barrier of $\approx 2 \, kcal/mol$ at the temperature $T = 300 \, K$. We used a Langevin thermostat to enforce the temperature $T = 300 \, K$. The LINear Constraint Solver (LINCS) algorithm [77] handled bond constraints while the particle-mesh Ewald scheme [78] was used to treat long-range electrostatic interactions. The non-bonded van der Waals cutoff radius was 0.8 nm.

The relatively low energy barriers allow the system to be sampled with both unbiased and biased simulations,
which gives a mean to assess the accuracy of the results obtained with enhanced sampling methods. We determined the free energy profile and diffusion coefficient by using either free simulations (500 ns) or US biased simulations and the DHAM approach. The starting structures for the US simulations were obtained by pulling the system along the dihedral angle $\Psi$ within the range $[-\pi, \pi]$. During the simulations, a snapshot was saved every 0.1 rad generating 60 windows. Each US window was subsequently run for 1 ns to allow equilibration, followed by additional 5 ns of the production run using an US force constant of 5 kcal mol$^{-1}$ rad$^{-2}$. In Fig. 7(b), (c), and (d) are shown the reconstructed FEPs and diffusion coefficients obtained within the DHAM framework, either with unbiased (red) or biased (blue) simulations. Both approaches gave similar results. In particular, we compared the diffusion profiles obtained with DHAM with the one obtained within the PACF framework (grey),

$$D(\Psi_k) = \frac{\var(\Psi)^2}{\int_0^\infty C_\Psi(t) \, dt}.$$  \hspace{1cm} (25)

In Eq. 25, $\langle \Psi \rangle_k$ is the average of the RC in the US window $k$, $\var(\Psi) = \langle \Psi^2 \rangle - \langle \Psi \rangle^2$ is its variance, and $C_\Psi(t) = \langle \delta \Psi(0) \delta \Psi(t) \rangle$ the PACF calculated directly from the time series. Following the original work of Hummer, we increased the strength of the biasing potential from $\approx 5$ kcal mol$^{-1}$ rad$^{-2}$ to $\approx 24$ kcal mol$^{-1}$ rad$^{-2}$ to satisfy the underlying assumption that the harmonic restraint be sufficiently large to render the underlying free energy surface a small perturbation on the harmonic potential $[4]$. As shown in Fig. 7(c) and (d), we observed good agreement between the results derived with DHAM either in the unbiased or biased simulations and the PACF method. To compare the results with those in the literature, we measured the effective diffusion coefficient

$$D_{\text{eff}}^{(2)} = \frac{1}{N_{\text{bin}}} \sum_\Psi D^{(2)}(\Psi),$$  \hspace{1cm} (26)

with $N_{\text{bin}}$ the number of bins used to discretize the dihedral angle $\Psi$ in the MSM. Eq. 26 yields the effective diffusion coefficient $D_{\text{eff}}^{(2)} \approx 0.16$ rad$^2$/ps, in agreement with the result obtained by Hummer et al. $[4, 67]$ ($\approx 0.15$ rad$^2$/ps) and Ma et al. $[5]$ ($\approx 0.34$ rad$^2$/ps). Furthermore, we assessed the reliability of our method with the comparison of the force profile, $F(\Psi)$ obtained directly from DHAM with the force profile obtained from the KM expansion, combining $D^{(1)}(\Psi) = F(\Psi)/\gamma(\Psi)$ and $D^{(2)}(\Psi) = k_B T/\gamma^2(\Psi)$. As shown in Fig. S4, the two profiles show good numerical agreement.

E. Membrane permeation in 1D

In this section we applied our method to study the permeation of Domperidone, a dopamine receptor antagonist, across a POPC lipid membrane $[11, 79, 80]$, as illustrated in Fig. 8(a). The simulation data were obtained from previous studies $[11, 80]$ and further details of the simulation methods and parameters can be found there. This system shows long time scale fluctuations inherent to the presence of the bilayer structure, which is representative of the transport of solute across heterogeneous media $[80]$. As described in the original work of Dickson et al. $[80]$, the starting structures for the US simulations were obtained by placing the ligand at the center of a POPC bilayer surrounded by water (72 POPC and 60 water molecules per lipid). To generate the US windows, the drug was pulled out from the center of the system to outside the membrane, for a total of 40 Å. Configurations were saved every 1 Å, from the center $z = 0$ Å to $z = 40$ Å generating 40 windows. Each US window was run for 20 ns to allow equilibration, followed by additional 80 ns of the production run using an US force constant of 2.5 kcal mol$^{-1}$ Å$^{-2}$.

The FEP and the diffusion profiles associated with the transport of Domperidone across the POPC bilayer, as shown in Fig. 8(b) and (c), respectively, were obtained using the DHAM. As the drug reaches the water/bilayer interface, the value of the diffusion coefficient significantly increases before remaining approximately constant until the molecule reaches the adjoining area of the two lipid layers, which is representative of the hydrophobic nature of the compound. These results can be interpreted within the inhomogeneous solubility-diffusion model commonly used to provide realistic description of membrane permeation $[81]$. In this model, the permeability, $P$, is written
as
\[
\frac{1}{P} = \int_{-d/2}^{+d/2} \frac{1}{K(Z)D(Z)} \int_{-d/2}^{+d/2} \exp \left( \frac{\Delta G(Z) / k_B T}{D(Z)} \right) dZ.
\]  

where \(K(Z), D(Z),\) and \(d\) are the position-dependent partition coefficient, the solute diffusion coefficient, and the membrane thickness respectively. \(\Delta G(Z)\) is the free energy difference, which is related to the partition coefficient \(K(Z) = \exp \left( -\Delta G(Z) / k_B T \right)\). Considering the results shown in Fig. 8, we obtained the log Perm value associated with the transport of the Domperidone molecule across the POPC bilayer, \(\log P = -2.71 \pm 0.15\), which is in agreement with experimental results (\(\log P_{\text{exp}} = -2.6\)) [11, 79, 80].

\section*{F. Membrane permeation in 2D}

We extended the 1D study above to the 2D analysis of the permeation of Domperidone across the POPC lipid membrane. To do so, we considered the rotational movement of the drug during its passage across the membrane as additional degree of freedom. We constructed a discretized two-dimensional grid to determine the MSMs along two reaction coordinates (see details in the SI). As our first reaction coordinate, we used the distance from the drug center of mass to bilayer center, \(Z\), as already considered in 1D. For our second coordinate, we used the projection of the molecular orientation vector onto the axis orthogonal to the bilayer axis, \(\Delta Z\), as a measure of the orientation of the drug with respect to the membrane, as described in Badaoui et al. [11]. We restrained the exploration of the free energy landscape to the area centered around the minimal energy well (\(Z \approx 10\AA\) in Figs. 8(a) and (b)). We measured the diagonal and off-diagonal KM coefficients defined in Eqs. 14 (see details in the SI). Subsequently, the associated elements of the diffusion tensor were obtained with weighted polynomial regression.

In Fig. 9(a), (b), and (c), we calculated the 2D projection of the diagonal and off-diagonal elements of the diffusion tensor, \(D_{11}(Z, \Delta z)\), \(D_{22}(Z, \Delta z)\), and \(D_{12}(Z, \Delta z)\), associated with the crossing of Domperidone across the POPC membrane bilayer. We restrained the exploration of the free energy landscape to the area centered around the minimal energy well (\(Z \approx 10\AA\) in Figs. 8(a) and (b)).

\section*{IV. Conclusion}

In the present study, we presented a general method for estimating multi-dimensional position-dependent diffusion coefficients equally valid in biased and unbiased MD simulations. For the first time, we combined the dynamic histogram analysis method (DHAM) and Kramers-Moyal expansion to link the underlying stochastic process obtained within the non-linear Langevin framework to its probabilistic description. Our approach neither requires prior assumptions regarding the form of the free energy landscape nor additional numerical integration scheme. We applied our numerical approach to one- and two-dimensional analytic potentials and data from explicit solvent molecular dynamics simulations, including the water-mediated conformations of the alanine dipeptide.

Importantly, we demonstrated the efficiency of our algorithm in studying the transport of solute across three-dimensional heterogeneous porous media, which is
known to show long time scale fluctuations potentially breaking ergodicity in sampling [9]. Specifically, our algorithm provided accurate assessment of the diffusion coefficient associated with the crossing of Domperidone across a lipid POPC membrane, which was not previously accessible with standard methods [82].

The method would allow the measure of the off-diagonal elements of the diffusion tensor, which is essential for assessing the importance of dynamic coupling between the reaction coordinates. As discussed by Ma and coworkers [5], the presence of significant off-diagonal element in the diffusion tensor could indeed impact the determination of the reactive eigenvector at play in the Kramers’ theory of reaction kinetics and then the measure of the rate associated with the kinetic evolution of the system. This is particularly important in toxicology and pharmacology, where the prediction of kinetic quantities of solutes, including drugs and small peptides provides fundamental understanding of numerous biochemical transport processes involved in the design of new drugs [82].

Additionally, using the calculated $D^{(1)}(X)$ and $D^{(2)}(X)$ profiles, we can also determine the force $F(X)$, which can be directly compared with the gradient of $V(x)$ from the MSM, offering a self-validation of the fitted KM coefficients. Alternatively, considering the potential of mean force, $V(x)$ from DHAM, which is typically more accurate and consistent over the range of lagtimes considered for the KM fitting, we could now reinforce this MSM-derived force exactly, and fit only $D^{(2)}(X)/D^{(1)}(X)$. In future developments, this relationship can therefore also be integrated with the determination of the KM coefficients to decrease the number of degrees of freedom that require numerical fit.

Eventually, one could consider the dynamic histogram analysis method extended to detailed balance (DHAMed) [41], where the transition rates are used as parameters to build MSMs from biased trajectories sampled at finite observation intervals. However, this would require to systematically eliminate transitions between non-contiguous states [41] in order to obtain a rate matrix consistent with a diffusive process, where no jump is allowed beyond the immediate neighbors of a state [4, 83, 84]. This roadmap will be considered in future work.

V. Acknowlegments

We acknowledge use of the research computing facility at King’s College London, Rosalind (https://www.rosalind.kcl.ac.uk). We thanks Robert Best and Gerhard Hummer for fruitful discussions and Magd Badaoui for his help in setting up the membrane bilayer system. We acknowledge the support of the UK Engineering and Physical Sciences Research Council (EPSRC), under grant number EP/R013012/1, and ERC project 757850 BioNet.

VI. Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI:

- Numerical details of the analytical potentials and position-dependent frictions considered in the Langevin dynamics.
- Numerical details for the MSM analysis of the Water-mediated conformations of Alanine Dipeptide and the membrane permeation.

[1] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge, Turbulent cascades in foreign exchange markets, Nature 381, 767 (1996).
[2] A. Naert, R. Friedrich, and J. Peinke, Fokker-planck equation for the energy cascade in turbulence, Phys. Rev. E 56, 6719 (1997).
[3] S. Lück, J. Peinke, and R. Friedrich, Uniform statistical description of the transition between near and far field turbulence in a wake flow, Phys. Rev. Lett. 83, 5495 (1999).
[4] G. Hummer, Position-dependent diffusion coefficients and free energies from bayesian analysis of equilibrium and replica molecular dynamics simulations, New J. Phys. 7, 34 (2005).
[5] A. Ma, A. Nag, and A. Dinner, Dynamic coupling between coordinates in a model for biomolecular isomerization, J. Chem. Phys. 124, 144911 (2006).
[6] A. Bahraminasab, F. Ghasemi, A. Stefanovska, P. McClintock, and R. Friedrich, Physics of brain dynamics: Fokker-planck analysis reveals changes in eeg $\delta - \theta$ interactions in anaesthesia, New J. Phys. 11, 103051 (2009).
[7] R. Best and G. Hummer, Coordinate-dependent diffusion in protein folding, Proc. Natl. Acad. Sci. USA 107, 1088 (2010).
[8] J. Comer, C. Chipot, and F. Gonzalez-Nilo, Calculating position-dependent diffusivity in biased molecular dynamics simulations, J. Chem. Theory Comput. 9, 876 (2013).
[9] C. Lee, J. Comer, C. Herndon, N. Leung, A. Pavlova, R. Swift, C. Tung, C. Rowley, R. Amaro, C. Chipot, Y. Wang, and J. Gumbart, Simulation-based approaches for determining membrane permeability of small compounds, J. Chem. Inf. Model. 56, 721 (2016).
[10] K. Gaalswyk, E. Awoonor-Williams, and C. Rowley, Generalized langevin methods for calculating transmembrane diffusivity, J. Chem. Theory Comput. 12, 5609 (2016).
[11] M. Badaoui, A. Kells, C. Molteni, C. Dickson, V. Hornak, and E. Rosta, Calculating kinetic rates and membrane
permeability from biased simulations, J. Phys. Chem. B 122, 11571 (2018).
[12] H. Kramers, Brownian motion in a field of force and the diffusion model of chemical reactions, Physica 7, 284 (1940).
[13] P. Hanggi, P. Talkner, and M. Borkovec, Reaction-rate theory: fifty years after kramers, Rev. Mod. Phys. 62, 251 (1990).
[14] F. Sicard, Computing transition rates for rare event: When kramers theory meets the free-energy landscape, Phys. Rev. E 98, 052408 (2018).
[15] E. Rosta, H. Woodcock, B. Brooks, and G. Hummer, Artificial reaction coordinate “tunneling” in free-energy calculations: The catalytic reaction of rnase h, J. Comput. Chem. 30, 1634 (2009).
[16] E. Rosta, M. Nowotny, W. Yang, and G. Hummer, Catalytic mechanism of rna backbone cleavage by ribonuclease h from quantum mechanics/molecular mechanics simulations, J. Am. Chem. Soc. 133, 8934–8941 (2011).
[17] F. Sicard and P. Senet, Reconstructing the free-energy landscape of met-enkephalin using dihedral principal component analysis and well-tempered metadynamics, J. Chem. Phys. 138, 235101 (2013).
[18] F. Sicard, N. Destainville, and M. Manghi, Dna denaturation bubbles: Free-energy landscape and nucleation/closure rates, J. Chem. Phys. 142, 034903 (2015).
[19] F. Twary and A. van de Walle, A Review of Enhanced Sampling Approaches for Accelerated Molecular Dynamics (Springer, Cham. Switzerland, 2016).
[20] F. Sicard, T. Bui, D. Monteiro, Q. Lan, M. Ceglio, C. Burress, and A. Striolo, Emergent properties of antiagglomerant films control methane transport: Implications for hydrate management, Langmuir 34, 9701 (2018).
[21] F. Sicard, J. Toro-Mendoza, and A. Striolo, Nanoparticles actively fragment armored droplets, ACS Nano 13, 9498 (2019).
[22] C. Camilloni and F. Pietrucci, Advanced simulation techniques for the thermodynamic and kinetic characterization of biological systems, Adv. Phys. X 3, 1477531 (2018).
[23] F. Sicard, N. Destainville, P. Rousseau, C. Tardin, and M. Manghi, Dynamical control of denaturation bubble nucleation in supercooled dna minicircles, Phys. Rev. E 101, 012403 (2020).
[24] A. Einstein, Investigations on the Theory of the Brownian Movement (Dover Books on Physics Series; Dover Publications, 1956).
[25] A. Ljubetić, I. Urbancič, and J. Štrancar, Recovering position-dependent diffusion from biased molecular dynamics simulations, J. Chem. Phys. 140, 084109 (2014).
[26] J. Bußerjahn, S. von Bulow, and G. Hummer, Optimal estimates of self-diffusion coefficients from molecular dynamics simulations, J. Chem. Phys. 153, 024116 (2020).
[27] R. Kubo, The fluctuation-dissipation theorem, Rep. Prog. Phys. 29, 255 (1966).
[28] A. Mamonov, M. Kurnikova, and R. Coalson, Diffusion constant of k+ inside granicidin a: A comparative study of four computational methods, Biophys. Chem. 124, 268 (2006).
[29] S.-J. Marrink and H. Berendsen, Simulation of water transport through a lipid membrane, J. Phys. Chem. 98, 4155 (1994).
[30] S.-J. Marrink and H. Berendsen, Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations, J. Phys. Chem. 100, 16729 (1996).
[31] A. Ghysele, R. Venable, R. Pastor, and G. Hummer, Position-dependent diffusion tensors in anisotropic media from simulation: Oxygen transport in and through membranes, J. Chem. Theory Comput. 13, 2962 (2017).
[32] B. Berne, M. Borkovec, and J. Straub, Classical and modern methods in reaction rate theory, J. Phys. Chem 92, 3711 (1988).
[33] T. Wolf and B. Roux, Conformational flexibility of o-phosphorylcholine and o-phosphorylethanolamine: A molecular dynamics study of solvation effects, J. Am. Chem. Soc. 116, 5916 (1994).
[34] S. Türkcan, A. Alexandrou, and J. Masson, A bayesian inference scheme to extract diffusivity and potential fields from confined single-molecule trajectories, Biophys. J. 102, 2288 (2012).
[35] H. Risken, The Fokker-Planck Equation (Springer, Berlin, 1966).
[36] C. Anteneodo and R. Riera, Arbitrary-order corrections for finite-time drift and diffusion coefficients, Phys. Rev. E 80, 031103 (2009).
[37] M. Petelczyc, J. Zebrowski, and R. Baranowski, Kramers-moyal coefficients in the analysis and modeling of heart rate variability, Phys. Rev. E 80, 031127 (2009).
[38] P. Rinat, P. Lind, M. Wachtler, and J. Peinke, The langevin approach: An r package for modeling markov processes, J. Open Research Soft. 4, e34 (2016).
[39] L. R. Gorgiao and F. Meirinhos, Kramers-moyal: Kramers-moyal coefficients for stochastic processes, J. Open Source Soft. 4, 1693 (2019).
[40] E. Rosta and G. Hummer, Free energies from dynamic weighted histogram analysis using unbiased markov state model, J. Chem. Theory Comput. 11, 276 (2015).
[41] L. Stelzl, A. Kells, E. Rosta, and G. Hummer, Dynamic histogram analysis to determine free energies and rates from biased simulations, J. Chem. Theory Comput. 13, 6328 (2017).
[42] S. Olsson, H. Wu, F. Paul, C. Clementi, and F. Noé, Combining experimental and simulation data of molecular processes via augmented markov models, Proc. Natl.Acad. Sci. USA 114, 8265 (2017).
[43] B. Husic and V. Pande, Markov state models: From an art to a science, J. Am. Chem. Soc. 140, 2386 (2018).
[44] M. Biswas, B. Lickert, and G. Stock, Metadynamics enhanced markov modeling of protein dynamics, J. Phys. Chem. B 122, 5508 (2018).
[45] O. Farago and N. Gronbech-Jensen, langevin dynamics in inhomogeneous media: Re-examining the itō-stratonovich dilemma, Phys. Rev. E 89, 031301 (2014).
[46] A. Bodrova, A. Chechkin, A. Cherstvy, H. Safdari, I. Sokolov, and R. Metzler, Underdamped scaled brownian motion: (non-)existence of the overdamped limit in anomalous diffusion, Sci. Rep. 6, 30520 (2016).
[47] S. Regev, N. Gronbech-Jensen, and O. Farago, Isothermal langevin dynamics in systems with power-law spatially dependent friction, Phys. Rev. E 94, 012116 (2016).
[48] A. Lau and T. Lubensky, State-dependent diffusion: Thermodynamic consistency and its path integral formulation, Phys.Rev.E 76, 011123 (2007).
[49] G.Bussi and M. Parrinello, Accurate sampling using langevin dynamics, Phys.Rev.E 75, 056707 (2007).
[50] O. Farago and N. Gronbech-Jensen, Fluctua-
tion–dissipation relation for systems with spatially varying friction, J. Stat. Phys. 156, 1093–1110 (2014).
[51] R. Friedrich and J. Peinke, Description of a turbulent cascade by a fokker-planck equation, Phys. Rev. Lett. 78, 863 (1997).
[52] S. Siegert, R. Friedrich, and J. Peinke, Analysis of data sets of stochastic systems, Phys. Lett. A 243, 275 (1998).
[53] R. Friedrich, J. Peinke, M. Sahimi, and M. Tabar, Approaching complexity by stochastic methods: From biological systems to turbulence, Phys. Rep. 506, 87 (2011).
[54] C. Honisch and R. Friedrich, Estimation of kramers-moyal coefficients at low sampling rates, Phys. Rev. E 84, 019906 (2011).
[55] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, V. Pande, K. Beauchamp, and G. Bowman, Everything you wanted to know about markov state models but were afraid to ask, Methods 52, 99 (2010).
[56] J.-H. Prinz, J. Chodera, and F. Noé, Markov models of molecular kinetics: Generation and validation, J. Chem. Phys. 134, 174105 (2011).
[57] J.-H. Prinz, J. Chodera, and F. Noé, Estimation and Validation of Markov Models, in An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation (Springer, New York, 2014).
[58] A. Kells, A. Annibale, and E. Rosta, Limiting relaxation times from markov state models, J. Chem. Phys. 149, 072324 (2018).
[59] G. R. Bowman, V. S. Pande, and F. Noé, An introduction to markov state models and their application to long timescale molecular simulation, in Advances in Experimental Medicine and Biology (Springer, Dordrecht, 2014).
[60] H. Wu, F. Paul, C. Wehmeyer, and F. Noé, Multiensemble markov models of molecular thermodynamics and kinetics, Proc. Natl. Acad. Sci. USA 113, E3221 (2016).
[61] L. Donati, C. Hartmann, and B. Keller, Girsanov reweighting for path ensembles and markov state models, J. Chem. Phys. 146, 244112 (2017).
[62] S. Kienlinger, L. Donati, and B. Keller, Dynamical reweighting methods for markov models, Curr. Opin. Struc. Bio. 61, 124 (2020).
[63] S. Medasani, J. Kim, and R. Krishnapuram, An overview of membership function generation techniques for pattern recognition, Int. J. Approx. Reason. 19, 391 (1998).
[64] A. Kells, Z. Miháka, A. Annibale, and E. Rosta, Mean first passage times in variational coarse graining using markov state models, J. Chem. Phys. 150, 134107 (2019).
[65] A. Kells, Z. Miháka, A. Annibale, and E. Rosta, Second-order integrators for langevin equations with holonomic constraints, Chem. Phys. Lett. 429, 310 (2006).
[66] J. Freund and T. Poschel, Stochastic Processes in Physics, Chemistry, and Biology (Springer, Berlin, Heidelberg, 2000).
[67] G. Hummer and I. Kevrekidis, Coarse molecular dynamics of a peptide fragment: Free energy, kinetics, and long-time dynamics computations, J. Chem. Phys. 118, 10762 (2003).
[68] W. Ren, E. Vanden-Eijnden, P. Maragakis, and E. Weinan, Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide, J. Chem. Phys. 123, 134109 (2005).
[69] J. Vymetal and J. Vondrasek, Metadynamics as a tool for mapping the conformational and free-energy space of peptides - the alanine dipeptide case study, J. Phys. Chem. B 114, 5632–5642 (2010).
[70] Y. Yonezawa, I. Fukuda, N. Kamiya, H. Shimoyama, and H. Nakamura, Free energy landscapes of alanine dipeptide in explicit water reproduced by the force-switching wolf method, J. Chem. Theory Comput. 7, 1484–1493 (2011).
[71] P. Tiwary and M. Parrinello, From metadynamics to dynamics, Phys. Rev. Lett. 111, 230602 (2013).
[72] J. Cuny, K. Korchagina, C. Menakbi, and T. Mineva, Metadynamics combined with auxiliary density functional and density functional tight-binding methods: Alanine dipeptide as a case study, J. Mol. Model 23, 72 (2017).
[73] G. Bussi, D. Donadio, and M. Parrinello, Canonical sampling through velocity rescaling, J. Chem. Phys. 126, 014101 (2007).
[74] D. Case, T. Cheatham, T. Darden, H. Gohlke, R. Luo, K. J. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. Woods, The amber biomolecular simulation programs, J. Comp. Chem. 26, 1668 (2005).
[75] W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, and M. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79, 926935 (1983).
[76] E. Lindahl, B. Hess, and D. V. D. Spoel, Gromacs 3.0: a package for molecular simulation and trajectory analysis, J. Mol. Model. 7, 306 (2001).
[77] B. Hess, H. Bekker, H. Berendsen, and J. Fraaije, Lincs: A linear constraint solver for molecular simulations, J. Comput. Chem. 18, 1463 (1997).
[78] T. Darden, D. York, and L. Pedersen, Particle mesh ewald: An nlog(n) method for ewald sums in large systems, J. Chem. Phys. 98, 10089 (1993).
[79] K. Eyer, F. Paech, F. Schuler, P. Kuhn, R. Kissner, S. Belli, P. Dittrich, and S. Krämer, A liposomal fluorescence assay to study permeation kinetics of drug-like weak bases across the lipid bilayer, J. Control. Release 173, 102 (2014).
[80] C. Dickson, V. Hornak, R. Pearlstein, and J. Duca, Structure-kinetic relationships of passive membrane permeation from multiscale modeling, J. Am. Chem. Soc. 139, 442 (2017).
[81] W. Shinoda, Permeability across lipid membranes, Biochim. Biophys. Acta 1858, 2254 (2016).
[82] R. Amaro and A. Mulholland, Multiscale methods in drug design bridge chemical and biological complexity in the search for cures, Nat. Rev. Chem. 2, 0148 (2018).
[83] D. Bicout and A. Szabo, Electron transfer reaction dynamics in non-debye solvents, J. Chem. Phys. 109, 2325 (1998).
[84] M. Innerbichler, G. Menzl, and C. Dellago, State-dependent diffusion coefficients and free energies for nucleation processes from bayesian trajectory analysis, Molecular Physics 116, 2987 (2018).
FIG. 10. TOC/Graphical Abstract