JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

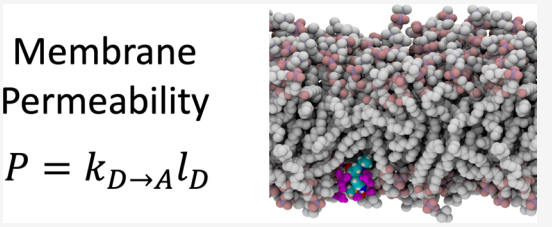
Mechanistic Insights into Passive Membrane Permeability of Drug-like Molecules from a Weighted Ensemble of Trajectories

She Zhang, Jeff P. Thompson, Junchao Xia, Anthony T. Bogetti, Forrest York, A. Geoffrey Skillman, Lillian T. Chong,* and David N. LeBard*

Cite This: https://doi.org/10.1021/acs.jcim.1c01540 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Passive permeability of a drug-like molecule is a critical property assayed early in a drug discovery campaign that informs a medicinal chemist how well a compound can traverse biological membranes, such as gastrointestinal epithelial or restrictive organ barriers, so it can perform a specific therapeutic function. However, the challenge that remains is the development of a method, experimental or computational, which can both determine the permeation rate and provide mechanistic insights into the transport process to help with the rational design of any given molecule. Typically, one of the following three methods are used to measure the membrane permeability: (1) experimental permeation assays acting on either artificial or natural membranes; (2) quantitative structure−permeability relationship models that rely on experimental values of permeability or related pharmacokinetic properties of a range of molecules to infer those for new molecules; and (3) estimation of permeability from the Smoluchowski equation, where free energy and diffusion profiles along the membrane normal are taken as input from large-scale molecular dynamics simulations. While all these methods provide estimates of permeation coefficients, they provide very little information for guiding rational drug design. In this study, we employ a highly parallelizable weighted ensemble (WE) path sampling strategy, empowered by cloud computing techniques, to generate unbiased permeation pathways and permeability coefficients for a set of drug-like molecules across a neat 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphatidylcholine membrane bilayer. Our WE method predicts permeability coefficients that compare well to experimental values from an MDCK-LE cell line and PAMPA assays for a set of drug-like amines of varying size, shape, and flexibility. Our method also yields a series of continuous permeation pathways weighted and ranked by their associated probabilities. Taken together, the ensemble of reactive permeation pathways, along with the estimate of the permeability coefficient, provides a clearer picture of the microscopic underpinnings of small-molecule membrane permeation.



INTRODUCTION

The ability of a drug candidate to cross (or permeate) lipid membranes is essential for achieving the required absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox) profile.[1] Understanding the mechanism by which membrane permeation can occur is invaluable for the rational improvement of drug bioavailability. While active transport via transmembrane proteins can contribute to ADME/Tox properties[2] of charged endogenous compounds,[3,4] passive diffusion across lipid bilayers is believed to be the predominant mechanism for membrane transport of drug candidates in a variety of cell types.

Given the time-consuming and costly nature of in vitro experiments[5−8] for measuring passive membrane permeation, there has been great interest in theoretical strategies for predicting drug permeability. The first theoretical model of permeation was based on Overton's rule[9] developed over a century ago,[10] which proportionally relates passive membrane permeability to the oil−water partition coefficient.[11] In this model, permeability is correlated with the ability to partition into the lipid phase, assuming that permeation is driven only by a molecule's inherent lipophilicity. More sophisticated quantitative structure−permeability relationship (QSPR) models have also emerged that are built from experimentally derived permeability measurements, along with a set of input physiochemical descriptors from a database of training molecules.[12] QSPR models typically rely on properties such as the polar surface area, molecular weight, hydrogen bond count, and octanol−water partition coefficient to statistically infer new permeability coefficients for drug-like compounds.[13] Over the last few years, advanced machine learning (ML) techniques have also been applied to predict permeability from either molecular descriptors or fingerprints. Reviews of ML applications to predict permeability exist elsewhere[14] and will not be covered in detail here, yet some examples of ML approaches worth mentioning include models for Parallel Artificial Membrane

Received: December 21, 2021

Permeability Assay (PAMPA)-like assays,[5] models for Caco-2 cell permeability,[15] as well as blood−brain barrier[16] and CNS[17] permeability models. Although ML and QSPR methods can be fast, cheap, and accurate within the domain of chemical space of the training set, such models have not provided any mechanistic insights to guide drug design to favor permeation.

In principle, molecular dynamics (MD) simulations can provide the most detailed mechanistic insights on membrane permeation. While MD simulations have captured small-molecule permeation processes that occur within a microsecond,[18] membrane permeability of drug-like molecules can be orders of magnitude beyond this timescale. To access these longer timescales, many simulation studies have used a method proposed by Marrink and Berendsen[19] that is based on an inhomogeneous solubility-diffusion (ISD) model in which the permeability coefficient of a small molecule is estimated from the free energy and diffusion rate profiles across the membrane. Such studies have used enhanced sampling techniques that apply either external biasing forces or modified Hamiltonians [e.g., adaptive biasing force method,[20] free energy perturbation,[21] restrained potential of mean force (PMF) calculations,[22] constant pH simulations,[23] and umbrella sampling.][24] A caveat of such studies, however, is that the results rely on the free energy profile, which is a function of the chosen reaction coordinate and may miss potential rate-limiting steps in orthogonal coordinates.[25,26] Alternatively, a few simulation studies have employed methods that combine short discontinuous trajectories to estimate long-timescale observables (e.g., milestoning[27,28] and Markov state models[29]). While all the above studies have revealed some microscopic details of the permeation process, they have not provided pathways of membrane permeation processes that are both unbiased in the dynamics and continuous.

To bridge the gap between computation and experiment in a quantitative manner, we combine the WE path sampling strategy with cloud-based computing to not only provide direct estimates of permeability coefficients but also continuous, atomistic permeation pathways of drug-like compounds across a model lipid bilayer. We focus initially on tacrine, a rigid small-molecule inhibitor of acetylcholinesterase (zero rotatable bonds), to evaluate the efficiency of several WE protocols for estimating permeability coefficients. We then apply the most efficient protocol to (R)-zacopride and sotalol, which are substantially more complex, flexible molecules (three and six rotatable bonds, respectively), for detailed analysis of their membrane permeation mechanisms. All three compounds are weakly basic and obey Lipinski's "rule of five"[30,31] for orally active therapeutics.

Theoretical Background. In this section, we introduce a general model to predict permeability using the kinetic rate constant of membrane crossing. The permeation rate constant can be defined using the mean-first-passage time (MFPT) of the crossing event, as will be outlined with the Hill relation below. In principle, however, this rate constant could be obtained from any kinetic method.

Kinetic Model of Membrane Permeability. The passive membrane permeation of a drug-like molecule can be modeled as a two-state, first-order process, in which there is a large free energy barrier to desolvate the molecule upon entering the lipid bilayer. During this process, a molecule diffuses across a lipid bilayer to reach one of two aqueous compartments, namely, the donor (D) or the acceptor (A)

$$\mathrm{D} \underset{k_{\mathrm{D}\to\mathrm{A}}}{\overset{k_{\mathrm{A}\to\mathrm{D}}}{\rightleftarrows}} \mathrm{A} \tag{1}$$

Here, $k_{\mathrm{D}\to\mathrm{A}}$ is the forward rate constant for passive diffusion of the molecule from the donor to the acceptor compartment, while $_{\mathrm{A}\to\mathrm{D}}$ is the reverse rate constant. Assuming that the system is symmetric (i.e., the volumes of compartment D and A are identical) and C_{D} and C_{A} are the concentrations of the molecule in the two compartments, the rate of change in the concentrations (or populations) of the molecule can be described by the following set of ordinary differential equations

$$\frac{\mathrm{d}C_{\mathrm{A}}}{\mathrm{d}t} = \nu_{\mathrm{D}\to\mathrm{A}} - \nu_{\mathrm{A}\to\mathrm{D}} \tag{2}$$

$$\frac{\mathrm{d}C_{\mathrm{D}}}{\mathrm{d}t} = \nu_{\mathrm{A}\to\mathrm{D}} - \nu_{\mathrm{D}\to\mathrm{A}} \tag{3}$$

where ν's are volumetric permeation rates

$$\nu_{\mathrm{D}\to\mathrm{A}} = k_{\mathrm{D}\to\mathrm{A}} C_{\mathrm{D}}$$

$$\nu_{\mathrm{A}\to\mathrm{D}} = k_{\mathrm{A}\to\mathrm{D}} C_{\mathrm{A}} \tag{4}$$

Alternatively, these rates can be measured by the rate of change in the concentration of the respective species. For example

$$\nu_{\mathrm{D}\to\mathrm{A}} = \frac{1}{l_{\mathrm{D}} S}\frac{\mathrm{d}m_{\mathrm{D}}}{\mathrm{d}t} \tag{5}$$

where $\mathrm{d}m_{\mathrm{D}}$ is the amount of molecule (in terms of mass or molar mass), originating from compartment D, passing through the bilayer, and moving into compartment A within time dt. Here, S is the surface area of the membrane, and l_{D} is the depth of the effective "reaction" volume of compartment D or the "unstirred layer" as referred by Missner and Pohl.[11] Instead of those in the bulk water, molecules in this aqueous layer adjacent to the membrane "react" with the membrane surface, potentially leading to the initiation of the permeation process. The detailed balance between the concentrations of the molecules in bulk water and in this layer should be safely regarded as rapidly maintained by diffusion in a relatively uniform aqueous environment—a process that is much more rapid than permeation through a lipid membrane.[18]

Although the volumetric permeation rates in eq 4 provide one route to estimate the kinetics of molecular transport across the membrane, the most common quantification is the membrane permeability coefficient, P_{m}, typically measured in logarithmic units of cm/s. This coefficient is based on Fick's first law of diffusion and linearly connects the net flux of the molecule across the membrane at steady state, J_{m}, to the difference in concentrations of the molecule in compartments D and A

$$J_{\mathrm{m}} = P_{\mathrm{m}}(C_{\mathrm{D}} - C_{\mathrm{A}}) \tag{6}$$

By imposing the homogenous solubility-diffusion (HSD) model,[18] where the membrane is in a quasi-steady state, it can be readily demonstrated that the proportionality constant from eq 6 is related to physical properties of the permeant-membrane system. If the concentration change is assumed to be linear across the membrane, the permeability coefficient equals DK/h, where D is the diffusion constant of the molecule inside the membrane, K is the oil−water partition coefficient, and h is the membrane thickness. The relationship between P_{m} and DK has been experimentally verified over 6 log units,[32] with a few notable exceptions that are likely attributed to active transport.

https://doi.org/10.1021/acs.jcim.1c01540
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

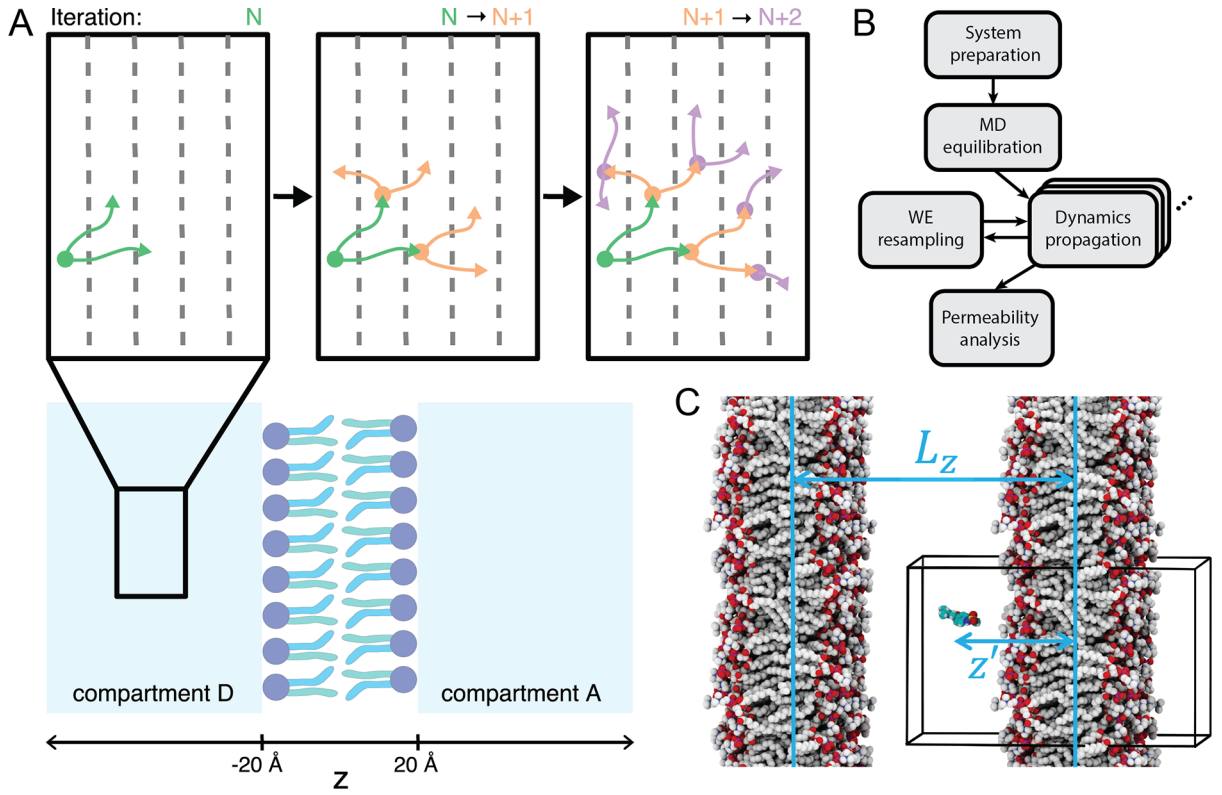


Figure 1. Basic weighted ensemble protocol and system setup. (A) Illustration of the WE protocol for a membrane permeability simulation in which a one-dimensional progress coordinate z (eq 17) is divided into bins, and iterative rounds of dynamics propagation and a resampling procedure are performed with the goal of providing even coverage along the coordinate. As seen in the upper left, two trajectories (*solid dots*) originating from the left-most bin each occupy a previously empty bin after N rounds of dynamics (*curved arrows*). The resampling procedure then replicates the trajectories in these newly occupied bins to maintain a target number of two trajectories per bin. (B) Simulation workflow used by permeability floe to directly calculate permeability coefficients, including one round of WE resampling (using WESTPA) and dynamics propagation for each WE iteration (see Figure S1 for further details). (C) Snapshot of the simulation system from a trajectory of a "rule of five" permeate, sotalol, crossing the periodic membrane. All water molecules have been removed for clarity. L_z is the z-component of the simulation box, while z' is the distance from the center of mass of the permeate and the center of mass of the membrane (*straight blue lines*).

In addition to eq 6, the membrane flux can also be written as the difference between the molecular influx, $J_{D\to A}$, and outflux, $J_{A\to D}$, through the membrane

$$J_m = J_{D\to A} - J_{A\to D} \tag{7}$$

Both the influx and outflux represent the amount of molecules that passes through a surface area, S, within a given amount of time. Mathematically, this can be expressed as

$$J_{D\to A} = \frac{1}{S}\frac{dm_D}{dt} \tag{8}$$

By comparing eqs 5 and 8, one can see that the in-/outflux and the volumetric permeation rates are related

$$J_{D\to A} = v_{D\to A} l_D$$

$$J_{A\to D} = v_{A\to D} l_A \tag{9}$$

Substituting eqs 4, 7, and 9 into eq 6 and assuming unequal concentrations in the two compartments, that is, $C_D \neq C_A$, we arrive at

$$P_m = \frac{k_{D\to A} l_D C_D - k_{A\to D} l_A C_A}{C_D - C_A} \tag{10}$$

Furthermore, either (1) under the assumption of a symmetric system, such as the case in the permeation of a small molecule through a neat membrane, where one can set $k_{D\to A} = k_{A\to D}$ and $l_D = l_A$, or (2) under the steady-state consideration of the WE strategy, where $C_A \equiv 0$ (see below for details), the form of the permeability coefficient in eq 10 can be reduced to

$$P_m = k_{D\to A} l_D = k_{A\to D} l_A \tag{11}$$

In our model, by default, we consider the size of the unstirred layer, l_D, to half the length of the aqueous part of the simulation box along the lipid bilayer normal. A more refined value of l_D can be determined by the size of the free energy basin of donor compartment D,[33] but preliminary tests suggest that the permeability coefficient is somewhat insensitive to the exact value of l_D, especially since permeability coefficients are usually expressed in log units (see Figure S2 for details). While eq 11 is derived here for a two-state permeation process, the same equation can be derived from a multistate model, where states D and A are connected by a series of intermediates inside the lipid bilayer to yield apparent rate constants for the overall process, $k_{D\to A}$ or $k_{A\to D}$, as described by Parisio et al.[33]

Calculation of Permeation Rate Constants Using Weighted Ensemble Simulations. The weighted ensemble path sampling strategy[34,35] enables simulations of processes that are orders of magnitude longer than the simulations themselves.[36−39] This greatly enhanced sampling results from an iterative resampling procedure (at fixed time intervals τ) that replicates trajectories to occupy less-visited regions of configurational space—typically defined by a progress coordinate toward

https://doi.org/10.1021/acs.jcim.1c01540
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

the target state (also known as a reaction coordinate, set of order parameters, or collective variables) that is divided into bins (Figure 1A). The WE resampling strategy is unbiased in two senses: (1) the underlying dynamics is not altered by the resampling and (2) the trajectories are assigned statistical weights that are rigorously tracked, such that the resampling procedure is statistically unbiased. The former allows us to potentially obtain an ensemble of continuous pathways for the process of interest, and the latter allows us to evaluate unbiased estimates of steady-state averages, such as the rate constants into any arbitrary state.[40,41] Thus, the WE strategy provides an ideal framework for direct simulations of drug membrane-permeability pathways and calculations of permeability coefficients.

According to the Hill relation, the permeation rate constant, or steady-state probability flux, into a target state of interest (compartment A) is exactly the inverse MFPT[35,42]

$$k_{D\to A} = \frac{1}{\mathrm{MFPT}(D \to A)} = \frac{f^{SS}_{D\to A}}{p_D} \tag{12}$$

For drug membrane permeation, $f^{SS}_{D\to A}$ is the steady-state (SS) probability flux arriving at state A from D, while p_D is the fraction of trajectories that are more recently in state D than in state A.

Within the WE framework, steady-state fluxes are mimicked by introducing an initial state, a target state, and a "recycling" condition, where a trajectory will be returned to the initial state once it enters the target state to prevent re-entry (i.e., first passage). In the context of the permeation system, the initial and the final states naturally correspond to compartments A and D. In this way, the weight of the recycled trajectory can be recorded as a probability flux into the target state. The time average of instantaneous fluxes arriving at state A from D, $\hat{f}_{D\to A}$, in a WE simulation provides an estimate of the steady-state flux

$$k_{D\to A} = \frac{\langle \hat{f}_{D\to A} \rangle}{\langle \hat{p}_D \rangle} = \langle \hat{f}_{D\to A} \rangle \tag{13}$$

The second equality comes from the recycling condition, which implies that $\hat{p}_D$, the instantaneous fraction of trajectories more recently in D than in A, is identically one. In practice, taking the time average using pre-steady-state data could result in a statistical bias toward events with short barrier-crossing times. To address this issue, we employed the rate from event durations (RED) scheme to apply a correction using the event-duration distribution to eq 13, to mitigate such bias.[43]

With regards to membrane permeability, the probabilities from eq 12 are proportional to the concentrations by a constant. Given the total concentration, $C_0 = m_{total}/V$

$$\tilde{p}_D = \frac{m_D/V_D}{m_{total}/V} = \frac{C_D}{C_0}$$

$$\tilde{p}_A = \frac{m_A/V_A}{m_{total}/V} = \frac{C_A}{C_0} \tag{14}$$

where $V_D = V_A = V$ due to the symmetry of the system and $m_{total} = m_D + \sum_{i\neq D,A} m_i + m_A$ is the total amount of the permeant. By assuming $m_D \gg \sum_{i\neq D,A} m_i$, we have $\tilde{p}_D \approx p_D$, and it can be seen from inspection that the rate constants in eq 4 are the same as eq 13, and eq 11 still holds true if populations of a molecule in compartment D and A are represented by probabilities. This procedure is like the transition rate-based counting method,[18] but instead of simply counting the number of crossing events in a simulation, the probabilistic weight of each crossing event is also considered.

Ultimately, the expression that we used to estimate the permeability coefficient from WE steady-state fluxes can be derived by combining eqs 11 and 13

$$P_m = k_{D\to A} l_D = \langle \hat{f}_{D\to A} \rangle l_D \tag{15}$$

METHODS

System Preparation and MD Equilibration. The final system used in the WE simulations was prepared using several individually constructed molecular systems that were pieced together in the following way. First, ParmEd[44] was used to generate input files containing Open Force Field Parsley v1.3.1a1 force field[45] parameters for the drug-like molecule as well as Amber LIPID17[46,47] parameters for the POPC membrane. Next, an initial solvated POPC membrane configuration was generated using CHARMM-GUI v2.0[48] with the Membrane Builder[49] input generator module and 50 POPC molecules per leaflet in a solution of TIP3P[50] water molecules. The solvated membrane was equilibrated using a slightly modified set of CHARMM-GUI scripts such that the OpenMM 7.5 MD[51] engine could be used with a single GTX1080 GPU to equilibrate the system for 0.5 μs in the NPT ensemble. For each drug-like molecule, a graph representation of the molecule was converted to a three-dimensional structure using OEChem Toolkit 3.2.0.0[52,53] followed by the generation of a diverse set of conformers using Omega Toolkit 4.1.2.0.[54−56] The top 20 conformers ranked by Omega were each randomly oriented in compartment D relative to the pre-equilibrated lipid bilayer and solvated by a 2 nm layer of water using PACKMOL[57] at a density of 1 gm/cm^3. The 2D chemical structures were visualized using Picto 4.5.3.0,[58] and 3D molecular structures including the snapshots and movies shown in this study were rendered using VMD 1.9.4.[59]

Each of the 20 solvated drug-membrane systems were subjected to energy minimization until convergence using the L-BFGS method and then equilibrated in the NPT ensemble in two stages. In the first stage, each system was gradually heated from 0 to 308 K over 0.01 ns, while applying a weak harmonic restraining potential with a force constant of 2 kcal/mol/Å^2 to all atoms of the drug-like molecule. In the second stage, the force constant was gradually reduced from 2.0 to 0.1 kcal/mol/Å^2 over 0.06 ns of equilibration. The resulting 20 equilibrated systems were used as starting conformations for WE simulations.

Weighted Ensemble Simulations. WE simulations were run in the following manner using a Python API of the WESTPA 2.0 software package.[60] To maintain non-equilibrium steady state conditions, trajectories that reached a target state of compartment A (i.e., z = 3.0 nm, see Figure 1) were "recycled", starting a new trajectory from the initial state (compartment). A one-dimensional progress coordinate was divided into bins using two different schemes: a manual, fixed binning scheme and the minimal adaptive binning schemes(MAB) scheme[61] (see below). A resampling interval of 0.1 ns was applied with a target number of 5 trajectories per bin. The simulations required 50 ns, which corresponds to 500 WE iterations, to reach reasonable convergence of the permeability coefficient.

Progress Coordinate and State Definitions. An effective progress coordinate for a WE simulation captures the slowest motion that is relevant to the rare-event process of interest. For membrane permeation, an effective progress coordinate is the

https://doi.org/10.1021/acs.jcim.1c01540
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

distance between the center of mass of the molecule ($\boldsymbol{q}_M$) to that of the lipid bilayer ($\boldsymbol{q}_L$) along the unit vector normal to the bilayer surface, $\hat{\mathbf{z}}$

$$z' = (\boldsymbol{q}_M - \boldsymbol{q}_L)\cdot\hat{\boldsymbol{z}} \tag{16}$$

Given a lipid bilayer with a width of ∼40 Å, a molecule with $z' < -20$ Å is in compartment D, a molecule within -20 Å $\leq z' \leq$ 20 Å is inside the membrane with $z' = 0$, indicating the center of the membrane, and a molecule with $z' > 20$ Å is in compartment A (Figure 1A). A change in z' from anywhere smaller than −20 Å to +30 Å is considered a membrane crossing event. Note that the extra 10 Å beyond +20 Å accounts for the interfacial solvation layer that will have different properties compared to bulk water.

The progress coordinate defined in eq 16 works well in an ideal setup, where there is a single lipid bilayer sandwiched between two aqueous compartments of infinite volumes. However, correction is needed to account for the periodic boundary conditions imposed by the simulation protocol. This correction can be written as

$$z = \left|z' + \frac{L_z}{2}\right| - \frac{L_z}{2} \tag{17}$$

where L_z is the length of the simulation box along $\hat{\mathbf{z}}$. The corrected progress coordinate, z, is guaranteed to change sign correctly from negative to positive when the molecule crosses any lipid bilayer, which can be in either the unit simulation cell or its periodic neighbor (Figure 1C).

Binning Schemes. The fixed binning scheme separates the water compartments ($z < -20$ Å or $z > 20$ Å) into fixed sized bins of 2 Å and the membrane region (-20 Å $< z <$ 20 Å) into 0.5 Å-wide bins. We used 30 dynamic linear bins for all our simulation using the MAB scheme.[61]

Dynamics Propagation. Dynamics were propagated using the OpenMM 7.5 MD engine in the NPT ensemble. As required for WE simulations, a stochastic thermostat was used to maintain the temperature at 308 K, that is, a Langevin thermostat with a collision frequency of 1 ps^{-1}. Constant pressure was maintained at 1 atm by anisotropically coupling the Monte Carlo membrane barostat to the system with 30 fs between attempts to adjust the system volume. A 1 nm cutoff was used for short-range nonbonded interactions, while the particle mesh Ewald (PME) method[62] was applied for the treatment of both long-range electrostatics and Lennard-Jones interactions. To enable a 2-fs timestep, the SHAKE[63] or SETTLE[64] algorithm was used to constrain the lengths of bonds to hydrogens. The trajectories were processed and analyzed using MDTraj.[65]

Reweighting Trajectories for a Steady State. To accelerate convergence of the WE simulation to a steady state, trajectory weights were adjusted using a WE steady-state (WESS) reweighting procedure that makes use of rates among "arbitrarily" defined bins.[35]

To reweight trajectories for a steady state, we applied the WESS reweighting procedure (see Bhatt et al.[35] for full details). Briefly, the WESS procedure uses the rates k_{ij} of transition between bins (i and j) to solve for the steady-state (SS) bin populations p_i^{SS} using the standard SS relation $\sum_j p_i^{SS} k_{ij} = \sum_j p_j^{SS} k_{ji}$. Trajectory weights are scaled iteratively at every fixed resampling interval τ_w to match the p_i^{SS} values. In the present study, the bins used for the reweighting procedure are the bins used for the manual, fixed binning scheme and reweighting was performed every τ_w = 5 ns.

Estimating Uncertainties. A Monte Carlo bootstrap procedure was used to estimate uncertainties in calculated permeability coefficients.[34] In short, given a simulation with a total of N iterations, Monte Carlo bootstrapping was performed on the original WE dataset of fluxes into state A accumulated from individual iterations, $\{\hat{f}^{(i)}_{D\to A} | i \in 1, 2, ..., m\}$, to generate 1000 subsamples. The reported rate constant is the mean among the subsamples (eq 13), and the associated uncertainty is the 95% confidence interval.

Free Energy Profiles. Free energy profiles have been generated as a function of the WE progress coordinate z and "auxiliary" coordinates that are orthogonal to the progress coordinate by symmetrizing the probability distribution as a function of the z-coordinate such that the resulting two sets of non-equilibrium steady state trajectories in opposite directions could be combined to form a set of equilibrium trajectories.[66] The auxiliary coordinates are the following:

1 The cosine of the angle relative to $\hat{\mathbf{z}}$, which is defined through the vector product of the unit electric dipole moment of the molecule and $\hat{\mathbf{z}}$, quantifies the relative orientation of the molecule through the membrane.
2 The number of hydrophobic contacts is defined to be the number of aliphatic atoms of the lipid tails within the 10 Å distance of any hydrophobic atom of the drug molecule, following Rogers and Geissler in their studies of lipid insertion.[25] The hydrophobic atoms used in this study are highlighted in Figure S6.
3 The number of hydrogen bonds between the drug and the membrane was identified using the Baker-Hubbard definition[67] as implemented in the MDTraj package. According to this definition, any donating NH or OH is assumed to be in a hydrogen bond with any accepting N or O if the bond angle is greater than 120° and the bond distance is less than 2.5 Å.
4 The end-to-end distance of each molecule was calculated based on the largest separated atoms of the molecule identified in the longest axis. The atoms used for each drug-like molecule are highlighted in Figure S6.

Auxiliary coordinates are separate from the WE progress coordinate and were calculated after the simulations were completed. If the auxiliary coordinates are orthogonal to the progress coordinate, the sampling along these auxiliary coordinates is equivalent to that of unbiased, conventional "brute force" MD. If the auxiliary coordinates are not orthogonal to the progress coordinate, the WE strategy would enhance the sampling of these coordinates. Even if the sampling of auxiliary coordinates is not extensive, the results of our permeability simulations would not be as impacted as the results of force-biasing methods by the choice of progress coordinate. In addition, auxiliary coordinates can serve as useful tools for gaining mechanistic insights into how a molecule crosses (or does not cross) the membrane and which permeation pathways would lead to the permeability estimated from the simulations.

Cloud Computing in Orion. Scientific computing in OpenEye's Orion cloud platform is patterned after the concept of flow-based programming (FBP).[68] In FBP, a series of compute kernels are defined to perform actions on input data and pass (transformed) output data to subsequent kernels. FBP is designed around the idea that information will flow from one kernel to another through connections known as ports, with complex operations occurring in each kernel process. Compute kernels within Orion's FBP vocabulary are known as "Cubes".

https://doi.org/10.1021/acs.jcim.1c01540
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

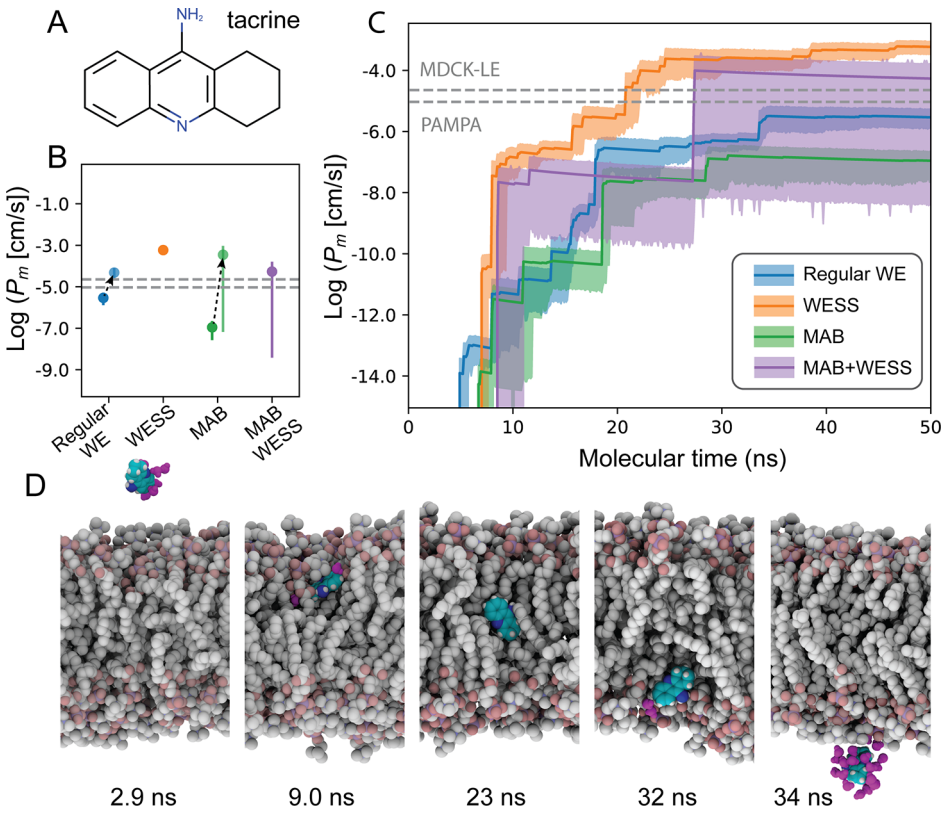


Figure 2. Estimated permeability coefficients of tacrine. (A) Chemical structure of tacrine. (B) Average (log) permeability coefficients (with 95% confidence intervals) calculated from each of the four WE simulations with different protocols [see the legend in panel (C)]. The gray dashed lines indicate the observed values from MDCK-LE and PAMPA experiments. Dashed arrows indicate the raised permeability estimates by applying WESS reweighting to Protocol 1 and 3 after the simulations were completed (50 ns). (C) Time evolution of estimated logarithm of the permeability coefficients (cm/s) for tacrine from four WE simulations, using a fixed binning scheme with (regular, *blue*) or without WESS (*orange*), and the MAB scheme with (*green*) or without WESS (*purple*), respectively. The solid lines indicate the mean values of the estimates, and the shaded areas indicate 95% confidence intervals. The dashed gray line indicates the permeability coefficient measured by an MDCK-LE experiment.[69] The molecular time is represented as $N\tau$, where N is the number of WE iterations and τ is the fixed time interval (100 ps) of each WE iteration. (D) Snapshots of the tacrine molecule (*cyan*) passing through the lipid bilayer (*gray*) at selected molecular times. Water molecules in close contact with the molecule are highlighted in *magenta*.

Cubes are connected to each other through a series of ports, whereby data can flow from one Cube to another in the form of one of several strongly typed data structures. Within the Orion FBP framework, finalized workflows with a set of logically connected cubes are known as "Floes". All Cubes and Floes are written in Python 3, where a Cube is an instance of a Cube class and Floes are in many ways the equivalent of a Python script. In Orion, all compute nodes are sourced from Amazon Web Services (AWS). Each Cube runs on its own AWS instance that itself runs in an isolated computing unit called a Docker container. The Orion platform handles all sourcing and scheduling of Cubes onto AWS instances. Floes are uploaded as Python-like packages, which can depend on other Python packages sourced from Anaconda or public pip repositories.

OpenEye Permeability Floe. The OpenEye Permeability floe in Orion contains a series of Cubes that each performs one of the following functions: system preparation, MD equilibration, WE simulation, and permeability analysis of the membrane-permeate system (see Figure 1B for an example of the flow relationship diagram of the compute kernels and above sections for details). The Simulation Manager and Segment Runner Cube, respectively, handle most of the WE logic and the MD propagation and, therefore, are connected to each other in a cyclic fashion to enable bidirectional communication between the WE driver and the MD engine.

The Permeability Simulation Floe is hosted on the Orion platform where all the simulation setup and actual computation (including system preparation and WE/MD simulation) took place. The Floe takes a molecule of interest as input, which can be either a graph representation (2D) of a molecule sketched into the molecular editor or a pre-generated 3D structure of a molecule. If a 2D molecule was given as input, the Floe will automatically convert it to a 3D structure using the OEChem Toolkit. The stereochemistry of the molecule is automatically handled by the Omega Toolkit, which will respect pre-defined stereochemistry if such information is provided in the molecular graph. The Floe also exposes various parameters for system preparation and the WE simulation including the option to turn on/off the MAB scheme for adaptive binning[61] or the bin probability reweighting for faster convergence to the steady state[35] (Figure S3). The input molecules and the WE protocols used in this study were set using the floe's graphical user interface (GUI).

Finally, the WE simulation is automatically parallelized and performed on CPUs or GPUs from either spot or non-spot AWS instances sourced by the Orion computing platform. Typically, the simulation will be automatically scaled-up by Orion to several hundreds of GPUs or thousands of CPUs per WE iteration. Permeability simulations are analyzed on-the-fly, and a simulation report describing important features of the reactive

https://doi.org/10.1021/acs.jcim.1c01540
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 1. Efficiencies of Different WE Protocols in Predicting the Membrane Permeability (Log P_m) of Tacrine

compound	MAB	WESS	platform	total simulation time (μs)	wall clock time (Orion days)	predicted log P_m (cm/s)
tacrine			GPU	23.0	15.5	-5.54 ± 0.13 (-4.32 ± 0.11)[a]
		×	GPU	23.0	16.2	-3.23 ± 0.09
	×		GPU	8.2	7.1	-6.96 ± 0.16 (-3.46 ± 0.23)[a]
	×	×	GPU	8.1	7.9	-4.27 ± 0.24
	×	×	CPU	5.6	11.4	-5.20 ± 0.28

[a]Permeability estimates in parentheses were calculated by applying the WESS reweighting procedure after the simulation was finished.

trajectory data is generated upon completion (Figure S4). The report provides the time-evolution of the permeability coefficient with a 90% confidence interval, probability distribution as a function of the progress coordinate z, and visualization of trajectories—both fully downloadable trajectories as well as visual schematics of membrane crossing—that successfully reached the acceptor (A) compartment with their associated probabilities.

RESULTS

Here, we present the results from fully automated permeability simulations performed on three "rule of five" molecules (tacrine, zacopride, and sotalol) using the OpenEye Permeability Floe package in the Orion cloud computing environment. These drug-like molecules are weakly basic primary or secondary amines that vary in size, shape, and number of rotatable bonds.

Evaluation of WE Protocols. To determine an effective WE protocol for simulating the membrane permeability for a drug-like molecule, we focused on tacrine, the simplest compound in this study with zero rotatable bonds (Figure 2A). In particular, we assessed the advantages of applying an adaptive binning scheme (MAB scheme) and WE steady-state (WESS) reweighting procedure by testing four WE protocols on GPUs: (1) standard WE with a manual, fixed binning scheme, (2) WE with the MAB scheme, (3) WE with the WESS reweighting procedure, and (4) WE with the MAB scheme and WESS reweighting (see Methods for full details). We also ran a WE simulation using protocol #4 on CPU cores rather than GPUs to perform a cost-benefit analysis of using GPUs over CPU cores. All the WE protocols yielded permeability coefficients (log P_m: -6.95–-3.23) that are in reasonable agreement with the value measured by MCDK-LE (-4.64)[69] and PAMPA (-5.02 ± 0.2),[70] Figure 2B and Table 1).

Relative to WE simulation with a manual binning scheme and no reweighting, use of the MAB scheme reduces the required total simulation time by ~65% (by comparing Protocols 3 and 4) and use of the WESS reweighting procedure reduces the required total simulation time by ~60% (by comparing Protocols 1 and 2, where Protocol 2 reached the final estimate of Protocol 1 at around 20 ns. See Figure 2C and Table 1). The combined use of the MAB scheme and WESS reweighting procedure reduces the required total simulation time by roughly threefold. Reasonably converged permeability coefficients were obtained within 50 ns of molecular time, which is defined as $N\tau$, where N is the number of WE iterations and τ is the fixed time interval (100 ps) of each WE iteration. Without the application of the WESS reweighting scheme, the permeabilities estimated from the simulations using the fixed binning (Protocol 1) and the adaptive binning (Protocol 3) differ by about 1.5 log units. In contrast to protocols that applied reweighting, the results from the different binning schemes (Protocols 2 and 4) differ by only one log unit (Table 1). The smaller difference between estimated permeability coefficients in the latter case is due to reweighting to a steady state. Once the reweighting procedure was applied after 50 ns of simulation using Protocols 1 and 3, the estimated permeability coefficients became comparable to those from Protocols 2 and 4 (Figure 2B). The fact that these successful membrane-crossing trajectories are orders of magnitude shorter than the mean first-passage time for membrane permeation indicates that the membrane-crossing events are indeed rare events. Said another way, the trajectories include solely the relatively fast transitions with low probability between stable states, leaving out the long waiting times in the initial stable state for the low-probability transitions.

In addition, we calculated the evolution of the progress coordinate (eq 17) of trajectories that successfully reached the target state in compartment A. This estimate is based on thousands of crossing events forked from about 10 *independent* events using standard (regular) WE with a fixed binning scheme (Figure S4C). One of the top-weighted, successful trajectories (probabilistic weight: 6.0×10^{-6}) was extracted for visualization at the atomic level (Figure 2D and Movie S1). With the MAB scheme, we were able to obtain comparable permeability estimates using much fewer bins compared to a fixed binning scheme (34 bins with MAB vs 100+ with fixed bins). Because of the reduced number of bins, MAB-based simulations require about 3 times less aggregate simulation time compared to fixed binning as shown in Table 1 (~8 μs for MAB vs ~23 μs for fixed bins). To do so, we can analyze the fixed binning data at around 8 μs aggregate time, which for Protocol 1 (fixed binning; no reweighting) would yield a permeability estimate of -6.55 on the log scale at iteration 194. This value is comparable to the permeability coefficient (-6.95) estimated by Protocol 3 at iteration 500 (MAB; no reweighting); however, the (log) permeability estimate from Protocol 2 (fixed binning, with reweighting) is still -5.56 at iteration 195. The fixed binning estimates with only 8 μs of aggregate simulation time are nearly 2 log units away from the MDCK-LE experiments (-4.64), whereas MAB with reweighting can produce a more accurate permeability estimate (-4.27) in a much shorter amount of aggregate simulation time. Nonetheless, the large lower CI bound for simulation with the MAB and WESS protocol was a result of scattered incoming fluxes into the target state due to significant shorter aggregate time as compared to other protocols (see total simulation time in Table 1).

Mechanisms of Membrane Permeation for Tacrine, Sotalol, and Zacopride. To evaluate the free energy profile of the molecule along the z-coordinate from the non-equilibrium steady-state ensembles generated by our simulations, we symmetrized the resulting probability distribution as a function of the z-coordinate (see Methods; for the original, unsymmetrized probability distribution for tacrine, see Figure S5). Using tacrine as an example, the free energy profile in Figure 3A shows that the largest barrier of the permeation process of the molecule was located at around z = 0 Å, that is, the center of the membrane. ~30 ns was needed for the free energy profile to

https://doi.org/10.1021/acs.jcim.1c01540
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

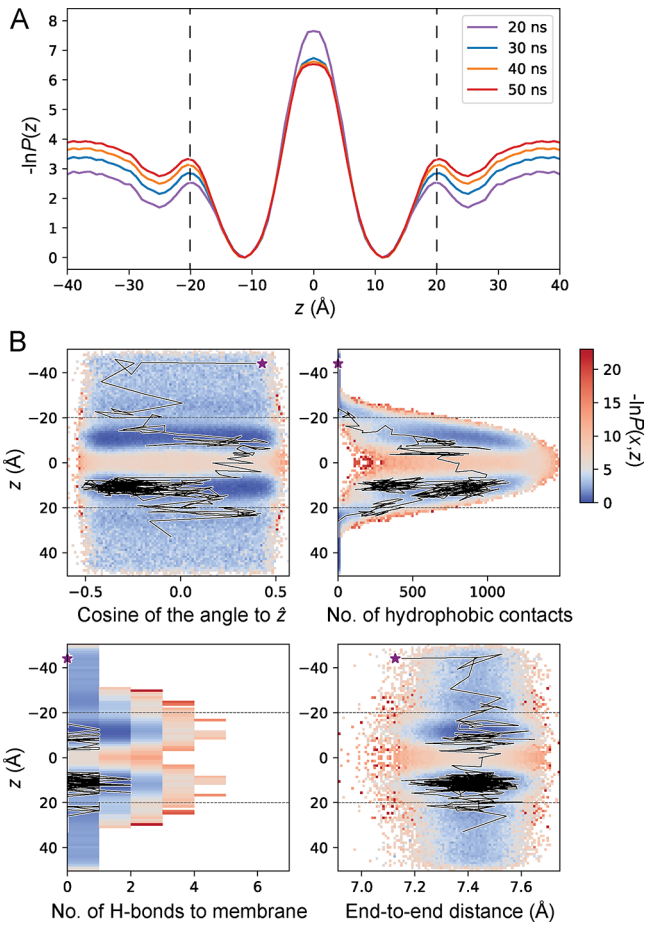


Figure 3. Free energy profile along the lipid normal (z) for tacrine using the fixed binning and reweighting WE protocol (WESS). (A) Free energy profile (in units of k_BT) of tacrine along the bilayer normal, z. The probability distribution along z, $P(z)$, was extracted from the simulation at different molecular times and symmetrized (see Methods; Figure S5B for the unsymmetrized distribution). (B) Free energy profile along z and the cosine of the angle of the molecule (dipole moment) with respect to z (*top left*), hydrophobic contacts between the molecule and the membrane (*top right*), the number of hydrogen bonds between the molecule and the membrane (*bottom left*), and the end-to-end distance of the molecule (*bottom* right, *blue*: $<5k_BT$, *red*: $>5k_BT$). The *solid black line* represents the top weighted trajectory (statistical weight: 6.0×10^{-6}) and the *purple star* indicates the start location. The approximate range of the membrane region is indicated by *black dashed lines* (-20 Å $< z <$ 20 Å). For all 2D distributions, the probabilities are symmetrized across the membrane to obtain the free energy profiles.

converge to the final profile determined by the total 50 ns simulation. There is a smaller barrier at the membrane surface ($z = \pm 20$ Å), followed by an energy minimum near the center of each leaflet ($z = \pm 10$ Å).

We also calculated four auxiliary progress coordinates, namely, the cosine of the angle of the unit electric dipole moment to $\hat{\mathbf{z}}$, the number of hydrophobic contacts, the number of hydrogen bonds between the molecule and the membrane, and the end-to-end distance of the molecule to evaluate molecular orientation, hydrophobic interactions, and hydrogen bond structure with respect to the main progress coordinate, z, for tacrine (Figure 3B), zacopride (Figure 4C), and sotalol (Figure 5C). Interestingly, all three molecules passed through the membrane with a 60−120° angle (±30° with respect to $\hat{\mathbf{z}}$), with a relatively constrained angle near the center and interface of the membrane, especially for zacopride and sotalol. All three molecules formed increasing numbers of hydrophobic contacts with the membrane as the molecules approach the center of the membrane and losing them when exiting from the center (Figure S6). All three molecules formed more hydrogen bonds near the headgroup region than in the center of the membrane. Tacrine, as expected, did not undergo a large conformational change crossing the membrane according to the end-to-end distances of the molecule (7.2−7.6 Å) from the simulation using Protocol 2 (fixed binning + WESS), but it was observed "flipping," as previously predicted, as a model for membrane permeation.[29] Similar energy barriers and bottleneck regions can be observed for tacrine from the simulation using the adaptive binning protocol (MAB + WESS, Figure S7). Owing to rigid nature of the molecule, zacopride seems to have crossed the membrane adopting a mostly extended form (end-to-end distance ∼10.7 Å), whereas sotalol seems to be relatively more flexible inside the membrane interior (end-to-end distance between 9 and 12 Å).

Overall, these 2D free energy profiles suggest that our WE simulations with a single (and simple) progress coordinate were able to sample a variety of conformations that allowed the molecules to pass through the membrane. The distributions shown here also reinforce the idea that choice in reaction coordinates may influence the interpretation of a rate-limiting step, which could be a problem for methods based on the ISD model. Depending on whether hydrogen bond count to the membrane, end-to-end distance, or the number of hydrophobic contacts were chosen as a reaction coordinate, one might believe a rate-limiting step that occurred at different positions within z, for example, in the center of the membrane for hydrophobic contacts or near the membrane−water interface for the hydrogen bond count.

We also extracted the trajectory with the highest probability for each molecule (see Figures 2D, 4B, and 5B for critical snapshots and see Movies S1−3 for full trajectories). Overall, the molecule adopted pathways consistent with our population-level observations above. Additionally, we observed water (<5 molecules at once) deep in the membrane (near $z = 0$), which was either due to the drug "dragging" water across the bilayer (sotalol), or by distorting the curvature of the outer leaflet of the membrane to bring water toward the bilayer center (zacopride). The water has been captured in both the visualized snapshots of each of the molecules as well as the movies of the top weighted simulations of drug molecules crossing the membrane.

Efficiency of WE Simulations. To estimate the efficiency of WE simulations relative to conventional MD simulations, we make two comparisons. First, we compare the total computing time that would be required in Orion using conventional MD simulations to generate a single permeation event given the corresponding MFPT estimated by our WE simulations (see eq 12). Second, we make a similar comparison to Anton3 from D. E. Shaw Research, which is currently known as the fastest MD simulation engine in the world. As shown in Table 2, conventional simulations in Orion would require 1.5−177 years to generate a single permeation event depending on the molecule, which is roughly equivalent to 22 weeks to over 7 years on the latest Anton3. Our WE simulations generated the first permeation event in 1.1 (tacrine), 10.7 (sotalol), or 7.5 (zacopride) days and a total of 872 (tacrine), 66 (sotalol), or 56 (zacopride) permeation events were observed within a 7.9 (tacrine), 12.7 (sotalol), and 11.7 (zacopride) day period. Although many of the pathways for these events are correlated,

https://doi.org/10.1021/acs.jcim.1c01540
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

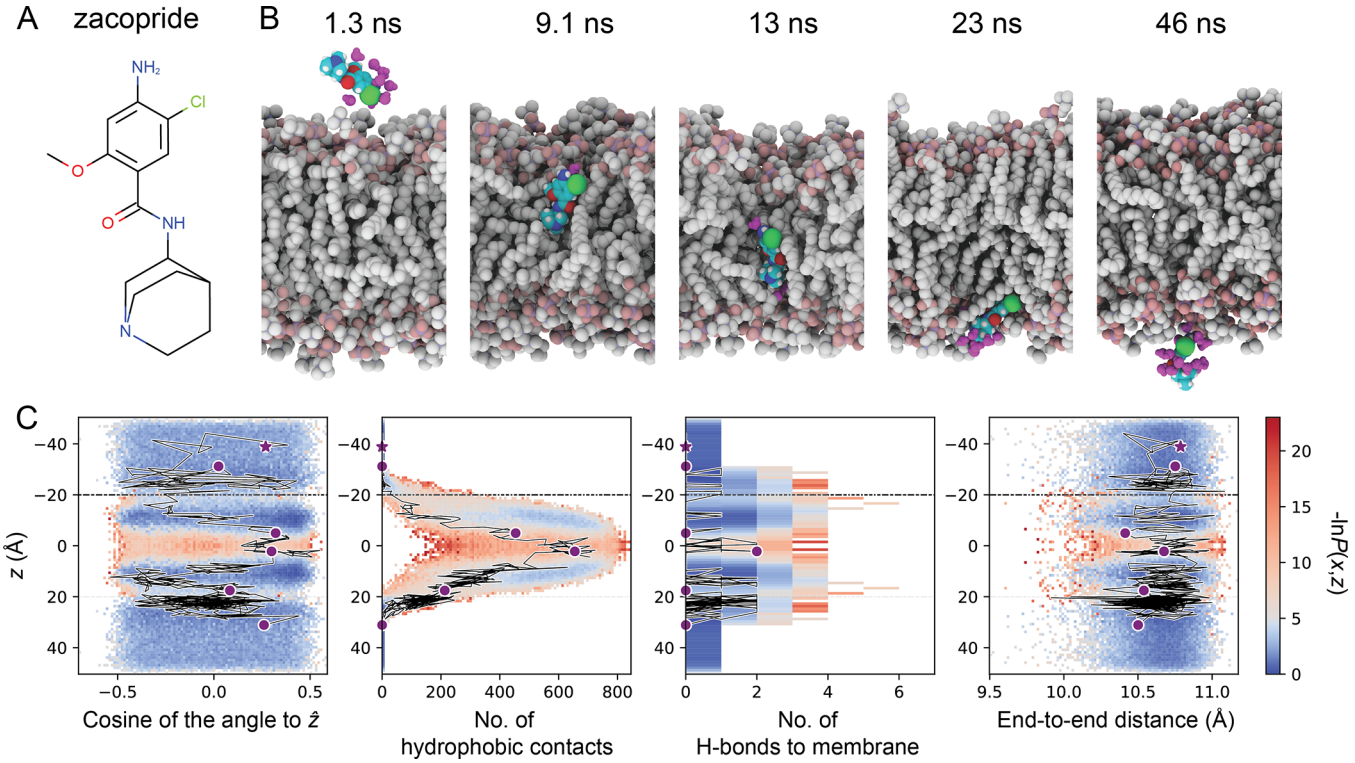


Figure 4. Mechanistic analysis of zacopride permeation. (A) Chemical structure of zacopride. (B) Snapshots of the zacopride molecule (*cyan*) passing through the lipid bilayer (*gray*) at selected molecular times (see the legend in Figure 2D). (C) Free energy profile (in units of k_BT) along the bilayer normal, z, the cosine of the angle of the molecule with respect to z (*top left*), hydrophobic contacts between the molecule and the membrane (*top right*), the number of hydrogen bonds between the molecule and the membrane (*bottom left*), and the end-to-end distance of the molecule (*bottom right, blue*: $<5k_BT$, *red*: $>5k_BT$). The *black line* represents the top weighted trajectory (probabilistic weight: 2.2×10^{-5}), with *purple dots* indicating the location of the snapshots in panel B and a *purple star* indicating the start location. The approximate range of the membrane region is indicated by *black dashed lines* (-20 Å $< z <$ 20 Å). For all 2D distributions, the probabilities are symmetrized across the membrane to obtain the free energy profiles.

sharing common trajectory segments, reasonably converged permeation coefficients were obtained.

DISCUSSION

Despite our use of a model membrane system, our estimated permeability coefficients from WE simulations for the three molecules (tacrine: -4.27 ± 0.24, zacopride: -6.35 ± 0.22, sotalol: -5.32 ± 0.22) are in reasonable agreement with experimentally measured values (tacrine: -4.64[69] or -5.03 ± 0.2,[70] zacopride: -5.23,[69] sotalol: -6.02,[69] -5.58 or -6.74;[71] see Figures 6 and S1 for detail). Absolute agreement of our calculated permeability coefficients with experiment would not, however, be expected due to several complexities of real cell membranes that are lacking in our simulation setup, particularly, for cell-line assays such as MDCK and CaCo-2. These complexities include the presence of multiple lipid species, (especially cholesterol), transmembrane proteins, and membrane rafts. Such membrane-associated structures can be particularly important for modeling the permeation process of more complicated cell types like those involved in the blood–brain barrier. Additionally, paracellular transport exists as a potential method for drug delivery. In future efforts, we will expand our model membrane (currently, 50 POPC lipids per leaflet) to accommodate larger molecules that fall outside the spectrum of Lipinski's "rule of five", such as peptides, natural products, or de novo designed biologics. In addition, we will improve on the diversity of pathways for membrane permeation by including orthogonal dimensions to the progress coordinate that can distinguish between different conformations of the molecule or the lipid bilayer, for example, radius of gyration of the drug or the extent of membrane curvature. These efforts will greatly aid the simulation of membrane permeation for a modern lead series with more than 10 rotatable bonds, including molecular glues and macrocycles that are highly flexible with large molecular weights (>kDa).

Additional concerns with the current work could be either the small number of molecules used for model development or the minimal dynamic range of the permeability data itself (Log P_m varies between about -4 and -6). Both issues should be addressed. Regarding the number of compounds, it is true that three molecules alone are not enough to provide the statistical insight needed to reliably judge the predictive power of the model for any given molecular species. Even so, for the molecules presented here, our model was able to predict permeability coefficients within about log unit of experimental MDCK-LE and PAMPA measurements. Such a discrepancy with respect to experiment can be considered small, especially compared to free energy-based methods, where permeability coefficients may be several log units away from a reference value.[20] Furthermore, one may also note that the agreement between two permeability estimates for the same molecule using the same experimental technique may not be in perfect agreement. In fact, reported PAMPA data for sotalol using either the top-to-bottom (TtoB) or the bottom-to-top (BtoT) protocol[71] provided permeability coefficients that differed by more than a log unit (see Figure 6). Note that the apparent discrepancy between 95% CIs in Figure 2C and standard deviations in Figure 6 is because the original CIs and standard

https://doi.org/10.1021/acs.jcim.1c01540
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

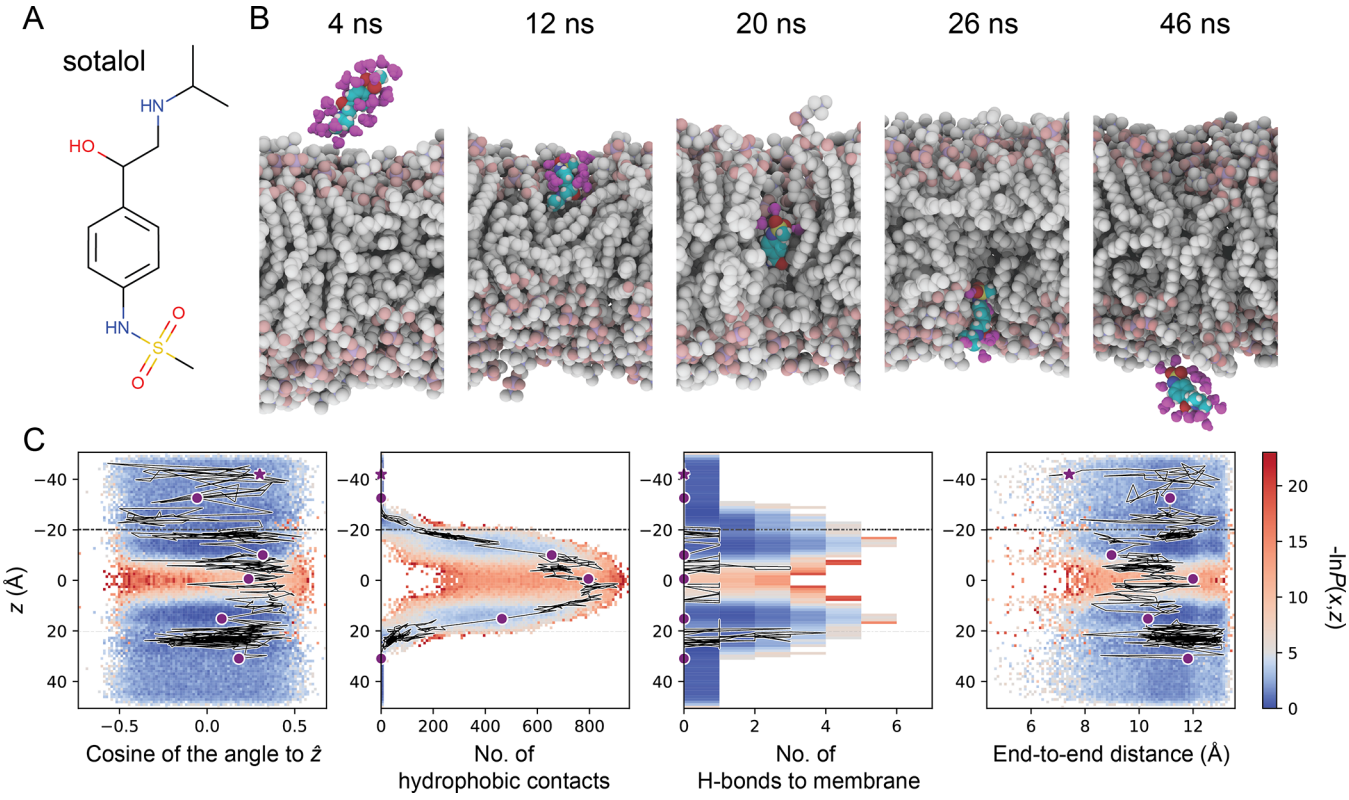


Figure 5. Mechanistic analysis of sotalol permeation. (A) Chemical structure of sotalol. (B) Snapshots of the sotalol molecule (*cyan*) passing through the lipid bilayer (*gray*) at selected molecular times (see the legend in Figure 2D). (C) Free energy profile (in units of k_BT) along the bilayer normal, z, and the cosine of the angle of molecule with respect to z (*top left*), number of hydrophobic contacts between the molecule and the membrane (*top right*), number of hydrogen bonds between the molecule and the membrane (*bottom left*), and the end-to-end distance of the molecule (*bottom right*, *blue*: $<5k_BT$, *red*: $>5k_BT$). The *black line* represents the top weighted trajectory (probabilistic weight: 9.3×10^{-6}), with *purple dots* indicating the location of the snapshots in panel (B) and a *purple star* indicating the start location. The approximate range of the membrane region is indicated by *black dashed lines* (-20 Å $< z <$ 20 Å). For all 2D distributions, the probabilities are symmetrized across the membrane to obtain the free energy profiles.

Table 2. Comparison of Wall-Clock Times Required by Weighted Ensemble MD Simulations Relative to Standard MD Simulations for Generating a Single Drug Membrane Crossing Event[a]

		weighted Ensemble MD				standard MD[b]	
		P_m estimation		a single crossing event		a single crossing event	
compound (no. of atoms)	predicated MFPT (ms)	total simulation time (μs)	wall clock time (Orion days)	simulation time (ns)	wall clock time (Orion days)	wall clock time (Orion years[c])	wall clock time (Anton3 years[d])
tacrine[e] (29)	4.6	8.1	7.9	1.4	1.1	1.5	0.06
sotalol (38)	52.1	6.3	12.7	3.7	10.7	16.5	0.71
zacopride (41)	559.2	6.0	11.7	3.0	7.5	177.3	7.7

[a]Wall-clock times are reported for the Orion cloud-computing platform on Amazon Web Services and the Anton3 special-purpose MD supercomputer, and they were estimated using the mean first passage time (MFPT) for the permeation process determined by weighted ensemble MD. The wall-clock time is also reported for the generation of a reasonably converged estimate of the permeability coefficient P_m, which is not practical to estimate using standard MD. [b]Expected times based on the MFPTs predicted by the WE simulations. [c]Using 100 GPUs at a time on Orion at the speed of 8600 ns per day. [d]Assuming 200 μs per day for a similar sized system (~33k atoms), as reported using the 64-node Anton3 performance data for DHFR and ApoA1[74] from D. E. Shaw Research. [e]Data from the tacrine run using MAB and WESS.

deviations were defined in the linear scale and converted to the log scale for visual representation.

As shown by Rogers and Geissler,[25] the choice of using z as the reaction coordinate for the free energy-based methods could have an enormous impact on the perceived free energy barrier, which may contribute to multi-log unit differences in the absolute comparison to permeability experiments mentioned above. While a reaction (progress) coordinate is typically defined for the WE strategy, the computed rates are independent of the chosen coordinate. Therefore, even though the efficiency of the computation can be reduced by a poor choice of reaction coordinate, the results are much less sensitive to it compared to free energy-based methods. In the most extreme case, when a reaction coordinate is suboptimal, WE simulations can surmount barriers along orthogonal coordinates in a "brute force" manner, whereas free energy-based methods have no means of recovering. Even if there is no enhanced sampling of auxiliary coordinates, the predicted permeability coefficients derived from permeation rates are in reasonable agreement with the experiment. Meanwhile, permeation rates computed from free energy methods depend exponentially on the barrier height, which is in turn sensitive to the choice of the progress

https://doi.org/10.1021/acs.jcim.1c01540
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

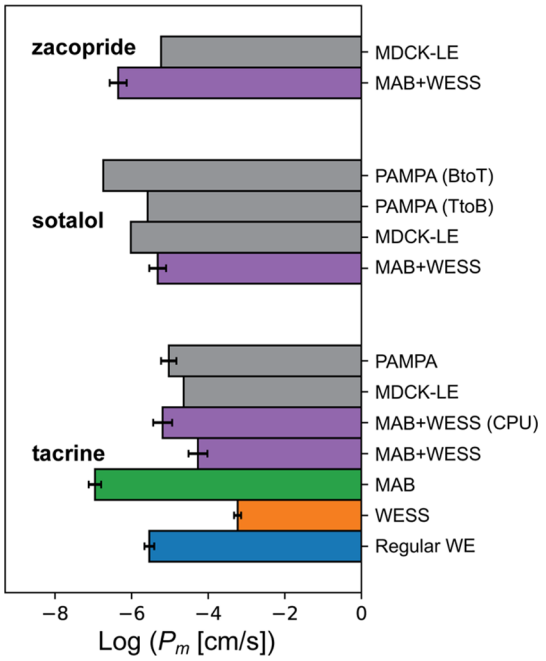


Figure 6. Membrane permeabilities (Log P_m) calculated using various WE simulation protocols for tacrine, sotalol, and zacopride. Uncertainties represent standard deviations, which are evaluated as 1/4th the difference between the 95% CI upper bound and the lower bound. Experimentally measured values are shown in gray [MDCK-LE: Dickson et al. (2019)],[69] PAMPA for tacrine: Katt et al. (2016),[70] and PAMPA for sotalol: Liu et al. (2012).[71] See also Table S1 in the Supporting Information.

coordinate. That said, as a future direction, we will apply a committor analysis technique[72] to characterize the mechanism of drug membrane permeation processes. We will also apply a history-augmented Markov state model analysis[73] instead of the WESS reweighting procedure to further enhance the efficiency of obtaining steady-state observables.

While this work advances the methodology for computational prediction of permeability coefficient, several limitations prevent using the current algorithm for high-throughput screening. First, even though the amount of computation is a fraction of that required of brute force MD simulations (see Table 2), the computational cost is still too large for a virtual screen where billions of molecules can be routinely analyzed to find a lead candidate. Second, the amount of required wall-clock time per permeability estimate is roughly one week with our current WE setup employing the MAB scheme for adaptive binning and the WESS reweighting scheme. One week per compound is outside the time budget of a typical virtual screen since many orders of magnitude more compounds could be completed within the same period using a faster, albeit less accurate, computational technique. Still, we could imagine our WE method being used in later stages of the drug development cycle, primarily in the lead optimization phase where more expensive computational techniques are routinely applied. For instance, it is widely accepted that the high attrition rate for small-molecule drug candidates is largely due to ADME/Tox liabilities. Therefore, a method like ours that provides detailed mechanistic insights into the membrane permeation process could be employed at the lead optimization stage to reduce such high drug candidate attrition.

The OpenEye Permeability Floe used for our WE simulations is designed to be a user-friendly cloud application that can assist modelers in designing drugs for increased bioavailability. The cloud-based Floe is system-agnostic, enabling users to run and analyze simulations on any workstation with almost no hardware requirement aside from Internet connection. In our present study, the simulations required one to two weeks to finish depending on the WE protocol. This wall-clock time can be further reduced by horizontal scaling, that is, recruiting more nodes or scaling groups. Each WE iteration typically requires a few hundred trajectory segments in total, and each trajectory segment only requires 2 min of wall clock time to complete with a single GPU of the Amazon EC2 G4 instance types. Ideally, if each trajectory segment was assigned to a dedicated node, each WE iteration would complete in only 2 min, and a 500-iteration simulation would complete in only 16 h with a significantly increased demand of computing resources. A future direction is to further optimize the balance of the required computation and run time. The Floe also features a GUI that facilitates the set up for a permeability simulation with their drug-like molecule of interest. The user can increase or decrease the simulation length for analysis if needed, and an automatic detection of convergence in the estimated permeability coefficient is under investigation that will trigger the floe's termination. As mentioned above, the WE strategy not only yields a direct calculation of the passive permeability coefficient but also provides fully continuous trajectories revealing how molecules cross through the membrane.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c01540.

Structural layout of the OpenEye Permeability Floe, estimated permeability as a function of the compartment width, floe setup GUI of the OpenEye Permeability Floe, simulation report figures generated by OpenEye Permeability Floe, original, unsymmetrized probability distribution along the lipid normal ($\hat{\mathbf{z}}$) for tacrine using the reweighting WE protocol, chemical structures of tacrine, sotalol, and zacopride, free energy profiles along the lipid normal ($\hat{\mathbf{z}}$) and auxiliary coordinates for tacrine using the MAB scheme and reweighting WE protocol, and predicted and experimentally determined permeabilities (PDF)

Simplified molecular input line entry system (TXT)

Permeation of tacrine (MP4)

Permeation of zacopride (MP4)

Permeation of sotalol (MP4)

■ AUTHOR INFORMATION

Corresponding Authors

Lillian T. Chong − *Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States;* orcid.org/0000-0002-0590-483X; Email: ltchong@pitt.edu

David N. LeBard − *OpenEye Scientific, Santa Fe, New Mexico 87508, United States;* orcid.org/0000-0002-0979-800X; Email: david.lebard@eyesopen.com

https://doi.org/10.1021/acs.jcim.1c01540
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Authors

She Zhang − *OpenEye Scientific, Santa Fe, New Mexico 87508, United States;* orcid.org/0000-0001-9265-4498

Jeff P. Thompson − *OpenEye Scientific, Santa Fe, New Mexico 87508, United States;* orcid.org/0000-0003-3134-9451

Junchao Xia − *OpenEye Scientific, Santa Fe, New Mexico 87508, United States*

Anthony T. Bogetti − *Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States;* orcid.org/0000-0003-0610-2879

Forrest York − *OpenEye Scientific, Santa Fe, New Mexico 87508, United States*

A. Geoffrey Skillman − *OpenEye Scientific, Santa Fe, New Mexico 87508, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.1c01540

Notes

The authors declare the following competing financial interest(s): S.Z., J.P.T, J.X., A.T.B., F.Y., A.G.S., and D.N.L. are current or former employees of OpenEye Scientific, and L.T.C. is a Scientific Advisory Board member at OpenEye Scientific and an Open Science Fellow at Roivant Sciences.
Data and Software Availability: All simulation data are stored on Orion and will be shared upon request due to their enormous file sizes. OpenEye Permeability Floe is available on the academic stack of Orion, which is also available upon request.

■ ACKNOWLEDGMENTS

The authors would like to thank Anthony Nicholls (OpenEye Scientific), Alan Grossfield (University of Rochester), and Jesper Sørensen (OpenEye Scientific) for reviewing and giving feedback on this work; Jesus Izaguirre (Roivant Sciences) for stimulating discussions; Gaetano Calabro (OpenEye Scientific) and Christopher Bayly (OpenEye Scientific) for sharing the MD Orion Cube package; and OpenEye Scientific for computing time provided through the Orion cloud computing platform. This work was supported by an NIH grant (R01 GM115805) to L.T.C. and a University of Pittsburgh Andrew Mellon Graduate Fellowship to A.T.B.

■ REFERENCES

(1) Li, A. P. Screening for Human ADME/Tox Drug Properties in Drug Discovery. *Drug Discov.* **2001**, *6*, 357−366.

(2) Dobson, P. D.; Kell, D. B. Carrier-Mediated Cellular Uptake of Pharmaceutical Drugs: An Exception or the Rule? *Nat. Rev. Drug Discovery* **2008**, *7*, 205−220.

(3) Goldberg, G. S.; Valiunas, V.; Brink, P. R. Selective Permeability of Gap Junction Channels. *Biochim. Biophys. Acta Biomembr.* **2004**, *1662*, 96−101.

(4) Harris, A. L. Connexin Channel Permeability to Cytoplasmic Molecules. *Prog. Biophys. Mol. Biol.* **2007**, *94*, 120−143.

(5) Sun, H.; Nguyen, K.; Kerns, E.; Yan, Z.; Yu, K. R.; Shah, P.; Jadhav, A.; Xu, X. Highly Predictive and Interpretable Models for PAMPA Permeability. *Bioorg. Med. Chem.* **2017**, *25*, 1266−1276.

(6) Irvine, J. D.; Takahashi, L.; Lockhart, K.; Cheong, J.; Tolan, J. W.; Selick, H. E.; Grove, J. R. MDCK (Madin-Darby Canine Kidney) Cells: A Tool for Membrane Permeability Screening. *J. Pharm. Sci.* **1999**, *88*, 28−33.

(7) Hubatsch, I.; Ragnarsson, E. G. E.; Artursson, P. Determination of Drug Permeability and Prediction of Drug Absorption in Caco-2 Monolayers. *Nat. Protoc.* **2007**, *2*, 2111−2119.

(8) Taillardat-Bertschinger, A.; Carrupt, P.-A.; Barbato, F.; Testa, B. Immobilized Artificial Membrane HPLC in Drug Research. *J. Med. Chem.* **2003**, *46*, 655−665.

(9) Overton, C. E. *Über Die Osmotischen Eigenschaften Der Lebenden Pflanzen-Und Tierzelle*; Fäsi & Beer, 1895.

(10) Menichetti, R.; Bereau, T. Revisiting the Meyer-Overton Rule for Drug-Membrane Permeabilities. *Mol. Phys.* **2019**, *117*, 2900−2909.

(11) Missner, A.; Pohl, P. 110 Years of the Meyer-Overton Rule: Predicting Membrane Permeability of Gases and Other Small Compounds. *ChemPhysChem* **2009**, *10*, 1405−1414.

(12) Egan, W. J.; Merz, K. M.; Baldwin, J. J.; Baldwin, J. J. Prediction of Drug Absorption Using Multivariate Statistics. *J. Med. Chem.* **2000**, *43*, 3867−3877.

(13) Moss, G. P.; Dearden, J. C.; Patel, H.; Cronin, M. T. D. Quantitative structure-permeability relationships (QSPRs) for percutaneous absorption. *Toxicol. in Vitro* **2002**, *16*, 299−317.

(14) Maltarollo, V. G.; Gertrudes, J. C.; Oliveira, P. R.; Honorio, K. M. Applying Machine Learning Techniques for ADME-Tox Prediction: A Review. *Expet Opin. Drug Metabol. Toxicol.* **2015**, *11*, 259−271.

(15) Sharma, A. K.; Srivastava, G. N.; Roy, A.; Sharma, V. K. ToxiM: A Toxicity Prediction Tool for Small Molecules Developed Using Machine Learning and Chemoinformatics Approaches. *Front. Pharmacol.* **2017**, *8*, 880.

(16) Saxena, D.; Sharma, A.; Siddiqui, M. H.; Kumar, R. Blood Brain Barrier Permeability Prediction Using Machine Learning Techniques: An Update. *Curr. Pharm. Biotechnol.* **2019**, *20*, 1163−1171.

(17) Doniger, S.; Hofmann, T.; Yeh, J. Predicting CNS Permeability of Drug Molecules: Comparison of Neural Network and Support Vector Machine Algorithms. *J. Comput. Biol.* **2002**, *9*, 849−864.

(18) Venable, R. M.; Krämer, A.; Pastor, R. W. Molecular Dynamics Simulations of Membrane Permeability. *Chem. Rev.* **2019**, *119*, 5954.

(19) Marrink, S.-J.; Berendsen, H. J. C. Simulation of Water Transport through a Lipid Membrane. *J. Phys. Chem.* **1994**, *98*, 4155−4168.

(20) Comer, J.; Schulten, K.; Chipot, C. Permeability of a Fluid Lipid Bilayer to Short-Chain Alcohols from First Principles. *J. Chem. Theory Comput.* **2017**, *13*, 2523−2532.

(21) Tse, C. H.; Comer, J.; Sang Chu, S. K.; Wang, Y.; Chipot, C. Affordable Membrane Permeability Calculations: Permeation of Short-Chain Alcohols through Pure-Lipid Bilayers and a Mammalian Cell Membrane. *J. Chem. Theory Comput.* **2019**, *15*, 2913−2924.

(22) Bennion, B. J.; Be, N. A.; McNerney, M. W.; Lao, V.; Carlson, E. M.; Valdez, C. A.; Malfatti, M. A.; Enright, H. A.; Nguyen, T. H.; Lightstone, F. C.; Carpenter, T. S. Predicting a Drug's Membrane Permeability: A Computational Model Validated With in Vitro Permeability Assay Data. *J. Phys. Chem. B* **2017**, *121*, 5228−5237.

(23) Yue, Z.; Li, C.; Voth, G. A.; Swanson, J. M. J. Dynamic Protonation Dramatically Affects the Membrane Permeability of Drug-like Molecules. *J. Am. Chem. Soc.* **2019**, *141*, 13421−13433.

(24) Lee, C. T.; Comer, J.; Herndon, C.; Leung, N.; Pavlova, A.; Swift, R. v.; Tung, C.; Rowley, C. N.; Amaro, R. E.; Chipot, C.; Wang, Y.; Gumbart, J. C. Simulation-Based Approaches for Determining Membrane Permeability of Small Compounds. *J. Chem. Inf. Model.* **2016**, *56*, 721−733.

(25) Rogers, J. R.; Geissler, P. L. Breakage of Hydrophobic Contacts Limits the Rate of Passive Lipid Exchange between Membranes. *J. Phys. Chem. B* **2020**, *124*, 5884−5898.

(26) Bolhuis, P. G.; Dellago, C.; Chandler, D. Reaction Coordinates of Biomolecular Isomerization. *Proc. Natl. Acad. Sci.* **2000**, *97*, 5877−5882.

(27) Votapka, L. W.; Lee, C. T.; Amaro, R. E. Two Relations to Estimate Membrane Permeability Using Milestoning. *J. Phys. Chem. B* **2016**, *120*, 8606−8616.

(28) Cardenas, A. E.; Elber, R. Modeling Kinetics and Equilibrium of Membranes with Fields: Milestoning Analysis and Implication to Permeation. *J. Chem. Phys.* **2014**, *141*, 054101.

(29) Dickson, C. J.; Hornak, V.; Pearlstein, R. A.; Duca, J. S. Structure-Kinetic Relationships of Passive Membrane Permeation from Multiscale Modeling. *J. Am. Chem. Soc.* **2017**, *139*, 442−452.

(30) Lipinski, C. A. Lead- and Drug-like Compounds: The Rule-of-Five Revolution. *Drug Discov. Today Technol.* **2004**, *1*, 337−341.
(31) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings 1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3-25. 1. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3−26.
(32) Orbach, E.; Finkelstein, A. The Nonelectrolyte Permeability of Planar Lipid Bilayer Membranes. *J. Gen. Physiol.* **1980**, *75*, 427−436.
(33) Parisio, G.; Stocchero, M.; Ferrarini, A. Passive Membrane Permeability: Beyond the Standard Solubility-Diffusion Model. *J. Chem. Theory Comput.* **2013**, *9*, 5236.
(34) Huber, G. A.; Kim, S. Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. *Biophys. J.* **1996**, *70*, 97−110.
(35) Bhatt, D.; Zhang, B. W.; Zuckerman, D. M. Steady-State Simulations Using Weighted Ensemble Path Sampling. *J. Chem. Phys.* **2010**, *133*, 014110.
(36) Sztain, T.; Ahn, S.-H.; Bogetti, A. T.; Casalino, L.; Goldsmith, J. A.; Seitz, E.; McCool, R. S.; Kearns, F. L.; Acosta-Reyes, F.; Maji, S.; Mashayekhi, G.; McCammon, J. A.; Ourmazd, A.; Frank, J.; McLellan, J. S.; Chong, L. T.; Amaro, R. E. A Glycan Gate Controls Opening of the SARS-CoV-2 Spike Protein. *Nat. Chem.* **2021**, *13*, 963−968.
(37) Dixon, T.; Uyar, A.; Ferguson-Miller, S.; Dickson, A. Membrane-Mediated Ligand Unbinding of the PK-11195 Ligand from TSPO. *Biophys. J.* **2021**, *120*, 158−167.
(38) Saglam, A. S.; Chong, L. T. Protein-protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. *Chem. Sci.* **2019**, *10*, 2360−2372.
(39) Adhikari, U.; Mostofian, B.; Copperman, J.; Subramanian, S. R.; Petersen, A. A.; Zuckerman, D. M. Computational Estimation of Microsecond to Second Atomistic Folding Times. *J. Am. Chem. Soc.* **2019**, *141*, 6519−6526.
(40) Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. The "Weighted Ensemble" Path Sampling Method Is Statistically Exact for a Broad Class of Stochastic Processes and Binning Procedures. *J. Chem. Phys.* **2010**, *132*, 054107.
(41) Aristoff, D. An Ergodic Theorem for the Weighted Ensemble Method. *J. Appl. Probab.* **2019**, *59*, 1.
(42) Hill, T. L. *Free Energy Transduction and Biochemical Cycle Kinetics*; Springer: New YorkNew York, NY, 1989.
(43) DeGrave, A. J.; Bogetti, A. T.; Chong, L. T. The RED Scheme: Rate-Constant Estimation from Pre-Steady State Weighted Ensemble Simulations. *J. Chem. Phys.* **2021**, *154*, 114111.
(44) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. Lessons Learned from Comparing Molecular Dynamics Engines on the SAMPL5 Dataset. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 147−161.
(45) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Hahn, D. F.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Tjanaka, B.; Tresadern, G.; Lucas, X.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P. Development and Benchmarking of Open Force Field v1.0.0-the Parsley Small-Molecule Force Field. *J. Chem. Theory Comput.* **2021**, *17*, 6262−6280.
(46) Gould, I. R.; Skjevik, A. A.; Dickson, C. J.; Madej, B. D.; Walker, R. C. Lipid17: A Comprehensive AMBER Force Field for the Simulation of Zwitterionic and Anionic Lipids. Manuscript in preparation **2018**.
(47) Dickson, C. J.; Madej, B. D.; Skjevik, Å. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 865−879.
(48) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859−1865.
(49) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. CHARMM-GUIMembrane Buildertoward realistic biological membrane simulations. *J. Comput. Chem.* **2014**, *35*, 1997−2004.
(50) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(51) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Comp. Biol.* **2017**, *13*, e1005659.
(52) *OEChem Toolkit*; OpenEye Scientific: Santa Fe, NM 2021.
(53) Walters, W. P.; Green, J.; Weiss, J. R.; Murcko, M. A. What Do Medicinal Chemists Actually Make? A 50-Year Retrospective. *J. Med. Chem.* **2011**, *54*, 6405−6416.
(54) *Omega Toolkit*; OpenEye Scientific: Santa Fe, NM 2021.
(55) Hawkins, P. C. D.; Nicholls, A. Conformer Generation with OMEGA: Learning from the Data Set and the Analysis of Failures. *J. Chem. Inf. Model.* **2012**, *52*, 2919−2936.
(56) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572−584.
(57) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30*, 2157−2164.
(58) *Picto*; OpenEye Scientific: Santa Fe, NM 2021.
(59) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.
(60) Russo, J. D.; Zhang, S.; Leung, J. M. G.; Bogetti, A. T.; Thompson, J. P.; DeGrave, A. J.; Torrillo, P. A.; Pratt, A. J.; Wong, K. F.; Xia, J.; et al. WESTPA 2.0: High-Performance Upgrades for Weighted Ensemble Simulations and Analysis of Longer-Timescale Applications. *J. Chem. Theor. Comput.* **2021**, *18*, 638.
(61) Torrillo, P. A.; Bogetti, A. T.; Chong, L. T. A Minimal, Adaptive Binning Scheme for Weighted Ensemble Simulations. *J. Phys. Chem. A* **2021**, *125*, 1642.
(62) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *Chem. Phys.* **1995**, *103*, 8577−8593.
(63) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(64) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(65) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528−1532.
(66) Suárez, E.; Lettieri, S.; Zwier, M. C.; Stringer, C. A.; Subramanian, S. R.; Chong, L. T.; Zuckerman, D. M. Simultaneous Computation of Dynamical and Equilibrium Information Using a Weighted Ensemble of Trajectories. *J. Chem. Theory Comput.* **2014**, *10*, 2658−2667.
(67) Baker, E. N.; Hubbard, R. E. Hydrogen Bonding in Globular Proteins. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97−179.
(68) Morrison, J. P. Flow-Based Programming. *Proceeding 1st International Workshop on Software Engineering for Parallel and Distributed Systems*; International Thomson Computer Press, 1994; pp 25−29.
(69) Dickson, C. J.; Hornak, V.; Bednarczyk, D.; Duca, J. S. Using Membrane Partitioning Simulations To Predict Permeability of Forty-Nine Drug-Like Molecules. *J. Chem. Inf. Model.* **2019**, *59*, 236−244.
(70) Katt, M. E.; Xu, Z. S.; Gerecht, S.; Searson, P. C. Human Brain Microvascular Endothelial Cells Derived from the BC1 IPS Cell Line

https://doi.org/10.1021/acs.jcim.1c01540
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Exhibit a Blood-Brain Barrier Phenotype. *PLoS One* **2016**, *11*, No. e0152105.
(71) Liu, W.; Okochi, H.; Benet, L. Z.; Zhai, S.-D. Sotalol Permeability in Cultured-Cell, Rat Intestine, and PAMPA System. *Pharm. Res.* **2012**, *29*, 1768−1774.
(72) Russo, J. D.; Copperman, J.; Aristoff, D.; Simpson, G.; Zuckerman, D. M. Unbiased Estimation of Equilibrium, Rates, and Committors from Markov State Model Analysis. **2021**, arXiv:2105.13402.
(73) Russo, J.; Copperman, J. T.; Aristoff, D.; Simpson, G.; Zuckerman, D. M. Iterative Steady State Restarting to Accelerate Weighted Ensemble Convergence. *Biophys. J.* **2022**, *121*, 274a.
(74) Shaw, D. E.; Adams, P. J.; Azaria, A.; Bank, J. A.; Batson, B.; Bell, A.; Bergdorf, M.; Bhatt, J.; Butts, J. A.; Correia, T., et al. Anton 3: Twenty Microseconds of Molecular Dynamics Simulation before Lunch. *Proceeding The International Conference for High Performance Computing, Networking, Storage and Analysis*; ACM, 2021; pp 1−11.

https://doi.org/10.1021/acs.jcim.1c01540
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX