

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

She Zhang<sup>1</sup>, Jeff P. Thompson<sup>1</sup>, Junchao Xia<sup>1</sup>, Anthony T. Bogetti<sup>2</sup>, Forrest York<sup>1</sup>, A. Geoffrey Skillman<sup>1</sup>, Lillian T. Chong<sup>\*,2</sup>, David N. LeBard<sup>\*,1</sup>

\* To whom correspondence should be addressed

<sup>1</sup>OpenEye Scientific, Santa Fe, NM 87508

<sup>2</sup>Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260

## Abstract

Passive permeability of a drug-like molecule is a critical assay 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 for any given molecule. Typically, one of three methods are used to measure membrane permeability: (1) experimental permeation assays acting on either artificial or natural membranes; (2) quantitative structure-permeability relationship (QSPR) models that rely on experimental values of permeability or related pharmacokinetic properties of a range of molecules to infer those for new molecules; (3) estimates 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 (POPC) 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.<sup>1</sup> 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<sup>2</sup> of charged endogenous compounds,<sup>3,4</sup> 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<sup>5-8</sup> 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<sup>9</sup> developed over a century ago,<sup>10</sup> which proportionally relates passive membrane permeability to the oil-water partition coefficient<sup>11</sup>. 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 qualitative 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.<sup>12</sup> QSPR models typically rely on properties like polar surface area, molecular weight, hydrogen bond count, and the octanol-water partition coefficient to statistically infer new permeability coefficients for drug-like compounds.<sup>13</sup> Over the last few years, advanced machine learning techniques have also been applied to predict permeability from either molecular descriptors or fingerprints. Reviews of ML applications to predict permeability exist elsewhere<sup>14</sup> and will not be covered in detail here, yet some examples of ML approaches worth mentioning include models for PAMPA-like assays,<sup>5</sup> models for Caco-2 cell permeability,<sup>15</sup> as well as blood-brain barrier<sup>16</sup> and CNS<sup>17</sup> 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 of membrane permeation. While MD simulations have captured small-molecule permeation processes that occur within a  $\mu$ s,<sup>18</sup> 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 by Marrink and Berendsen<sup>19</sup> 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,<sup>20</sup> free energy perturbation,<sup>21</sup> restrained potential of mean force (PMF) calculations,<sup>22</sup> constant pH simulations<sup>23</sup>, and umbrella sampling<sup>24</sup>). 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.<sup>25,26</sup> Alternatively, a few simulation studies have employed methods that combine short discontinuous trajectories to estimate long-timescale observables (e.g., milestoning<sup>27,28</sup> and Markov State Models<sup>29</sup>). 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 zalcopride and sotalolol, which are substantially more complex, flexible molecules (3 and 6 rotatable bonds, respectively), for detailed analysis of their membrane permeation mechanisms. All three compounds are weakly basic and obey Lipinski’s “rule of five”<sup>30,31</sup> 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 desolvating 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):



Here,  $k_{D \rightarrow A}$  is the forward rate constant for passive diffusion of the molecule from the donor to the acceptor compartment, while  $k_{A \rightarrow D}$  is the reverse rate constant. Assuming 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:

$$\frac{dC_A}{dt} = v_{D \rightarrow A} - v_{A \rightarrow D}, \quad (2)$$

$$\frac{dC_D}{dt} = v_{A \rightarrow D} - v_{D \rightarrow A}, \quad (3)$$

where  $v$ 's are volumetric permeation rates,

$$\begin{aligned} v_{D \rightarrow A} &= k_{D \rightarrow A} C_D, \\ v_{A \rightarrow D} &= k_{A \rightarrow D} C_A. \end{aligned} \quad (4)$$

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

$$v_{D \rightarrow A} = \frac{1}{l_D S} \frac{dm_D}{dt}, \quad (5)$$

where  $dm_D$  is the amount of molecule (in terms of mass or molar mass), originating from compartment D then passing through the bilayer, and moving into the compartment A within time  $dt$ . Here,  $S$  is the surface area of the membrane, and  $l_D$  is the depth of the effective volume of compartment D.

Although the volumetric permeation rates in Eq. 4 provide one route to estimating the kinetics of molecular transport across the membrane, the most common quantification is the 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:

$$J_m = P_m (C_D - C_A). \quad (6)$$

By imposing the homogenous solubility-diffusion (HSD) model<sup>18</sup> 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,<sup>32</sup> with a few notable exceptions that are likely attributed to active transport.

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

$$J_m = J_{D \rightarrow A} - J_{A \rightarrow D}. \quad (7)$$

Both the in- and outflux represent the amount of molecule that pass through a surface area,  $S$ , within a given amount of time. Mathematically this can be expressed as:

$$J_{D \rightarrow A} = \frac{1}{S} \frac{dm_D}{dt}, \quad (8)$$

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

$$\begin{aligned} J_{D \rightarrow A} &= v_{D \rightarrow A} l_D, \\ J_{A \rightarrow D} &= v_{A \rightarrow D} l_A. \end{aligned} \quad (9)$$

Substituting Eq. 4, 7 and 9 into Eq. 6, assuming unequal concentrations in the two compartments, i.e.,  $C_D \neq C_A$ , we arrive at

$$P_m = \frac{k_{D \rightarrow A} l_D C_D - k_{A \rightarrow D} l_A C_A}{C_D - C_A}. \quad (10)$$

Furthermore, under the assumption of a symmetric system, such as the case in the permeation of a small molecule through a neat membrane, one can set  $k_{D \rightarrow A} = k_{A \rightarrow D}$  and  $l_D = l_A$ . Thus, Eq. 10 reduces the permeability coefficient to:

$$P_m = k_{D \rightarrow A} l_D = k_{A \rightarrow D} l_A. \quad (11)$$

In our model, by default,  $l_D$  is set to half the length of the aqueous part of the simulation box along the lipid bilayer normal (preliminary tests suggest that the permeability coefficient is somewhat insensitive to the exact value of  $l_D$ , see [Supplementary Figure S2](#) for details). While Eq. 11 is derived here for a two-state permeation process, the same equation can be derived from a multi-

state 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,  $k_{D \rightarrow A}$  or  $k_{A \rightarrow D}$ , as described by Parisio et al.<sup>33</sup>

### Calculation of permeation rate constants using weighted ensemble simulations

The weighted ensemble path sampling strategy<sup>34,35</sup> enables simulations of processes that are orders of magnitude longer than the simulations themselves.<sup>36–39</sup> This greatly enhanced sampling results from an iterative resampling procedure (at fixed time intervals  $\tau$ ) that replicates trajectories to occupy less-visited regions of configurational space – typically defined by a progress coordinate towards 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.<sup>40,41</sup> 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<sup>35,42</sup>:

$$k_{D \rightarrow A} = \frac{1}{MFPT(D \rightarrow A)} = \frac{f_{D \rightarrow A}^{SS}}{p_D}, \quad (12)$$

For drug membrane permeation,  $f_{D \rightarrow A}^{SS}$  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 reentry (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,  $\langle \hat{f}_{D \rightarrow A} \rangle$ , in a WE simulation provides an estimate of the steady-state flux,

$$k_{D \rightarrow A} = \frac{\langle \hat{f}_{D \rightarrow A} \rangle}{\langle \hat{p}_D \rangle} = \langle \hat{f}_{D \rightarrow A} \rangle. \quad (13)$$

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.<sup>43</sup>

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

$$\begin{aligned} \tilde{p}_D &= \frac{m_D/V_D}{m_{total}/V} = \frac{C_D}{C_0}, \\ \tilde{p}_A &= \frac{m_A/V_A}{m_{total}/V} = \frac{C_A}{C_0}, \end{aligned} \quad (14)$$

where  $V_D = V_A = V$  due to the symmetry of the system.  $m_{total} = m_D + \sum_{i \neq D,A} m_i + m_A$  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,<sup>18</sup> but instead of simply counting the number of crossing events in a simulation, the probabilistic weight of each crossing event is also considered.

## 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<sup>44</sup> was used to generate input files containing Open Force Field Parsley v1.3.1a1 force field<sup>45</sup> parameters for the drug-like molecule, as well as Amber LIPID17<sup>46,47</sup> parameters for the POPC membrane. Next, an initial solvated POPC membrane configuration was generated using the CHARMM-GUI v2.0 with the Membrane Builder input generator module, and 50 POPC molecules per leaflet in a solution of TIP3P water molecules. The solvated membrane was equilibrated using a slightly modified set of CHARMM-GUI scripts, such that the OpenMM 7.5 MD engine could be used with a single

GTX1080 GPU to equilibrate the system for 0.5  $\mu$ s 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<sup>48,49</sup> followed by the generation of a diverse set of conformers using Omega Toolkit 4.1.2.0.<sup>50–52</sup> 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<sup>53</sup> at a density of 1 gm/cm<sup>3</sup>. The 2D chemical structures were visualized using Picto 4.5.3.0<sup>54</sup> and 3D molecular structures including the snapshots and movies shown in this study were rendered using VMD 1.9.4.<sup>55</sup>

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

### Weighted ensemble simulations

WE simulations were run in the following manner using the Python API of the WESTPA 2.0 software package.<sup>56</sup> To maintain non-equilibrium steady state conditions, trajectories that reached a target state of compartment A or D (*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 (MAB) scheme<sup>57</sup> (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 distance between the center of mass of the molecule ( $\mathbf{q}_M$ ) to that of the lipid bilayer ( $\mathbf{q}_L$ ) along the unit vector normal to the bilayer surface,  $\hat{\mathbf{z}}$ ,

$$z' = (\mathbf{q}_M - \mathbf{q}_L) \cdot \hat{\mathbf{z}}. \quad (15)$$



Given a lipid bilayer with a width of  $\sim 40$  Å, a molecule with  $z' < -20$  Å is in compartment D, a molecule within  $-20 \text{ Å} \leq z' \leq 20 \text{ Å}$  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. 15 works well in an ideal setup where there is a single lipid bilayer sandwiched between two aqueous compartments of infinite volumes. However, a correction is needed to account for the periodic boundary conditions imposed by the simulation protocol. This correction can be written as

$$z = \left| z' + \frac{L_z}{2} \right| - \frac{L_z}{2}, \quad (16)$$

where  $L_z$  is the length of the simulation box along  $\hat{z}$ . 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 ( $-20 \text{ Å} < z < 20 \text{ Å}$ ) into 0.5 Å-wide bins. We used 30 dynamic linear bins for all our simulation using the MAB scheme.<sup>57</sup>

*Dynamics propagation.* Dynamics were propagated using the OpenMM 7.5 MD engine in the NPT ensemble. A 1-nm cutoff was used for short-range nonbonded interactions, while the particle mesh Ewald (PME) method<sup>58</sup> was applied for the treatment of both long-range electrostatics and Lennard-Jones interactions. To enable a 2-fs timestep, the SHAKE<sup>59</sup> or SETTLE<sup>60</sup> algorithm was used to constrain the lengths of bonds to hydrogens. The trajectories were processed and analyzed using MDTraj<sup>61</sup>.

*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.<sup>35</sup> In this work, reweighting was performed every 50 WE iterations.

## Auxiliary coordinates

Two-dimensional probability distributions have been generated as a function of the WE progress coordinate  $z$  and “auxiliary” coordinates that are orthogonal to the progress coordinate, including:

- 1) The angle relative to  $\hat{\mathbf{z}}$ , which is defined through the vector product of the unit electric dipole moment of the molecule and  $\hat{\mathbf{z}}$ , and quantifies the relative orientation of the molecule through the membrane.
- 2) The number of hydrophobic contacts, which is defined to be the number of aliphatic atoms of the lipid tails within the 10 Å distance of any hydrophobic atoms of the drug molecule, following Rogers and Geissler in their studies of lipid insertion.<sup>25</sup> The hydrophobic atoms used in this study are highlighted in [Supplementary Figure S5](#).
- 3) The number of hydrogen bonds between the drug and the membrane, which was identified using the Baker-Hubbard definition<sup>62</sup> as implemented in the MDTraj package.<sup>61</sup> 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 degrees and the bond distance is less than 2.5 Å.
- 4) The end-to-end distance of each molecule, which 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 [Supplementary Figure S5](#).

Notably, these auxiliary coordinates do not play any role in progressing the WE simulation and were only calculated for post-analysis.

## Cloud computing in Orion

Scientific computing in OpenEye's Orion cloud platform is patterned after the concept of flow-based programming (FBP).<sup>63</sup> 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". 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 exposes various parameters for system preparation and the WE simulation including the option to turn on/off the MAB scheme for automatic binning<sup>57</sup> or the bin probability reweighting for faster convergence to the steady state<sup>35</sup> ([Supplementary 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

CPU per WE iteration. Permeability simulations are analyzed on-the-fly, and a simulation report describing important features of the reactive trajectory data is generated upon completion ([Supplementary 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, 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 to -3.23) that are in reasonable agreement with the value measured by MCDK-LE (-4.64)<sup>64</sup> and PAMPA ( $-5.02 \pm 0.2$ );<sup>65</sup> [Figure 2B](#) and [Table 1](#)).

Relative to a 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 protocol 3 and 4) and use of the WESS reweighting procedure reduces the required total simulation time by ~60% (by comparing protocol 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. The large lower CI bound for the 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](#)). 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  $\tau$  is the fixed time interval (100 ps) of each WE iteration. 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. 16) 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 (Supplementary Figure S4C). One of the top-weighted, successful trajectories (probabilistic weight:  $6.0 \times 10^{-6}$ ) was extracted for visualization at the atomic level (Figure 2D and Supplementary Movie S1).

**Mechanisms of membrane permeation for tacrine, sotalol, and zacopride.** We evaluated the free energy profile of the molecule along  $z$  by symmetrizing the probability distribution sampled by the WE simulations. 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, i.e., the center of the membrane. ~30 ns was needed for the free energy profile to 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 angle of the unit electric dipole moment to  $\hat{z}$ , 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 degree angle ( $\pm 30$  degree with respect to  $z$ ), with a relatively constrained angle near the center and interface of the membrane, especially for zacopride and sotalol. Tacrine formed a much larger number of hydrophobic contacts (~1000) with the membrane than zacopride (~300) and sotalol (~200), presumably due to its higher number of hydrophobic carbons (Supplementary Figure S5). 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), but it was observed “flipping” as previously predicted as a model for membrane permeation.<sup>29</sup> Interestingly, 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 - 12 Å).

Overall, these 2D probability distributions 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 coordinate 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 occurred at different positions within  $z$ , e.g., 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 [Figure 2D](#), [Figure 4B](#), and [Figure 5B](#) for critical snapshots, and see [Supplementary Movie 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 the distorting the curvature of the outer leaflet of the membrane to bring water towards 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 to 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), 11.7 (zacopride) day period. Although many of the pathways for these events are correlated, 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:  $-4.27 \pm 0.24$ , zacopride:  $-6.35 \pm 0.22$ , sotalol:  $-5.32 \pm 0.22$ ) are in reasonable agreement with experimentally measured values (tacrine:  $-4.64^{64}$  or  $-5.03 \pm 0.2$ ,<sup>65</sup> zacopride:  $-5.23$ ,<sup>64</sup> sotalol:  $-6.02$ ,<sup>64</sup>  $-5.58$  or  $-6.74$ ,<sup>66</sup> see [Figure 6](#) and [Supplementary Figure 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 like MDCK and CaCo-2. These complexities include the presence of multiple lipid species, including 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, e.g., 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.<sup>20</sup> 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<sup>66</sup> 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 deviations were defined in the linear scale and converted to the log scale for visual representation.

As shown by Rogers and Geissler,<sup>25</sup> 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. Moreover, rates computed from free energy methods depend exponentially on the barrier height, which is in turn very sensitive to the choice of the progress coordinate.

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 an Internet connection. 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 full continuous trajectories revealing how molecules cross through the membrane.

## Acknowledgments

The authors would like to thank Anthony Nicholls (OpenEye Scientific), Alan Grossfield (University of Rochester), 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 compute 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.



## Disclosures

S.Z., J.P.T., J.X., A.T.B., F.Y., A.N., 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. The OpenEye Permeability Floe is available on the academic stack of Orion, which is also available upon request.

## Reference

- (1) Li, A. P. Screening for Human ADME/Tox Drug Properties in Drug Discovery. *Drug Discovery Today* **2001**, 6 (7), 357–366. [https://doi.org/10.1016/S1359-6446\(01\)01712-3](https://doi.org/10.1016/S1359-6446(01)01712-3).
- (2) Dobson, P. D.; Kell, D. B. Carrier-Mediated Cellular Uptake of Pharmaceutical Drugs: An Exception or the Rule? *Nature reviews. Drug discovery* **2008**, 7 (3), 205–220. <https://doi.org/10.1038/nrd2438>.
- (3) Goldberg, G. S.; Valiunas, V.; Brink, P. R. Selective Permeability of Gap Junction Channels. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2004**, 1662 (1–2), 96–101. <https://doi.org/10.1016/j.bbamem.2003.11.022>.
- (4) Harris, A. L. Connexin Channel Permeability to Cytoplasmic Molecules. *Progress in Biophysics and Molecular Biology* **2007**, 94 (1–2), 120–143. <https://doi.org/10.1016/j.pbiomolbio.2007.03.011>.
- (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. *Bioorganic & Medicinal Chemistry* **2017**, 25 (3), 1266–1276. <https://doi.org/10.1016/j.bmc.2016.12.049>.
- (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. *Journal of Pharmaceutical Sciences* **1999**, 88 (1), 28–33. <https://doi.org/10.1021/js9803205>.
- (7) Hubatsch, I.; Ragnarsson, E. G. E.; Artursson, P. Determination of Drug Permeability and Prediction of Drug Absorption in Caco-2 Monolayers. *Nature Protocols* **2007**, 2 (9), 2111–2119. <https://doi.org/10.1038/nprot.2007.303>.
- (8) Taillardat-Bertschinger, A.; Carrupt, P.-A.; Barbato, F.; Testa, B. Immobilized Artificial Membrane HPLC in Drug Research. *Journal of Medicinal Chemistry* **2003**, 46 (5), 655–665. <https://doi.org/10.1021/jm020265j>.
- (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. *Molecular Physics* **2019**, 117 (20), 2900–2909. <https://doi.org/10.1080/00268976.2019.1601787>.

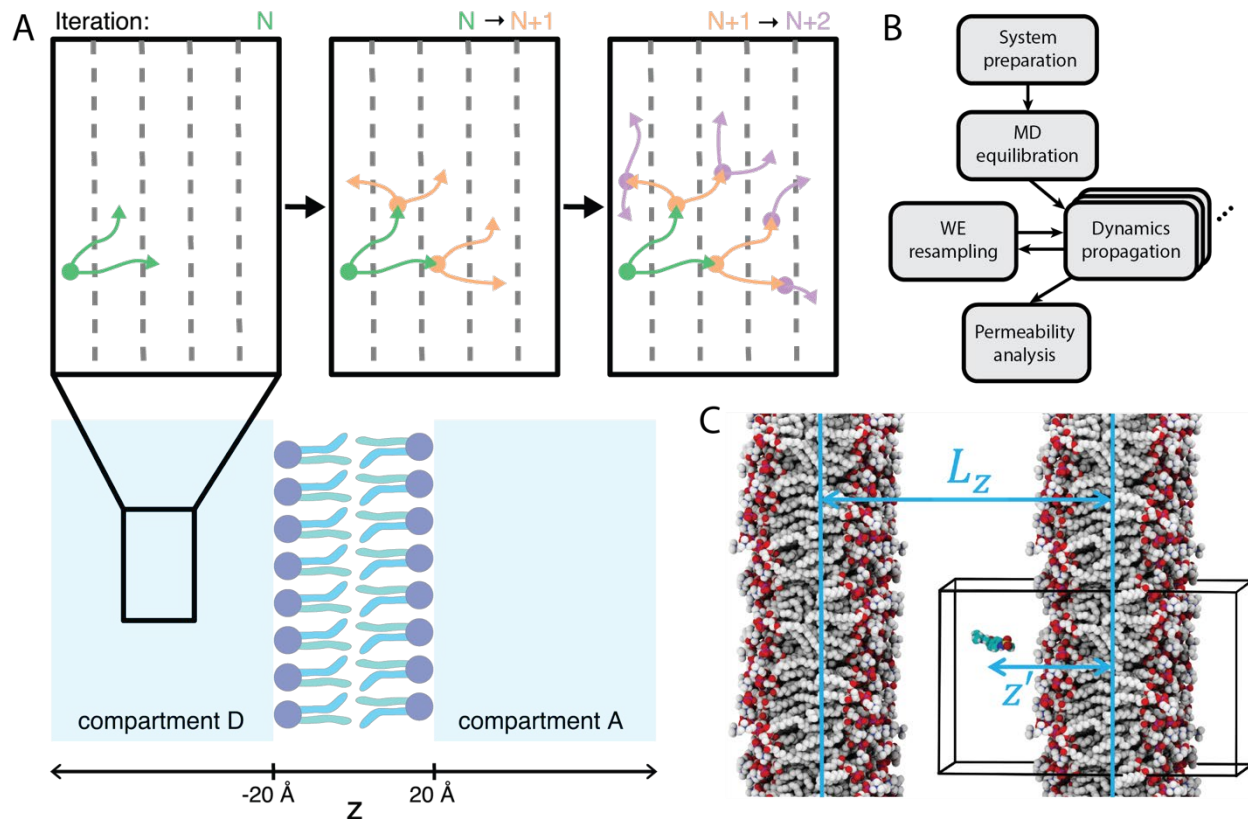
- (11) Missner, A.; Pohl, P. 110 Years of the Meyer-Overton Rule: Predicting Membrane Permeability of Gases and Other Small Compounds. *ChemPhysChem* **2009**, *10* (9–10), 1405–1414. <https://doi.org/10.1002/cphc.200900270>.
- (12) Egan, W. J.; Merz, , Kenneth M.; Baldwin, J. J. Prediction of Drug Absorption Using Multivariate Statistics. *Journal of Medicinal Chemistry* **2000**, *43* (21), 3867–3877. <https://doi.org/10.1021/jm000292e>.
- (13) Moss, G. P.; Dearden, J. C.; Patel, H.; Cronin, M. T. D. Quantitative Structure–Permeability Relationships (QSPRs) for Percutaneous Absorption. *Toxicology in Vitro* **2002**, *16* (3), 299–317. [https://doi.org/10.1016/S0887-2333\(02\)00003-6](https://doi.org/10.1016/S0887-2333(02)00003-6).
- (14) Maltarollo, V. G.; Gertrudes, J. C.; Oliveira, P. R.; Honorio, K. M. Applying Machine Learning Techniques for ADME-Tox Prediction: A Review. *Expert Opinion on Drug Metabolism & Toxicology* **2015**, *11* (2), 259–271. <https://doi.org/10.1517/17425255.2015.980814>.
- (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. *Frontiers in Pharmacology* **2017**, *8*. <https://doi.org/10.3389/fphar.2017.00880>.
- (16) Saxena, D.; Sharma, A.; Siddiqui, M. H.; Kumar, R. Blood Brain Barrier Permeability Prediction Using Machine Learning Techniques: An Update. *Current Pharmaceutical Biotechnology* **2019**, *20* (14), 1163–1171. <https://doi.org/10.2174/1389201020666190821145346>.
- (17) Doniger, S.; Hofmann, T.; Yeh, J. Predicting CNS Permeability of Drug Molecules: Comparison of Neural Network and Support Vector Machine Algorithms. *Journal of Computational Biology* **2002**, *9* (6), 849–864. <https://doi.org/10.1089/10665270260518317>.
- (18) Venable, R. M.; Krämer, A.; Pastor, R. W. Molecular Dynamics Simulations of Membrane Permeability. *Chemical Reviews* **2019**, *119* (9). <https://doi.org/10.1021/acs.chemrev.8b00486>.
- (19) Marrink, S.-J.; Berendsen, H. J. C. Simulation of Water Transport through a Lipid Membrane. *The Journal of Physical Chemistry* **1994**, *98* (15), 4155–4168. <https://doi.org/10.1021/j100066a040>.
- (20) Comer, J.; Schulten, K.; Chipot, C. Permeability of a Fluid Lipid Bilayer to Short-Chain Alcohols from First Principles. *Journal of Chemical Theory and Computation* **2017**, *13* (6), 2523–2532. <https://doi.org/10.1021/acs.jctc.7b00264>.
- (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. *Journal of Chemical Theory and Computation* **2019**, *15* (5), 2913–2924. <https://doi.org/10.1021/acs.jctc.9b00022>.
- (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. *The Journal of Physical Chemistry B* **2017**, *121* (20), 5228–5237. <https://doi.org/10.1021/acs.jpcc.7b02914>.
- (23) Yue, Z.; Li, C.; Voth, G. A.; Swanson, J. M. J. Dynamic Protonation Dramatically Affects the Membrane Permeability of Drug-like Molecules. *Journal of the American Chemical Society* **2019**, *141* (34), 13421–13433. <https://doi.org/10.1021/jacs.9b04387>.

- (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. *Journal of Chemical Information and Modeling* **2016**, *56* (4), 721–733. <https://doi.org/10.1021/acs.jcim.6b00022>.
- (25) Rogers, J. R.; Geissler, P. L. Breakage of Hydrophobic Contacts Limits the Rate of Passive Lipid Exchange between Membranes. *The Journal of Physical Chemistry B* **2020**, *124* (28), 5884–5898. <https://doi.org/10.1021/acs.jpcb.0c04139>.
- (26) Bolhuis, P. G.; Dellago, C.; Chandler, D. Reaction Coordinates of Biomolecular Isomerization. *Proceedings of the National Academy of Sciences* **2000**, *97* (11), 5877–5882. <https://doi.org/10.1073/pnas.100127697>.
- (27) Votapka, L. W.; Lee, C. T.; Amaro, R. E. Two Relations to Estimate Membrane Permeability Using Milestoning. *The Journal of Physical Chemistry B* **2016**, *120* (33), 8606–8616. <https://doi.org/10.1021/acs.jpcb.6b02814>.
- (28) Cardenas, A. E.; Elber, R. Modeling Kinetics and Equilibrium of Membranes with Fields: Milestoning Analysis and Implication to Permeation. *The Journal of Chemical Physics* **2014**, *141* (5), 054101. <https://doi.org/10.1063/1.4891305>.
- (29) Dickson, C. J.; Hornak, V.; Pearlstein, R. A.; Duca, J. S. Structure–Kinetic Relationships of Passive Membrane Permeation from Multiscale Modeling. *Journal of the American Chemical Society* **2017**, *139* (1), 442–452. <https://doi.org/10.1021/jacs.6b11215>.
- (30) Lipinski, C. A. Lead- and Drug-like Compounds: The Rule-of-Five Revolution. *Drug discovery today. Technologies* **2004**, *1* (4), 337–341. <https://doi.org/10.1016/j.ddtec.2004.11.007>.
- (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. *Advanced drug delivery reviews* **2001**, *46* (1–3), 3–26. [https://doi.org/10.1016/s0169-409x\(00\)00129-0](https://doi.org/10.1016/s0169-409x(00)00129-0).
- (32) Orbach, E.; Finkelstein, A. The Nonelectrolyte Permeability of Planar Lipid Bilayer Membranes. *Journal of General Physiology* **1980**, *75* (4), 427–436. <https://doi.org/10.1085/jgp.75.4.427>.
- (33) Parisio, G.; Stocchero, M.; Ferrarini, A. Passive Membrane Permeability: Beyond the Standard Solubility-Diffusion Model. *Journal of Chemical Theory and Computation* **2013**, *9* (12). <https://doi.org/10.1021/ct400690t>.
- (34) Huber, G. A.; Kim, S. Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. *Biophysical Journal* **1996**, *70* (1), 97–110. [https://doi.org/10.1016/S0006-3495\(96\)79552-8](https://doi.org/10.1016/S0006-3495(96)79552-8).
- (35) Bhatt, D.; Zhang, B. W.; Zuckerman, D. M. Steady-State Simulations Using Weighted Ensemble Path Sampling. *The Journal of Chemical Physics* **2010**, *133* (1). <https://doi.org/10.1063/1.3456985>.
- (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. *Nature Chemistry* **2021**, *13* (10), 963–968. <https://doi.org/10.1038/s41557-021-00758-3>.

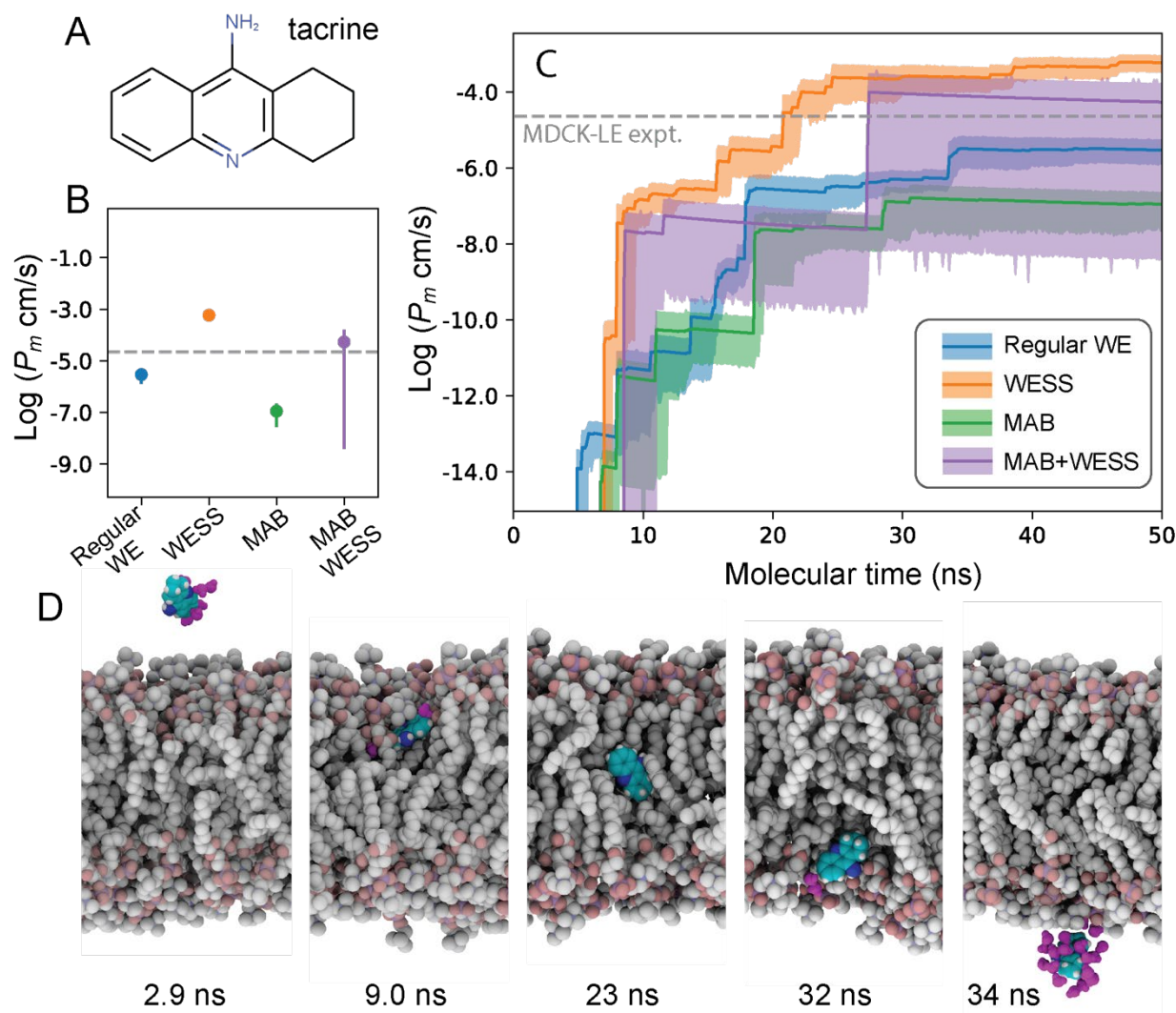
- (37) Dixon, T.; Uyar, A.; Ferguson-Miller, S.; Dickson, A. Membrane-Mediated Ligand Unbinding of the PK-11195 Ligand from TSPO. *Biophysical Journal* **2021**, *120* (1), 158–167. <https://doi.org/10.1016/j.bpj.2020.11.015>.
- (38) Saglam, A. S.; Chong, L. T. Protein–Protein Binding Pathways and Calculations of Rate Constants Using Fully-Continuous, Explicit-Solvent Simulations. *Chemical Science* **2019**, *10* (8), 2360–2372. <https://doi.org/10.1039/C8SC04811H>.
- (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. *Journal of the American Chemical Society* **2019**, *141* (16), 6519–6526. <https://doi.org/10.1021/jacs.8b10735>.
- (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. *The Journal of Chemical Physics* **2010**, *132* (5), 054107. <https://doi.org/10.1063/1.3306345>.
- (41) Aristoff, D. An Ergodic Theorem for the Weighted Ensemble Method. **2019**.
- (42) Hill, T. L. *Free Energy Transduction and Biochemical Cycle Kinetics*; Springer New York: New York, NY, 1989. <https://doi.org/10.1007/978-1-4612-3558-3>.
- (43) DeGrave, A. J.; Bogetti, A. T.; Chong, L. T. The RED Scheme: Rate-Constant Estimation from Pre-Steady State Weighted Ensemble Simulations. *The Journal of Chemical Physics* **2021**, *154* (11), 114111. <https://doi.org/10.1063/5.0041278>.
- (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. *Journal of Computer-Aided Molecular Design* **2017**, *31* (1), 147–161. <https://doi.org/10.1007/s10822-016-9977-1>.
- (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. *Journal of Chemical Theory and Computation* **2021**, *17* (10), 6262–6280. <https://doi.org/10.1021/acs.jctc.1c00571>.
- (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. *Journal of Chemical Theory and Computation* **2014**, *10* (2), 865–879. <https://doi.org/10.1021/ct4010307>.
- (48) OEChem Toolkit. OpenEye Scientific: Santa Fe, NM 2021.
- (49) Walters, W. P.; Green, J.; Weiss, J. R.; Murcko, M. A. What Do Medicinal Chemists Actually Make? A 50-Year Retrospective. *Journal of Medicinal Chemistry* **2011**, *54* (19), 6405–6416. <https://doi.org/10.1021/jm200504p>.
- (50) Omega Toolkit. OpenEye Scientific: Santa Fe, NM 2021.
- (51) Hawkins, P. C. D.; Nicholls, A. Conformer Generation with OMEGA: Learning from the Data Set and the Analysis of Failures. *Journal of Chemical Information and Modeling* **2012**, *52* (11), 2919–2936. <https://doi.org/10.1021/ci300314k>.
- (52) 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. *Journal of Chemical Information and Modeling* **2010**, 50 (4), 572–584.  
<https://doi.org/10.1021/ci100031x>.
- (53) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *Journal of Computational Chemistry* **2009**, 30 (13), 2157–2164. <https://doi.org/10.1002/jcc.21224>.
  - (54) Picto. OpenEye Scientific: Santa Fe, NM 2021.
  - (55) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, 14 (1), 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
  - (56) Russo, J.; Zhang, S.; Leung, J.; Bogetti, A.; Thompson, J.; DeGrave, A.; Torrillo, P.; Pratt, A.; Wong, K.; Xia, J.; others. WESTPA 2.0: High-Performance Upgrades for Weighted Ensemble Simulations and Analysis of Longer-Timescale Applications. *bioRxiv* **2021**. <https://doi.org/10.1101/2021.12.05.471280>.
  - (57) Torrillo, P. A.; Bogetti, A. T.; Chong, L. T. A Minimal, Adaptive Binning Scheme for Weighted Ensemble Simulations. *The Journal of Physical Chemistry A* **2021**, 125 (7). <https://doi.org/10.1021/acs.jpca.0c10724>.
  - (58) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *The Journal of Chemical Physics* **1995**, 103 (19), 8577–8593. <https://doi.org/10.1063/1.470117>.
  - (59) 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. *Journal of computational physics* **1977**, 23 (3), 327–341.
  - (60) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *Journal of Computational Chemistry* **1992**, 13 (8), 952–962. <https://doi.org/10.1002/jcc.540130805>.
  - (61) 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. *Biophysical Journal* **2015**, 109 (8), 1528–1532. <https://doi.org/10.1016/j.bpj.2015.08.015>.
  - (62) Baker, E. N.; Hubbard, R. E. Hydrogen Bonding in Globular Proteins. *Progress in Biophysics and Molecular Biology* **1984**, 44 (2), 97–179. [https://doi.org/10.1016/0079-6107\(84\)90007-5](https://doi.org/10.1016/0079-6107(84)90007-5).
  - (63) Morrison, J. P. Flow-Based Programming. In *Proc. 1st International Workshop on Software Engineering for Parallel and Distributed Systems*; 1994; pp 25–29.
  - (64) Dickson, C. J.; Hornak, V.; Bednarczyk, D.; Duca, J. S. Using Membrane Partitioning Simulations To Predict Permeability of Forty-Nine Drug-Like Molecules. *Journal of Chemical Information and Modeling* **2019**, 59 (1), 236–244. <https://doi.org/10.1021/acs.jcim.8b00744>.
  - (65) Katt, M. E.; Xu, Z. S.; Gerecht, S.; Searson, P. C. Human Brain Microvascular Endothelial Cells Derived from the BC1 IPS Cell Line Exhibit a Blood-Brain Barrier Phenotype. *PLOS ONE* **2016**, 11 (4), e0152105. <https://doi.org/10.1371/journal.pone.0152105>.
  - (66) Liu, W.; Okochi, H.; Benet, L. Z.; Zhai, S.-D. Sotalol Permeability in Cultured-Cell, Rat Intestine, and PAMPA System. *Pharmaceutical Research* **2012**, 29 (7), 1768–1774. <https://doi.org/10.1007/s11095-012-0699-3>.

- (67) Shaw, D. E.; Adams, P. J.; Azaria, A.; Bank, J. A.; Batson, B.; Bell, A.; Bergdorf, M.; Bhatt, J.; Butts, J. A.; Correia, T.; others. Anton 3: Twenty Microseconds of Molecular Dynamics Simulation before Lunch. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*; 2021; pp 1–11.

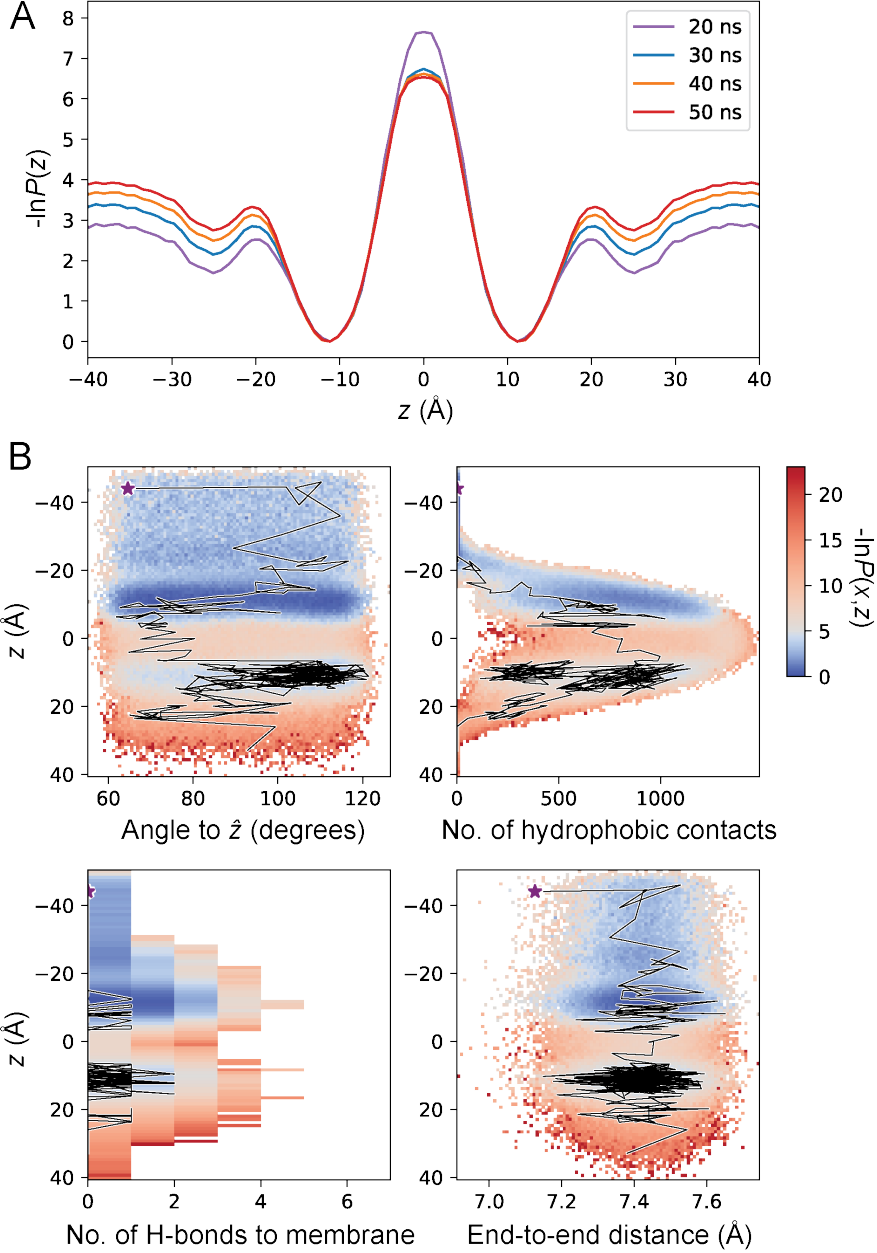


**Figure 1. Basic weighted ensemble protocol and system setup.** **A)** Illustration of the WE protocol for a membrane permeability simulation in which a one-dimensional progress coordinate  $z$  (Eq. 16) is divided into bins and iterative rounds of dynamics propagation and a resampling procedure are 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 2 trajectories per bin. **B)** Simulation workflow used by the Permeability Floe to directly calculate permeability coefficients, including one round of WE resampling (using WESTPA) and dynamics propagation for each WE iteration (see [Supplementary 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*).

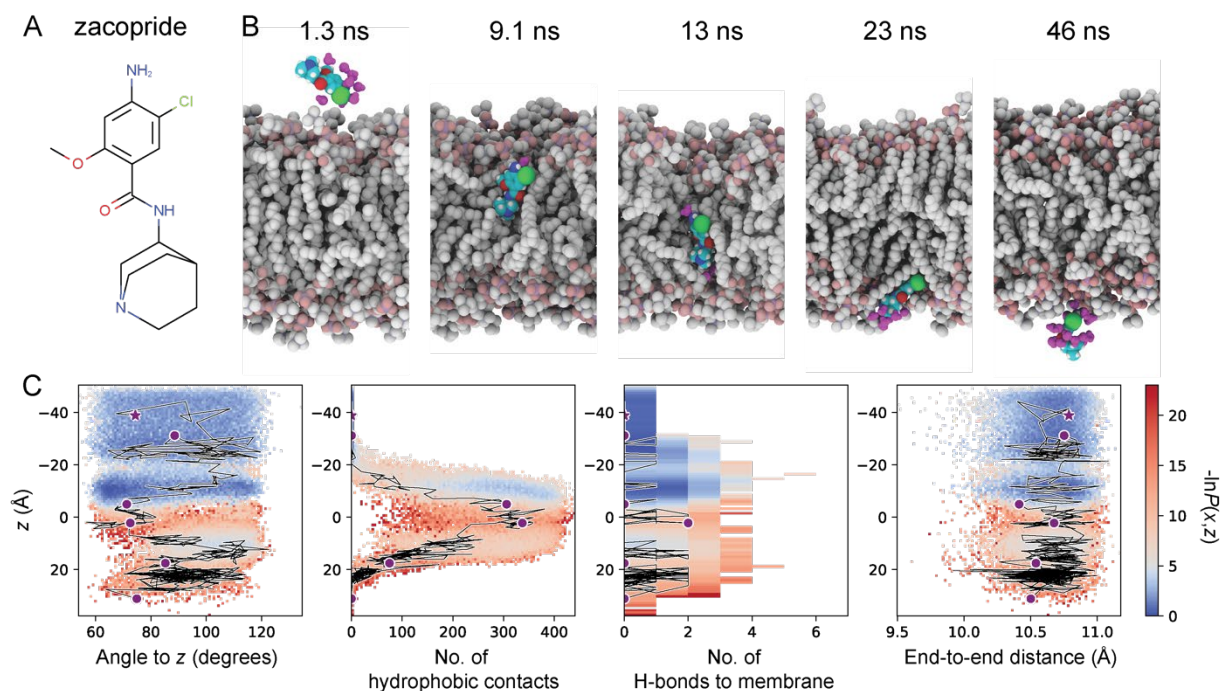


**Figure 2. Estimated permeability coefficients of tacrine.** **A)** Chemical structure of tacrine. **B)** Average (log) permeability coefficients (with 95% confidence intervals) calculated from each of the four WE simulations with different protocols (see legend in panel C). The gray dashed line indicates the experimentally observed value. **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<sup>64</sup>. The molecular time is defined as  $N\tau$ , where  $N$  is the number of WE iterations and  $\tau$  is the fixed time interval (100 ps) of each WE iteration. **D)** The 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*.

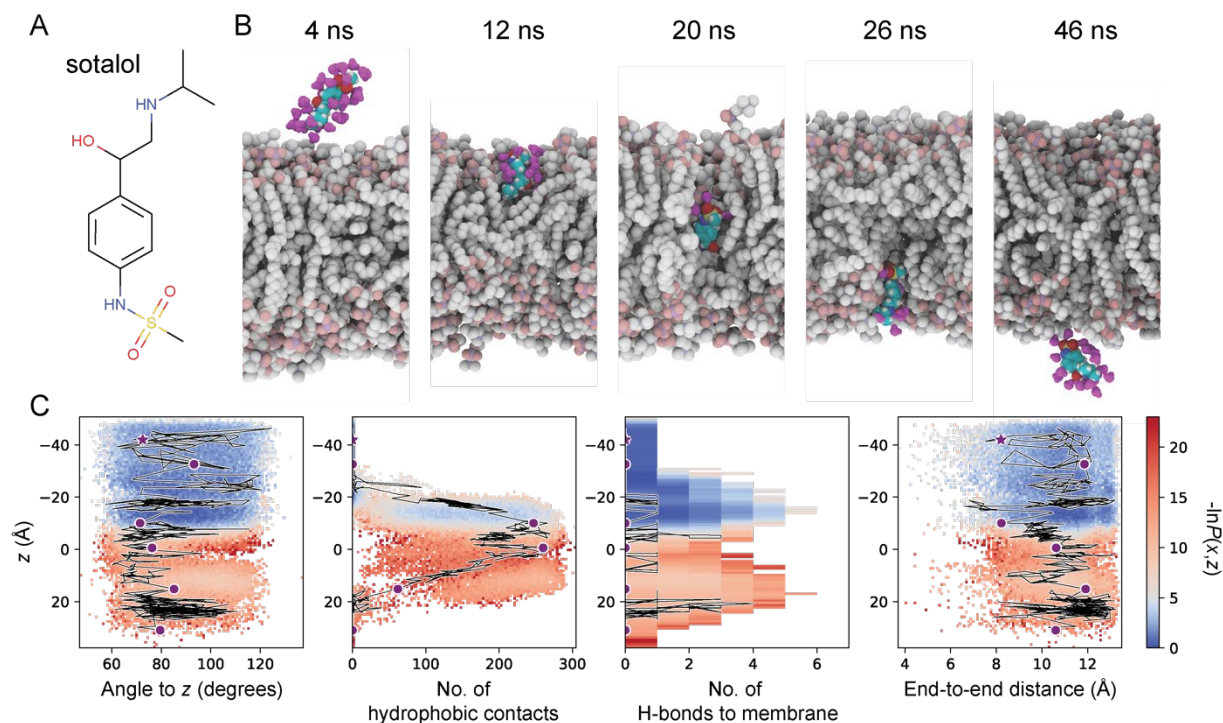




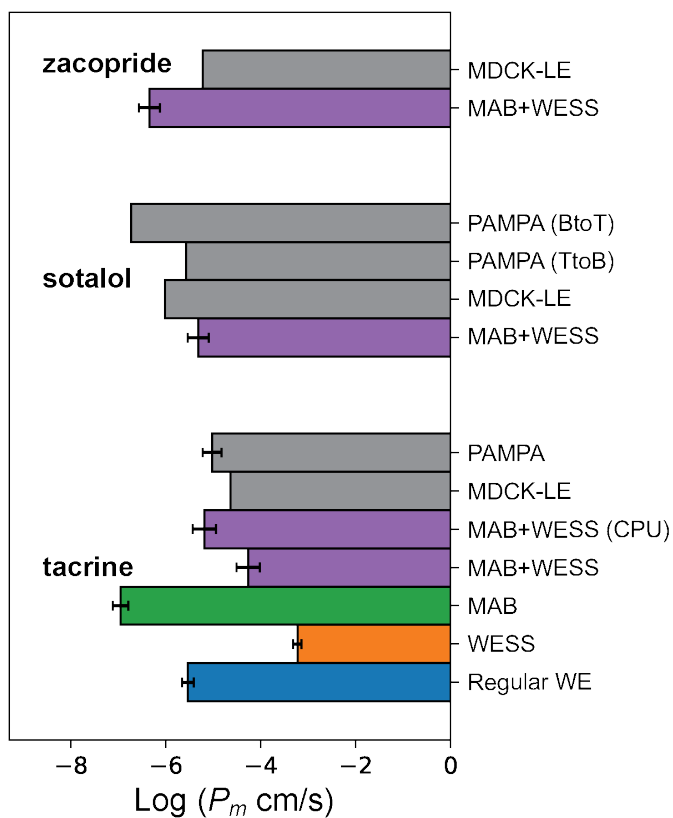
**Figure 3. Free energy profile along the lipid normal ( $z$ ) for tacrine using the reweighting WE protocol (WESS).** **A)** Free energy profile of tacrine along the bilayer normal,  $z$ . The probability distribution along  $z$ ,  $P(z)$ , was extracted from the simulation at different molecular times and symmetrized to account for the recycling condition imposed by the WE steady-state protocol. **B)** The 2D probability distributions (in units of  $k_B T$ ) of  $z$  and the angle of 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:  $< 5k_B T$ , red:  $> 5k_B T$ ). The *black line* represents the top weighted trajectory (probabilistic weight:  $6.0 \times 10^{-6}$ ), and the *purple star* indicates the start location. For all 2D distributions the probabilities are left unsymmetrized across the membrane.



**Figure 4. Mechanistic analysis of zacopride permeation.** **A)** Chemical structure of zacopride. **B)** The snapshots of the zacopride molecule (cyan) passing through the lipid bilayer (gray) at selected molecular times (see legend in Figure 2D). **C)** The 2D probability distributions (in units of  $k_B T$ ) of the bilayer normal,  $z$ , 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:  $< 5k_B T$ , red:  $> 5k_B T$ ). The black line represents the top weighted trajectory (probabilistic weight:  $2.2 \times 10^{-5}$ ), with purple dots indicating the location of the snapshots in panel B and a purple star indicating the start location.



**Figure 5. Mechanistic analysis of sotalol permeation.** **A)** Chemical structure of sotalol. **B)** The snapshots of the sotalol molecule (*cyan*) passing through the lipid bilayer (*gray*) at selected molecular times (see legend in [Figure 2D](#)). **C)** The 2D probability distributions (in units of  $k_B T$ ) of the bilayer normal,  $z$ , and 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:  $< 5k_B T$ , red:  $> 5k_B T$ ). The *black line* represents the top weighted trajectory (probabilistic weight:  $9.3 \times 10^{-6}$ ), with *purple dots* indicating the location of the snapshots in panel B and a *purple star* indicating the start location.



**Figure 6. Membrane permeabilities ( $\text{Log } P_m$ ) calculated using various WE simulation protocols for tacrine, sotalol, and zacopride.** Uncertainties represent standard deviations, which are evaluated as  $\frac{1}{4}$  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),<sup>64</sup> PAMPA for tacrine: Katt et al. (2016),<sup>65</sup> PAMPA for sotalol: Liu et al. (2012)<sup>66</sup>]. See also [Supplementary Table S1](#) in the Supporting Information.

**Table 1. Efficiencies of different WE protocols in predicting the membrane permeability (Log  $P_m$ ) of tacrine.**

Compound	MAB	WESS	Platform	Total simulation time ( $\mu$ s)	Wall clock time (Orion days)	Predicted log ( $P_m$ cm/s)
Tacrine			GPU	23.0	15.5	$-5.54 \pm 0.13$
		x	GPU	23.0	16.2	$-3.23 \pm 0.09$
	x		GPU	8.2	7.1	$-6.96 \pm 0.16$
	x	x	GPU	8.1	7.9	$-4.27 \pm 0.24$
	x	x	CPU	5.6	11.4	$-5.20 \pm 0.28$

**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.** Wall-clock times are reported for the Orion cloud-computing platform on Amazon Web Services and the Anton3 special-purpose MD supercomputer, and 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.

Compound (no. of atoms)	Predicated MFPT (ms)	Weighted Ensemble MD				Standard MD*	
		$P_m$ estimation		A single crossing event		A single crossing event	
		Total simulation time ( $\mu$ s)	Wall clock time (Orion days)	Simulation time (ns)	Wall clock time (Orion days)	Wall clock time (Orion years <sup>a</sup> )	Wall clock time (Anton3 years <sup>b</sup> )
Tacrine <sup>c</sup> (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

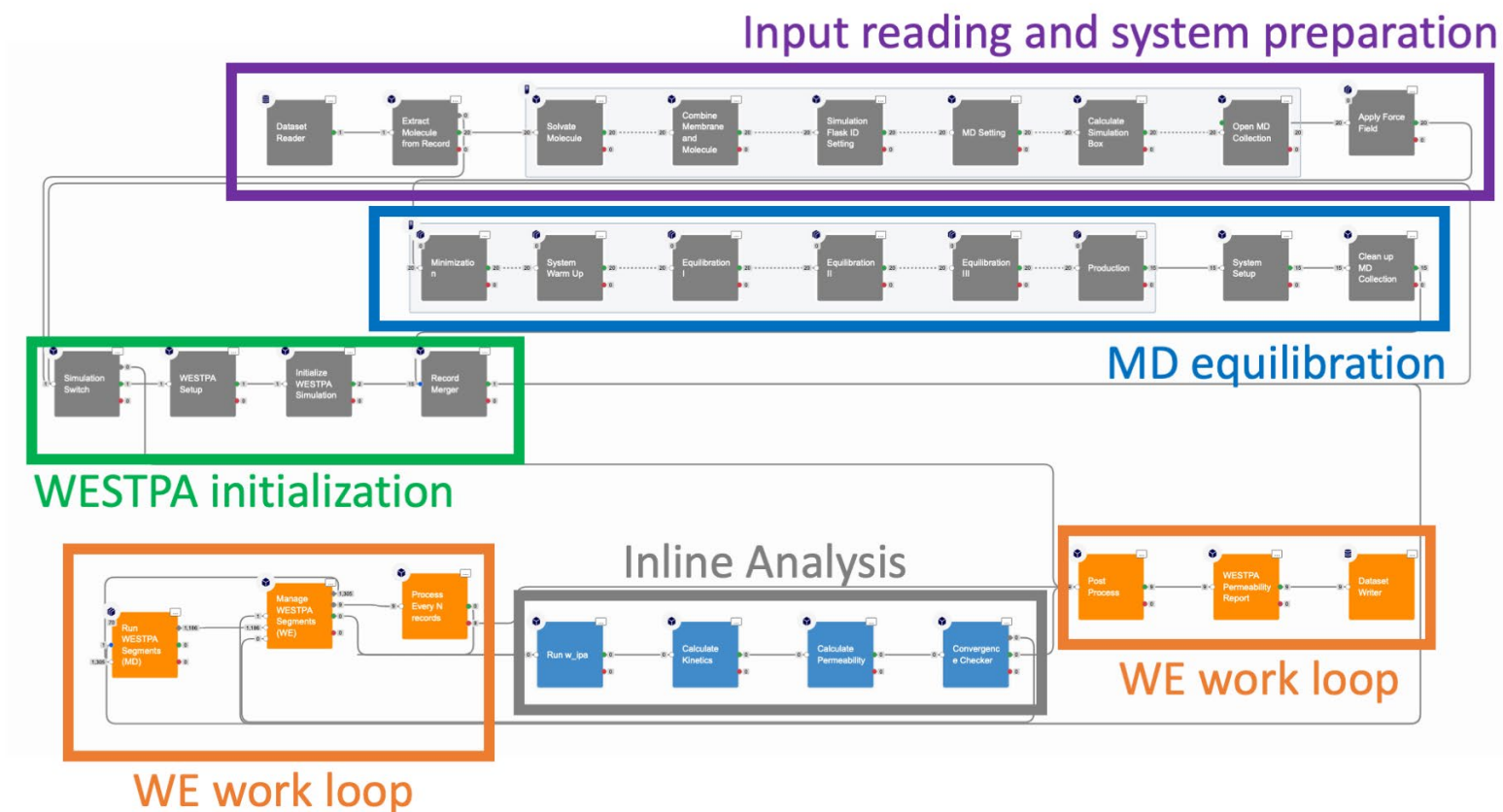
\* Expected times based on the MFPTs predicted by the WE simulations.

<sup>a</sup> Using 100 GPUs at a time on Orion at the speed of 8600 ns per day.

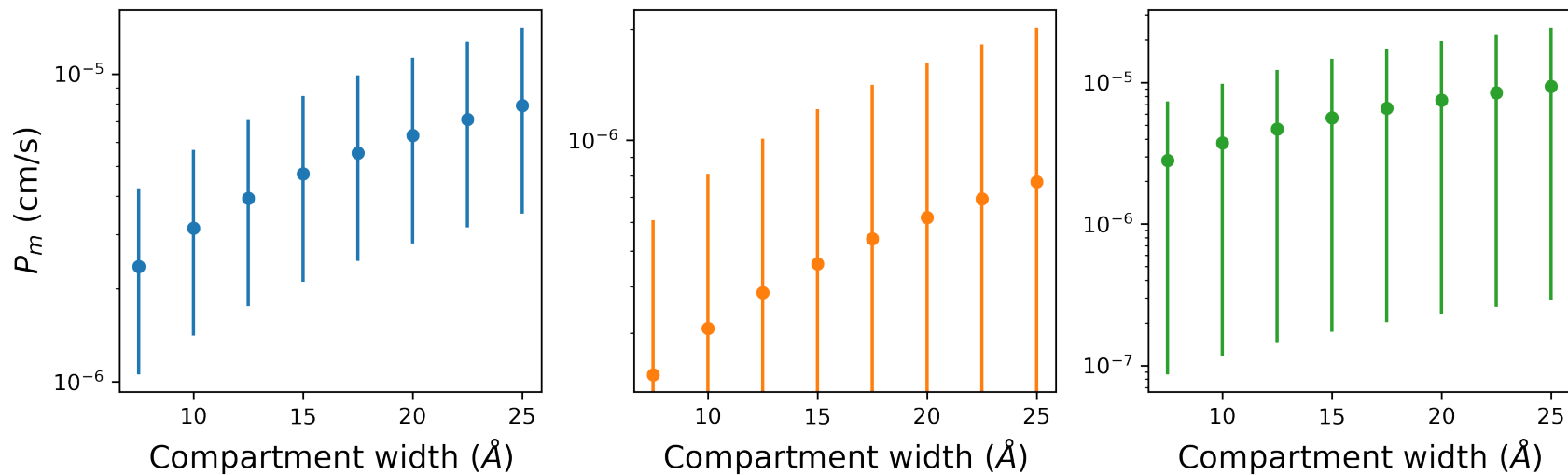
<sup>b</sup> Assuming 200 ms per day for a similar sized system (~33k atoms), as reported using the 64-node Anton3 performance data for DHFR and ApoA1<sup>67</sup> from D. E. Shaw Research.

<sup>c</sup> Data from the tacrine run using MAB and WESS.

## Supplementary Information



**Supplementary Figure S1.** Structural layout of the OpenEye Permeability Floe, which can be broken into 5 logical phases: (1) Preparation of input the permeate system (purple); (2) Equilibration using a standard MD protocol of the aqueous ligand/membrane system (blue); (3) Initialization of the WESTPA toolkit (green); (4) Loop of segment trajectory splitting/recycling events using the WE algorithm, and output trajectory storage (orange); (5) Inline analysis of the permeability simulation for kinetics and convergence criteria (gray).



**Supplementary Figure S2.** The estimated permeability as a function of the compartment width. The upper limit, 25 Å, is the full length of the water compartment in our simulation ( $(L_z - 40\text{Å})/2$ ). The lower limit, 7.5 Å, roughly corresponds to the thickness of the water layer near the membrane surface.



**Permeability - Run Permeability Simulation (GPU) (Permeability)**  
 A floe to calculate the passive membrane permeability of a drug-like molecule on GPUs

✔ **Job Properties**

Name:

Email me when this job completes: ☒ Yes ☐ No

Output path:    
The folder where this job's output will be saved

✔ **Job Cost Limits** Email: not set Terminate: not set >

✖ **Promoted Parameters**

✖ **Inputs**   
 Required input parameters  
 Input Dataset:    
The dataset(s) to read records from  
 \* Value is required

✖ **Outputs**   
 Required output parameters  
 Output Dataset:    
Output dataset to write to  
 \* Value is required

✔ **System Preparation Parameters** >

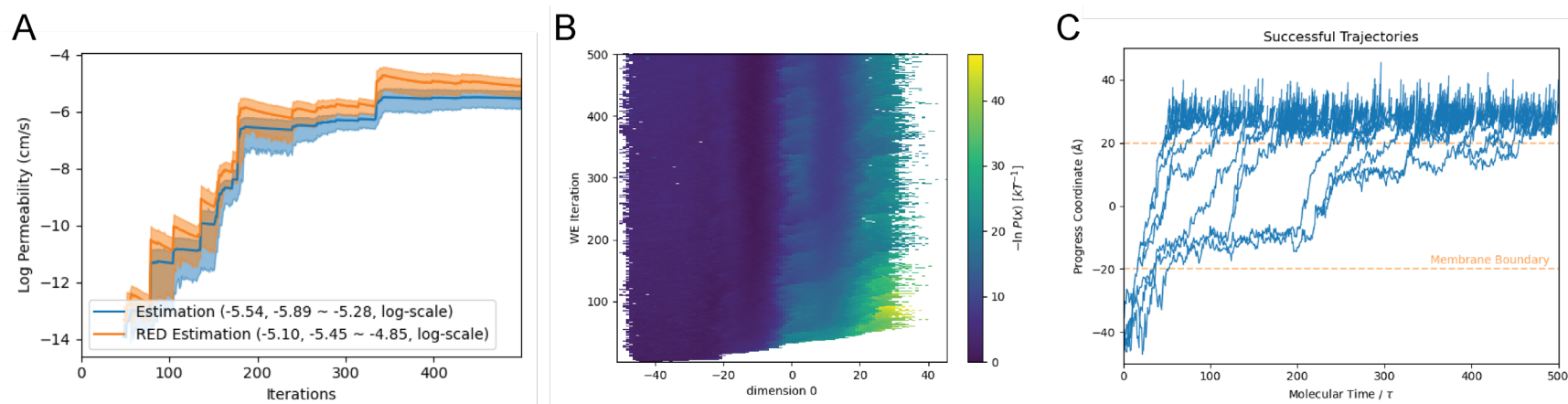
✔ **Weighted Ensemble Parameters**   
 Number Of Basis States:    
A set of Omega conformers used as input to permeability  
 MAB: ☒ On ☐ Off   
Whether enable the Minimal, Adaptive Binning (MAB) scheme  
 Iteration Interval (Tau):    
Length of each WESTPA iteration in picoseconds  
 Iterations:    
Number of iterations for the WESTPA simulation  
 Restart Simulation: ☐ Off   
Restart (instead of continue) the simulation if it is unfinished  
 Reweighting: ☒ On ☐ Off   
Whether or not reweighting of the walkers should occur to reduce the relaxation time in steady-state simulations.  
 Convergence Detection: ☐ Off   
Whether or not automatically detect convergence of the simulation.

✔ **Advanced Weighted Ensemble Parameters** >

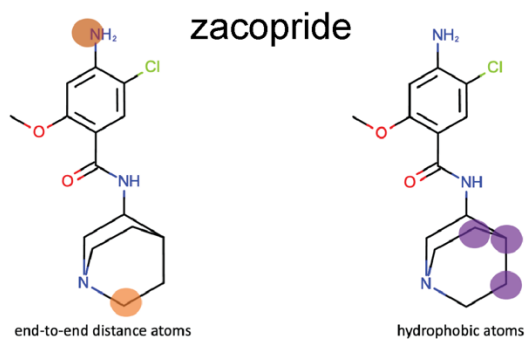
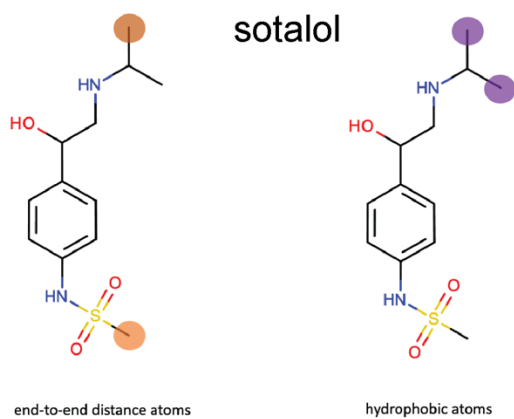
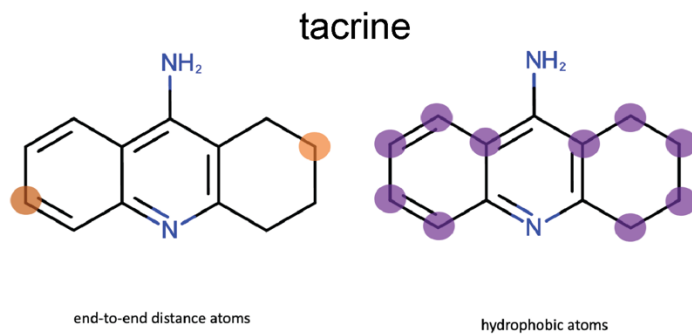
✔ **Miscellaneous Parameters** >

Show cube parameters: ☐ No ☒ Yes

**Supplementary Figure S3.** The Floe setup GUI of the OpenEye Permeability Floe. All the parameters have a brief description of their function and are grouped with other parameters of the same component (e.g., system preparation, WE simulation, etc.) for easier navigation on the user's end.



**Supplementary Figure S4.** Simulation report figures generated by OpenEye Permeability Floe. **A)** The time (iteration=molecular time/ $\tau$ ) evolution of the permeability estimate (*blue*) and that estimated using RED (*orange*). The curves represent the mean estimates, and the shaded areas represent 95% CI. **B)** The time evolution of the probability distribution (in units of  $k_B T$ ). **C)** All the recycled trajectories in the regular WE with fixed binning scheme run represented by the progress coordinate,  $z$  (y-axis), versus the molecular time in terms of the number of iterations (x-axis).



**Supplementary Figure S5.** Chemical structures of tacrine, sotalol, and zacopride. Atoms that were chosen to calculate the end-to-end distances were marked by orange circles, and the hydrophobic carbons were marked by purple circles.

**Supplementary Table S1. Predicted and experimentally determined permeabilities.**

<b>Compound</b>	<b>MAB Adaptive binning</b>	<b>WESS Reweight- ing</b>	<b>Platform</b>	<b>Predicted log (<math>P_m</math> cm/s)</b>	<b>Expt. log (<math>P_m</math> cm/s)</b>
Tacrine			GPU	$-5.54 \pm 0.13$	-4.64, -5.03 $\pm$ 0.2,
		x	GPU	$-3.23 \pm 0.09$	
	x		GPU	$-6.96 \pm 0.16$	
	x	x	GPU	$-4.27 \pm 0.24$	
	x	x	CPU	$-5.20 \pm 0.28$	
Sotalol	x	x	GPU	$-5.32 \pm 0.22$	-6.02, -5.58 (TtoB), -6.74 (BtoT)
Zacopride	x	x	GPU	$-6.35 \pm 0.22$	-5.23