Biophysical Journal, Volume *122*

Supplemental information

Path sampling with memory reduction and replica exchange to reach long permeation timescales

Wouter Vervust, Daniel T. Zhang, Titus S. van Erp, and An Ghysels

SUPPLEMENTARY INFORMATION

Path sampling with memory reduction and replica exchange to reach long permeation timescales

Wouter Vervust,[1] Daniel T. Zhang,[2] Titus S. van Erp,[2] and An Ghysels[1]

[1)] *IBiTech - BioMMedA research group, Faculty of Engineering and Architecture, Ghent University, Corneel Heymanslaan 10, Block B (entrance 36), 9000 Gent, Belgium*

[2)] *Department of Chemistry, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway*

I. THE 2D MAZE

A. Construction of the maze

The maze is first drawn as a pixel map, as shown in Fig. S1a. Maze pixels are labeled as a horizontal wall, a vertical wall, or as both horizontal and vertical walls thus creating a corner.

The walls of the maze generate repulsive forces acting on the Langevin particle. Consider for instance the vertical wall in Fig. S1b with wall thickness $2d_w$. The distance d is defined as the distance from the Langevin particle to the center of the wall. When d is large, beyond a cutoff value D, the particle experiences no interaction with the wall. When the particle approaches the wall and $d < D$, the wall exerts a repulsive force on the particle. On top of a wall when $d < d_w$, the Langevin particle has a constant potential energy and does not experience any force either. The force $F_{\text{wall}}(d)$ scales as a Gaussian function and continuously approaches 0 at $d = d_w$. For $d_w < d \leq D$, the force is given by

$$F_{\text{wall}}(d) = \frac{a}{c^2}(d - d_w)e^{-d^2/(2c^2)}, \tag{1}$$

while the force exerted by the wall is 0 elsewhere. The parameter c is a measure for the steepness of the wall's slope, and a is a measure for the height of the wall. The value of the cutoff distance D should be taken large enough such that the force becomes negligible at $d = D$. In Fig. S1a, black pixels represent very high walls with $a = 500$, while the red pixels represent a lower wall with $a = 25$. The steepness parameter was set to $c = \delta$, the wall thickness to $2d_w = \delta$, and the cutoff value $D = 4\delta$, where $\delta = 1/59$ is the pixel width in the maze design with 59×59 pixels covering an area of 1 by 1.

When the particle is within the cutoff D of two or more walls, the particle only feels the superposition of the closest horizontal and closest vertical wall. Consequently, the potential energy function remains continuous and continuously differentiable in the vicinity of corners, which are characterized as both a vertical and horizontal wall.

In our implementation of the maze potential, the cutoff D is exploited to reduce the CPU-time of the energy and force calculation. Assume the the Langevin particle is positioned at (x_p, y_p). First, the discrete pixel indices (i_p, j_p) of the particle are determined. Then, the detection area is set to a square grid of $(2n_D + 1) \times (2n_D + 1)$ pixels centered around the particle's pixel (i_p, j_p), with $n_D = D/\delta = 4$ in this paper. Only the walls in these

$(2n_D + 1)^2$ pixels are examined to determine the interaction with the particle, which keeps the simulation time-efficient even for large mazes.

The Langevin particle dynamics were integrated with the PyRETIS internal engine. In a Lennard-Jones type of unit system, based on argon[1], the following parameters were used: particle mass $m = 1$, temperature $T = 0.07$, friction coefficient $\gamma = 25$, and integration timestep $\Delta t = 0.01$. The interfaces are placed at $\lambda_A = 0.2$, 0.325, 0.55, 0.69, 0.75, and $0.90 = \lambda_b$.

For completeness, we report the two other factors in Eq. 7 of the main text to be $\xi = 0.48$ ($\pm 1.3\%$) and $\tau_{\mathrm{ref},[0^-\prime]}/\Delta z = 4.86$ ($\pm 0.7\%$).

The simulation input files, together with the maze potential code, can be found on github: WouterWV/pathsampling_toymodels.

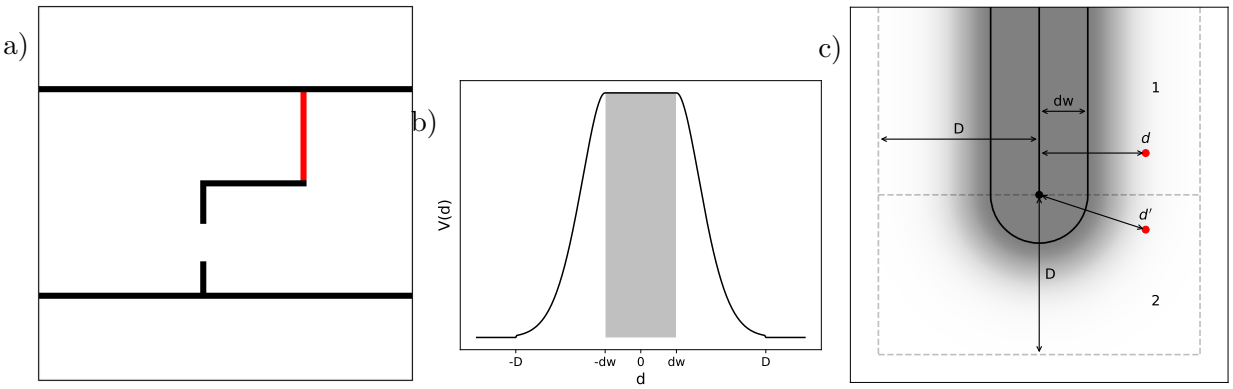

FIG. S1: (a) Pixel map (59 by 59 pixels) used to construct the maze potential in an area of 1 by 1. The red pixels form a lower energetic barrier in the upper channel, while the black pixels are high walls that are nearly impenetrable. (b) Wall potential corresponding to force in Eq. 1. Particles within a distance D of the wall center feel a force exerted by the wall. The top of the wall (distance $< d_w$) is a flat potential. (c) Heat-map of the potential of a vertical wall. A horizontal force is felt in the upper area (area 1). A particle (red) with distance d from the wall center is drawn. A radial force is felt in the lower area (area 2). The magnitude of the force in area 2 is determined by the distance d' to the 'wall center edge'. Particles do not feel a force beyond the cutoff D.

B. 2D histograms of (RE)PPTIS path ensembles

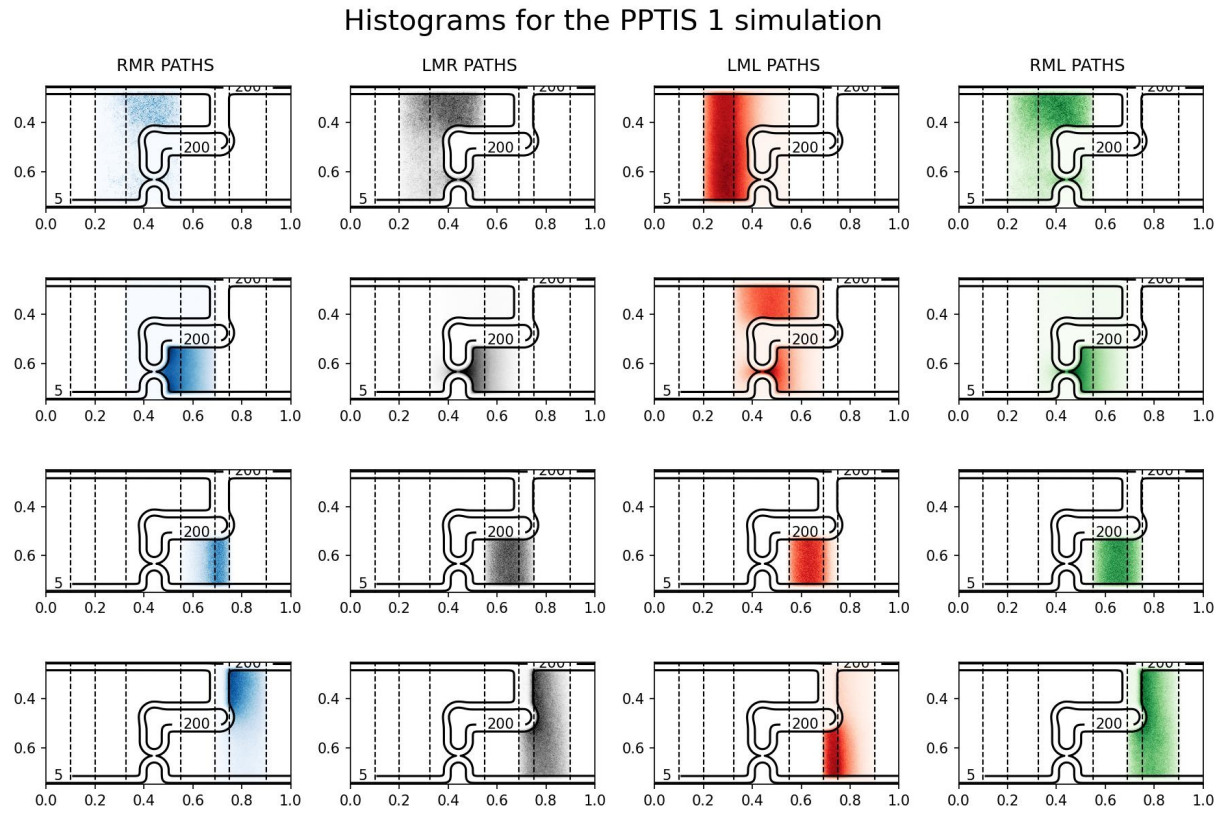

FIG. S2: Visualization of the different path types in each ensemble, for the PPTIS 1 simulation. For each path type, and for each ensemble, the phasepoints of all the trajectories were histogrammed and visualized using a heatmap. The rows from top to bottom correspond with ensembles $[1^{\pm}]$, $[2^{\pm}]$, $[3^{\pm}]$ and $[4^{\pm}]$, and the columns left to right correspond with path types RMR, LMR, LML and RML. The dashed vertical lines in each plot represent the interfaces $(\lambda_{-1}, \lambda_A, \ldots, \lambda_B)$

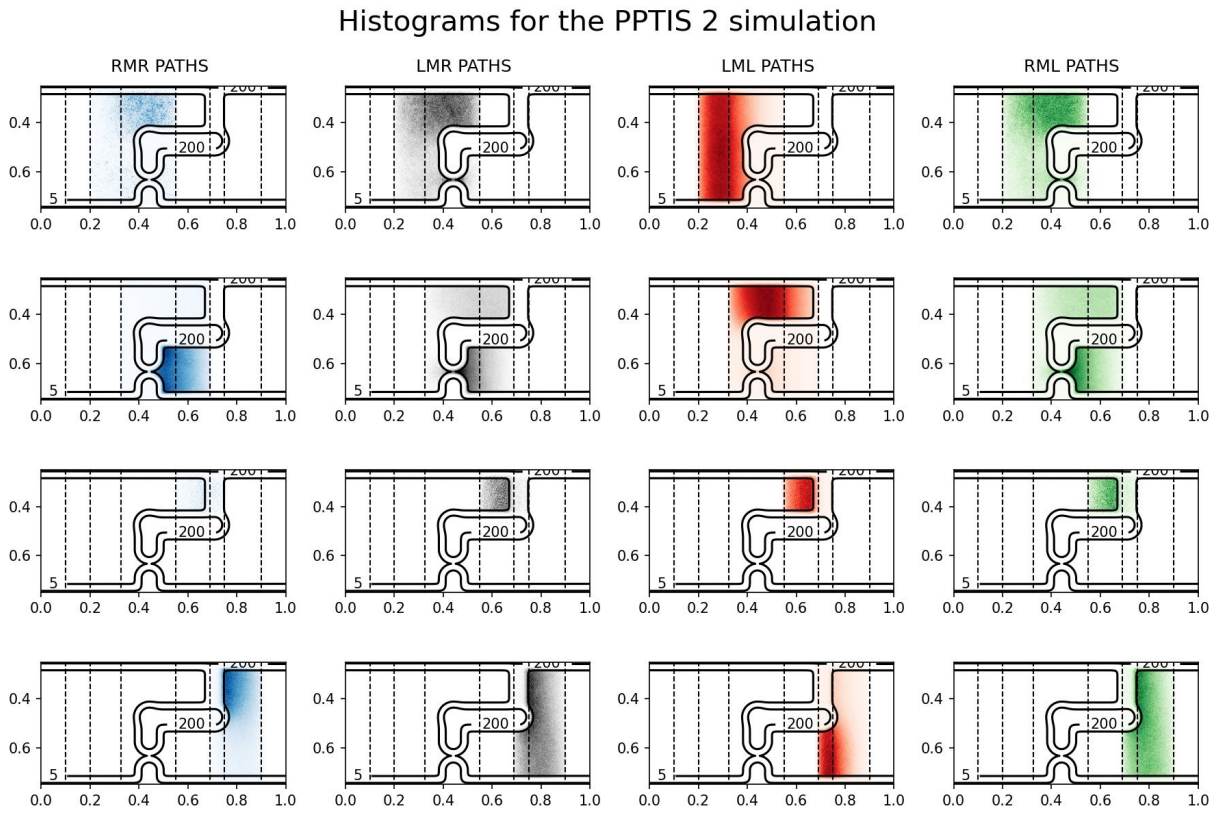

FIG. S3: Same as Fig. S2, for the PPTIS 2 simulation.

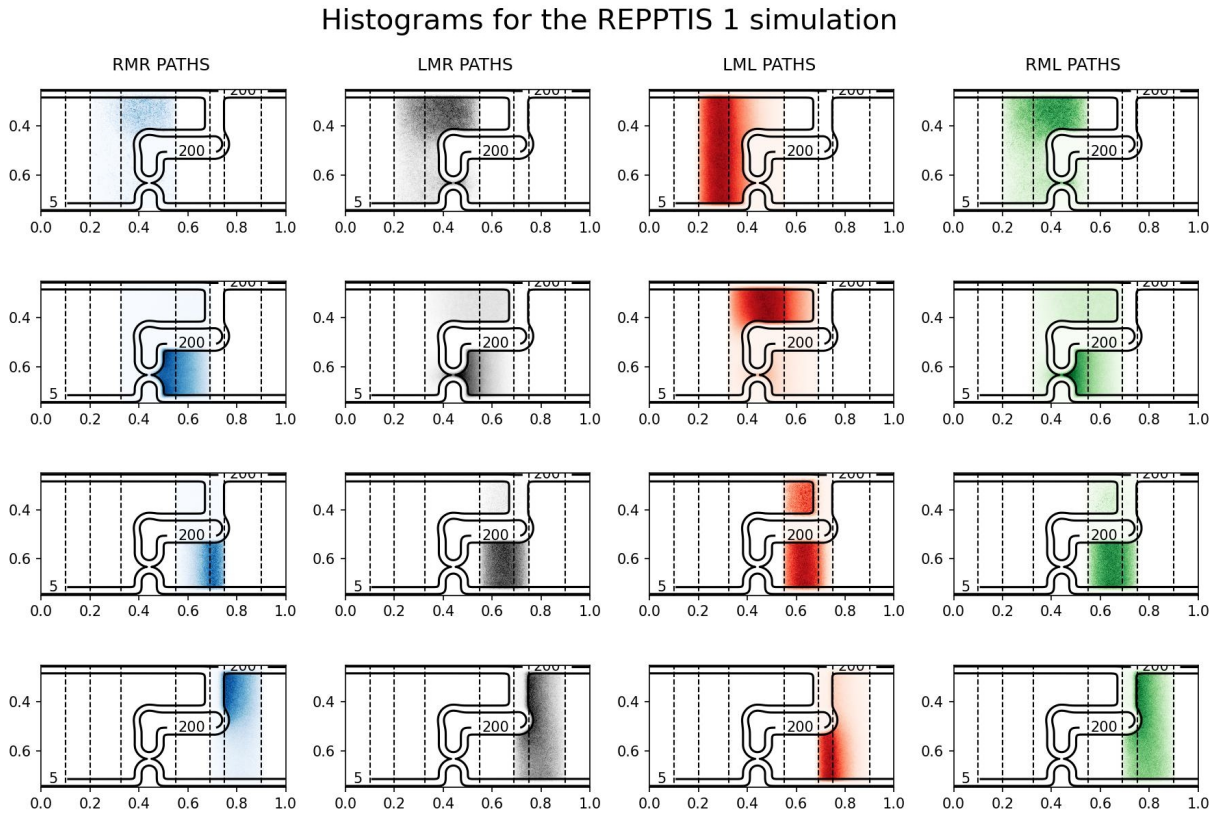

FIG. S4: Same as Fig. S2, for the REPPTIS 1 simulation.

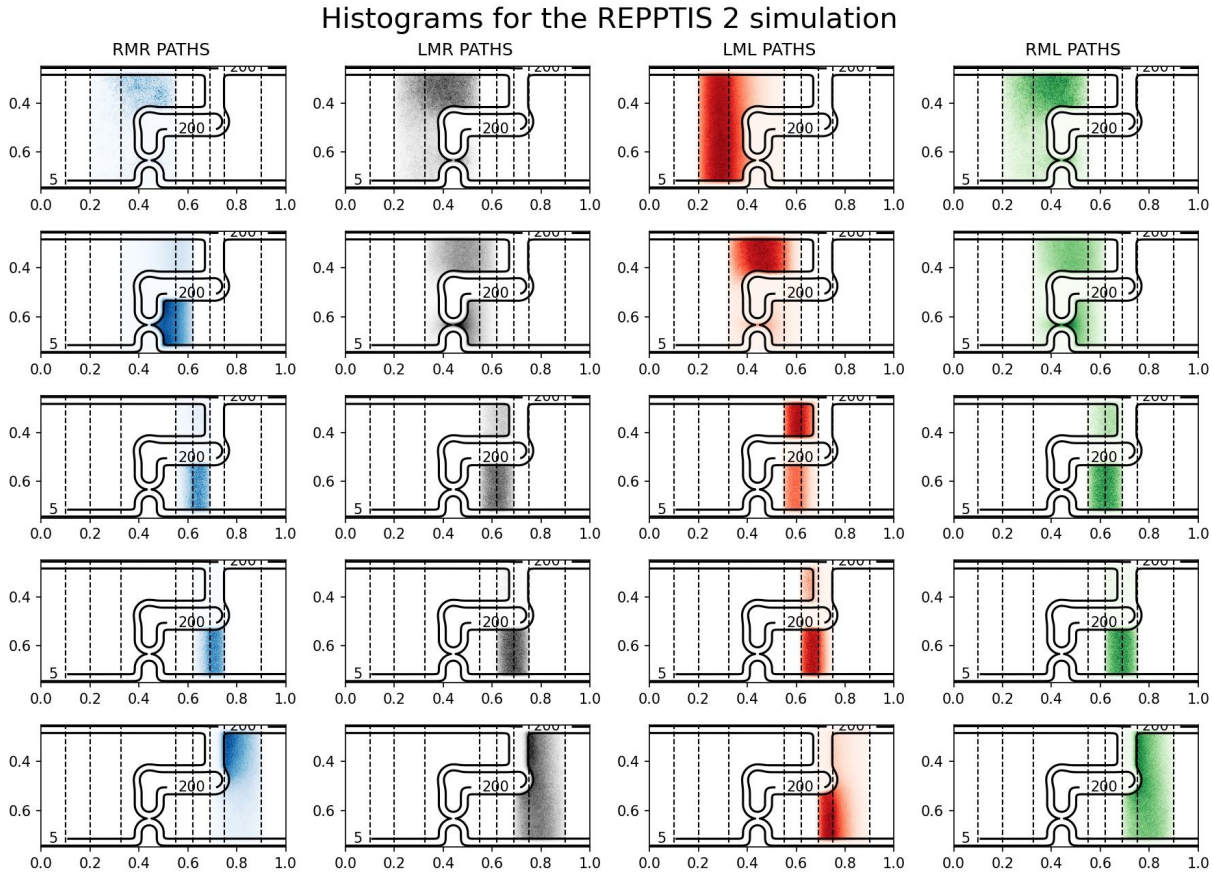

FIG. S5: Same as Fig. S2, for the REPPTIS 2 simulation. An extra row is included for REPPTIS 2, due to the extra interface $\lambda_{2.5}$. The rows from top to bottom correspond with ensembles $[1^{\pm}]$, $[2^{\pm}]$, $[2.5^{\pm}]$, $[3^{\pm}]$ and $[4^{\pm}]$, and the columns left to right correspond with path types RMR, LMR, LML and RML.

.

II. IBUPROFEN PERMEATION

A. Reconstructing the free energy 2D surface of ibuprofen

The two-dimensional free energy profile $F_{\text{true}}(z, \theta)$ of ibuprofen in a DOPC bilayer is reported by Jambeck et al. in Fig. 2a of Ref. 2, as obtained from all-atom molecular dynamics simulations. The authors also report one-dimensional slices of the profile as a function of z for the *cis* and *trans* configurations, and the free energy profile as a function of θ in water and hexadecane. Tracing these profiles and approximating them with analytical functions allowed us to reconstruct a free energy profile $F(z, \theta)$ resembling their $F_{\text{true}}(z, \theta)$ profile. We will now provide the details how this was done.

One-dimensional slices $F^{cis}_{\text{true}}(z) \equiv F_{\text{true}}(z, \theta = 0)$ and $F^{trans}_{\text{true}} \equiv F_{\text{true}}(z, \theta = \pi)$ are provided in Fig. 2b of Ref. 2. First, these profiles were extracted from the figure using in-house code, which resulted in 716 equidistant sample points $\mathcal{S}_M = \{z_i, F^s_i\}_{i=1\ldots716}$ ranging from 0 to 3.5 nm, with sample spacing $\Delta z_s = 3.5\,\text{nm}/715$. Next, two artificial extensions of the grid were added to the left and the right of this interval. To the left, 50 artificial sample points $\mathcal{S}_L = \{z_1 - i\Delta z_s, F^s_1\}_{i=50\ldots1}$ were added, such that oscillations introduced by the polynomial fit (Runge's phenomenom) are suppressed near $z \approx 0$. The same was done on the right, where 50 artificial sample points $\mathcal{S}_R = \{z_{716} + i\Delta z_s, F^s_{716}\}_{i=1\ldots50}$ were added. Hereafter, a polynomial fit of order 15 was performed to $\mathcal{S} = \mathcal{S}_L \cup \mathcal{S}_M \cup \mathcal{S}_R$. This resulted in the reconstructed profiles $F^{cis}(z) \equiv F(z, \theta = 0)$ and $F^{trans}(z) \equiv F(z, |\theta| = \pi)$, which are shown in Fig. S6a. The regions of the artificial extensions to the left and right are never used in the simulations, as only the domain $z \in [0, 3.4\,\text{nm}]$ is sampled in the PyRETIS path ensembles.

The free energy profiles of ibuprofen in water ($F^{\text{water}}_{\text{true}}(\theta)$) and in n-hexane ($F^{n\text{-hexane}}_{\text{true}}(\theta)$) are provided in Fig. S1 of Ref. 2. These profiles were extracted from the figure using in-house code, resulting in 713 equidistant sample points, with sample spacing $\Delta\theta_s = (2\pi)/712$. A Fourier reconstruction of order 11 was used to reconstruct the profiles from the sample points. As the profile is periodic in θ with period 2π, no Gibbs phenomenom occurs, and no artificial extensions were required. This resulted in the reconstructed profiles $F(z = 0, \theta)$ and $F(z = 3.5\,\text{nm}, \theta)$, which are shown in Fig. S6b.

The free energy profile $F(z, \theta)$ used in our simulations is defined as a linear combination

of the reconstructed one-dimensional profiles

$$F(z,\theta) = F^{cis}(z)\frac{\pi-\theta}{\pi} + F^{trans}(z)\frac{\theta}{\pi} + F^{\text{water}}(\theta)\frac{z}{3.5\,\text{nm}} + F^{n\text{-hexane}}(\theta)\frac{3.5\,\text{nm}-z}{3.5\,\text{nm}}, \qquad (2)$$

for $z \in [0, 3.4\,\text{nm}]$ and $\theta \in [0, \pi]$. Symmetry operations are used to obtain $F(z,\theta)$ for $z \in [-3.4\,\text{nm}, 0]$ and $\theta \notin [0, \pi]$.

The force vector is given by $\left(\frac{\partial F}{\partial z}, \frac{\partial F}{\partial \theta}\right)$. As the reconstructed profiles consist of either polynomials or trigonometric functions, the gradient of Eq. 2 could be computed analytically.

For the water phase region, the reconstruction is ideal, i.e. the original profile describes ibuprofen in water, and in the reconstructed profile it is also positioned in the water phase. For the bilayer interior region, the reconstruction might not be ideal, as n-hexane is possibly a better descriptor of the lipid tail regions $|z| \approx 1$, rather than the midplane region $z = 0$. Nevertheless, the reconstructed profile F more than suffices for the purpose of our simulations.

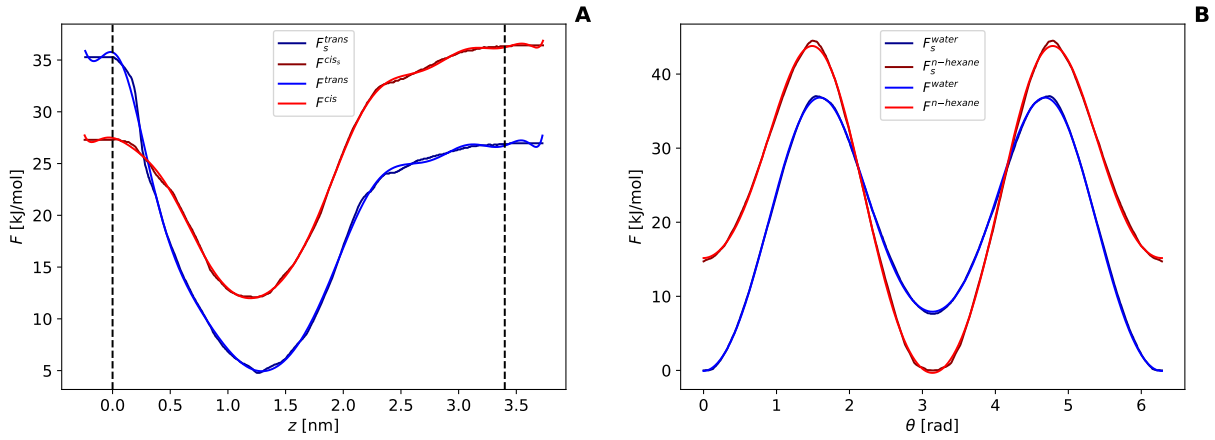

FIG. S6: Reconstructed profiles. For the z-dependent profiles on the left, only the region within the dashed lines is used for the RE(PP)TIS simulations.

B. Langevin dynamics for ibuprofen

The z and θ coordinates of ibuprofen were modeled as two 1D Langevin particles in the PyRETIS code with each their own 'mass'. This mass figures in the kinetic energy term and in the friction term in the Langevin equation. For z, which describes the center-of-mass motion of ibuprofen through the membrane, the associated mass is the total mass of the ibuprofen. For θ, the associated mass is the moment of inertia of the OH bond about the dihedral angle axis, which is the CO bond. In practice, this moment of inertia can be roughly estimated as $m_{\mathrm{H}} d_{\mathrm{OH}}^2$, with m_{H} equal to hydrogen mass, and d_{OH}^2 the average OH bond length. Coriolis effects were neglected, which is acceptable given the friction. Using one 2D Langevin particle would not have worked well here, as PyRETIS assigns one mass to one particle. The mass of ibuprofen, which determines the resistance to the center-of-mass (COM) displacement along z, differs greatly from the inertial moment determining the resistance to rotational displacement along θ, and these masses cannot be approximated by a single value.

The temperature was taken to match the temperature of Ref. 2. A typical friction value of $\gamma = 50\ \mathrm{ps}^{-1}$ was used in the Langevin dynamics which is assumed to be appropriate for aqueous solution.[3] The same friction value was used for z as for θ. GROMACS units (time in ps, length in nm, mass in u, temperature in K) were used to set the Langevin parameters for ibuprofen. The following parameters were used: integration timestep $\Delta t = 0.02\,\mathrm{ps}$, temperature $T = 303\,\mathrm{K}$, friction coefficient $\gamma = 50/\mathrm{ps}$, mass of particle one (z-displacement) $m_{\mathrm{COM}} = 206.31\,\mathrm{u}$, and mass of particle two (θ-displacement) $m_{\mathrm{angle}} = 0.0093\,\mathrm{u}$.

The simulation input files, together with the ibuprofen potential code, can be found on github: WouterWV/pathsampling_toymodels.

C. RE(PP)TIS simulations of ibuprofen

The order parameter λ is the z coordinate. Six simulations were run with PyRETIS:

1. RETIS entrance
2. RETIS internal barrier
3. REPPTIS internal barrier
4. RETIS escape

5. REPPTIS escape
6. REPPTIS full transit

Modeling the full transit is not possible with RETIS because of the two stable states in the leaflet, where trajectories would get trapped. Modeling the entrance was only done with RETIS, because this simulation only needs 3 interfaces and it does not suffer from long trapped paths; hence a REPPTIS simulation would not be that useful.

In practice, two runs were done for each of the six simulations: the simulations were started both from the *cis* and *trans* configuration of ibuprofen. It was verified that all simulations exhibited conformational changes between the *trans* and *cis* conformations, and that these changes occurred in all ensembles. This implies that the two runs are effectively modeling the same system. Therefore, in the reporting of the results, the data of the two runs were combined.

The interfaces used in each of the simulations are given in Tab. S1. REPPTIS uses more interfaces, as both the 'rising' and 'falling' edges of the free energy barrier require interfaces that are separated by $\sim 2k_BT$. RETIS only requires interfaces on the 'rising' edge (and one in stable state B). REPPTIS paths are, however, considerably shorter. The λ_A and λ_B interfaces are matched in RETIS and REPPTIS. For simulation 6, the symmetry of the free energy profile around $z = 0$ was used to (almost) cut the simulation domain in half. Indeed, the interfaces are given by $[-3\,\mathrm{nm}, \ldots, 0, \lambda_b, \lambda_c]$, where the 2 extra interfaces λ_b and λ_c were added such that the $[i^{\pm}]$ with $z = 0$ as middle interface was also sampled. The local crossing probabilities in the $[i^{\pm}]$ ensembles for $z > 0$ can be copied from their symmetric counterpart ensembles in $z < 0$. For run 1, $[7^{\pm}]$ is the PPTIS ensemble with $[\lambda_L, \lambda_M, \lambda_R] = [-0.384\,\mathrm{nm}, 0, 0.384\,\mathrm{nm}]$. The local crossing probabilities of $[(i+N)^{\pm}]$ are given by their symmetric counterpart values in $[(i-N)^{\pm}]$. For example, $p^{\pm}_{[6^{\pm}]} \equiv p^{\mp}_{[8^{\mp}]}$ and $p^{=}_{[5^{\pm}]} \equiv p^{\ddagger}_{[9^{\pm}]}$.

Denote N^i_j as the amount of RE(PP)TIS cycles for run i of simulation j. Then $N^1_1 = 91600$, $N^2_1 = 100000$, $N^1_2 = 24982$, $N^2_2 = 21950$, $N^1_3 = 43101$, $N^2_3 = 45236$, $N^1_4 = 31231$, $N^2_4 = 28966$, $N^1_5 = 100001$, $N^2_5 = 74430$, $N^1_6 = 16506$, and $N^2_6 = 17355$.

The crossing probabilities $P(\lambda_i|\lambda_A)$ for simulations 2, 3, 4 and 5 are given in Fig. S7. REPPTIS and RETIS give statistically identical results.

The crossing probabilities $P(\lambda_i|\lambda_A)$ for simulation 6 are given in Fig. S8. The crossing

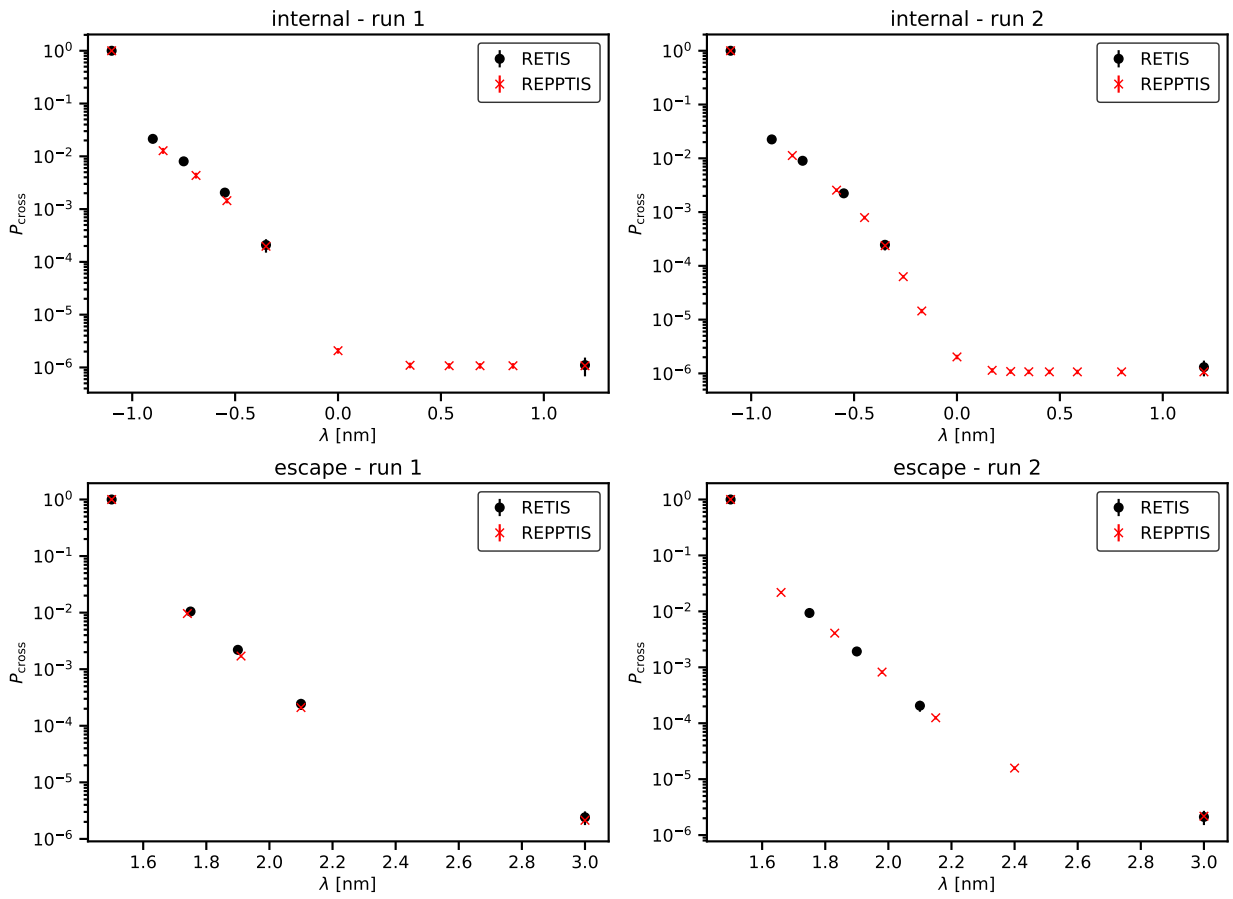

FIG. S7: Crossing probabilities for the internal barrier (top) and the escape (bottom), studied in simulations 2, 3, 4, and 5. The data of the individual runs (that have a different initial path) are shown. Error bars are standard errors from block averaging.

probability to reach the second leaflet is approximately half the crossing probability to reach the first leaflet. Indeed, when ibuprofen is located in the first leaflet, the probability of escaping to the water phase is approximately equal to the probability of overcoming the internal energy barrier towards the second leaflet. This is a direct consequence of the similar magnitudes of the free energy barriers of both (rare) transitions.

sim.	description	interfaces $[\lambda_A, \lambda_1, \ldots, \lambda_B]$ (in nm)	
		run 1 (initial path *cis*)	run 2 (initial path *trans*)
1†*	RETIS entrance	[-3, -2.059, -1.885]	[-3, -2.4, -2.15]
2	RETIS int. barr.	[-1.1, -0.9, -0.75, -0.55, -0.35, 1.2]	[-1.1, -0.9, -0.75, -0.55, -0.35, 1.2]
3	REPPTIS int. barr.	[-1.1, -0.85, -0.69, -0.54, -0.35, 0, 0.35, 0.54, 0.69, 0.85, 1.2]	[-1.1, -0.8, -0.585, -0.449, -0.349, -0.261, -0.171, 0, 0.171, 0.261,0.349, 0.449, 0.585, 0.8, 1.2]
4	RETIS escape	[1.5, 1.75, 1.9, 2.1, 3]	[1.5, 1.75, 1.9, 2.1, 3]
5	REPPTIS escape	[1.5, 1.74, 1.91, 2.1, 3]	[1.5, 1.66, 1.83, 1.98, 2.15, 2.4, 3]
6*	REPPTIS full transit	[-3, -2.059, -1.885, -1.29, -0.885, -0.717, -0.384, 0, 0.384, 0.717]	[-3, -2.4, -2.15, -1.98, -1.83, -1.66, -1.25, -0.8, -0.585, -0.449, -0.349, -0.261, -0.171, 0, 0.171, 0.261]

TABLE S1: Interfaces used in the different RE(PP)TIS simulations. The same λ_A and λ_B interfaces are used in simulations for RETIS as for REPPTIS. This is to ensure that the comparison of the P_{int} values from RETIS and REPPTIS make sense. Wirefencing was used for simulations 2 and 4.

$*$: simulations 1 and 6 use a $\lambda_{-1} = -3.4\,\text{nm}$ interface, while the other simulations use no λ_{-1} interface.

†: Simulation 1 uses a flat free energy region in the water phases, which was implemented by imposing $F(z, \theta) = F(3\,\text{nm}, \theta)$, $\forall |z| > 3\,\text{nm}$. This is done to obtain a better estimate for $\tau_{\text{ref},[0-']}$ and ξ. The crossing probabilities of the other simulations do not depend on F in the water phase.

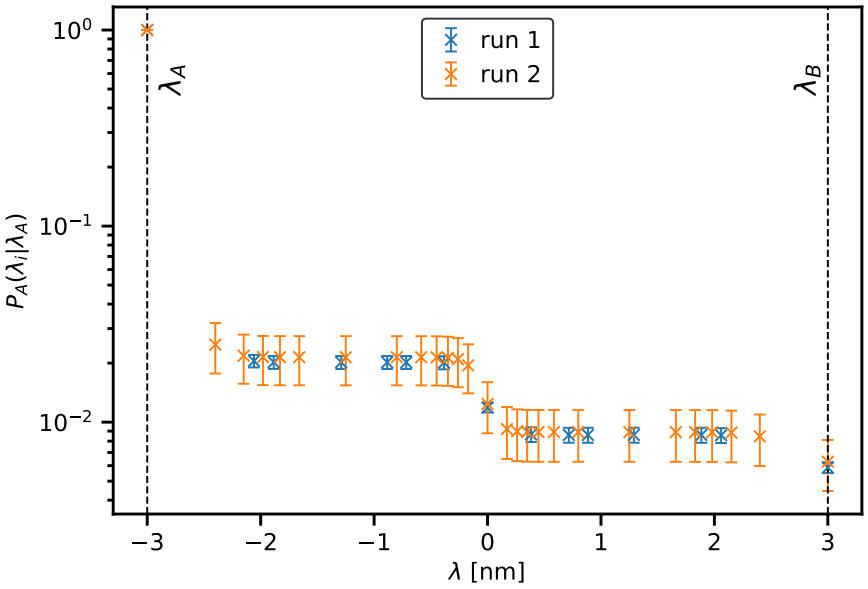

FIG. S8: The REPPTIS crossing probabilities $P_A(\lambda_i|\lambda_A)$ of the full membrane transit of ibuprofen, for the first (blue) and second (orange) runs. Due to *cis-trans* transitions, the dihedral angle configuration of the initial path has no effect on the end results. Error bars are standard errors from block averaging.

D. Markov model for permeation

As RETIS cannot be used for the full transit, the crossing probability of the full transition P_{trans} can be approximated using a Markov model. The Markov model consists of 4 states, which are vizualized in Fig. S9. State A and D are the left and right water phases, respectively. State B and C are the free energy minima in the first and second leaflets, respectively. Let P_{XY} denote the probability of going from state X to a neighbouring state Y. Let $P\left(\begin{smallmatrix} C \\ A \end{smallmatrix} \middle| B\right)$ denote the probability of reaching C for the first time, given that you are in B, and without passing through A. Similarly, let $P\left(\begin{smallmatrix} D \\ A \end{smallmatrix} \middle| C\right)$ denote the probability of reaching D for the first time, given that you are in C, and without passing through A. The full transit probability P_{trans} is then approximated by

$$P_{\text{trans}} = P_{AB} P\left(\begin{smallmatrix} C \\ A \end{smallmatrix} \middle| B\right) P\left(\begin{smallmatrix} D \\ A \end{smallmatrix} \middle| C\right). \tag{3}$$

$P\left(\begin{smallmatrix} C \\ A \end{smallmatrix} \middle| B\right)$ is simply given by P_{BC}, while $P\left(\begin{smallmatrix} D \\ A \end{smallmatrix} \middle| C\right)$ is found using a recursive relation

$$P\left(\begin{smallmatrix} D \\ A \end{smallmatrix} \middle| C\right) = P_{CB} P\left(\begin{smallmatrix} C \\ A \end{smallmatrix} \middle| B\right) P\left(\begin{smallmatrix} D \\ A \end{smallmatrix} \middle| C\right) + P_{CD}.$$

Substituting P_{BC} for $P\left(\begin{smallmatrix} C \\ A \end{smallmatrix} \middle| B\right)$, gives

$$P\left(\begin{smallmatrix} D \\ A \end{smallmatrix} \middle| C\right) = \frac{P_{CD}}{1 - P_{CB} P_{BC}}. \tag{4}$$

Plugging these expressions into Eq. 3, the transit probability is given by

$$P_{\text{transit}} = \frac{P_{AB} P_{BC} P_{CD}}{1 - P_{CB} P_{BC}}. \tag{5}$$

The relation between the transition probabilities P_{XY} and the characteristic crossing probabilities is now described. The entrance probability directly gives $P_{AB} = P_{DC} = P_{\text{entr}}$, where the first equality holds due to symmetry. The transition probabilities P_{BA} and $P_{BC} = 1 - P_{BA}$ are given by the relative ratios of the corresponding rates

$$\begin{cases} P_{BA} = \dfrac{k_{BA}}{k_{BA} + k_{BC}} = \dfrac{k_{\text{esc}}}{k_{\text{esc}} + k_{\text{int}}}, \\ P_{BC} = \dfrac{k_{BC}}{k_{BA} + k_{BC}} = \dfrac{k_{\text{int}}}{k_{\text{esc}} + k_{\text{int}}}. \end{cases}$$

Due to symmetry, $P_{BC} = P_{CB}$ and $P_{CD} = P_{BA}$. Plugging the values of the transition probabilities into Eq. 3, the final expression for the transit probability is obtained

$$P_{\text{trans}} = \frac{P_{\text{entr}} k_{\text{int}}}{k_{\text{esc}} + 2k_{\text{int}}}. \tag{6}$$

The rates k_{esc} and k_{int} are found by multiplying the crossing probabilities P_{esc} and P_{int} with their respective fluxes f_{esc} and f_{int}

$$\begin{cases} k_{\text{int}} = f_{\text{int}} P_{\text{int}}, \\ k_{\text{esc}} = f_{\text{esc}} P_{\text{esc}}. \end{cases}$$

This final step is not required, as the rate – together with the flux and crossing probability – are part of the RETIS output.

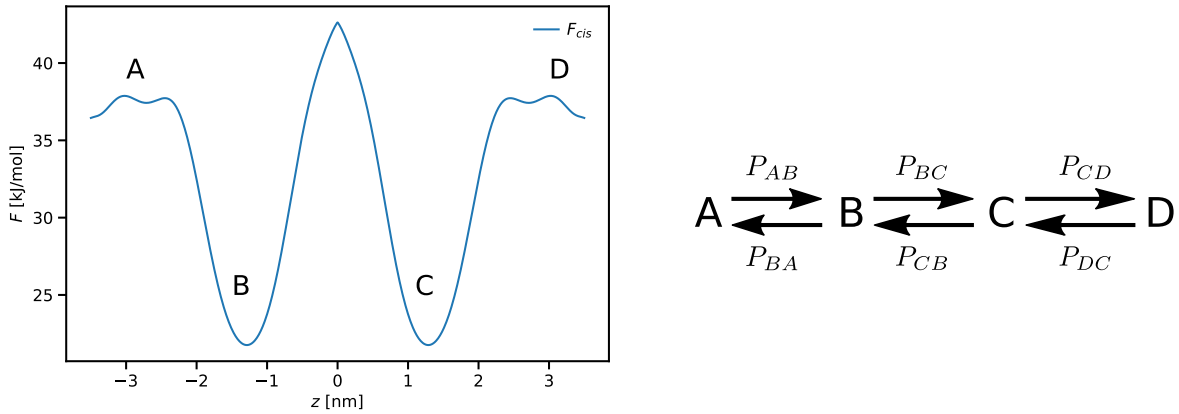

FIG. S9: The 4 states of the Markov model (left) and the transition probabilities between those states (right).

REFERENCES

[1]L. Verlet, "Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules," *Phys. Rev.*, vol. 159, pp. 98–103, 1967.

[2]J. P. Jambeck and A. P. Lyubartsev, "Exploring the free energy landscape of solutes embedded in lipid bilayers," *The Journal of Physical Chemistry Letters*, vol. 4, no. 11, pp. 1781–1787, 2013.

[3]R. Pastor, B. Brooks, and A. Szabo, "An analysis of the accuracy of Langevin and molecular-dynamics algorithms," *Mol. Phys.*, vol. 65, no. 6, pp. 1409–1419, 1988.