Lipid bilayer permeation of aliphatic amine and carboxylic acid drugs: Rates of insertion, translocation and dissociation from MD simulations

Supporting Information

Tuğçe Oruç, Sami Emre Küçük, and Deniz Sezer*
Faculty of Engineering and Natural Sciences, Sabancı University, Orhanlı-Tuzla, 34956 Istanbul, Turkey
(Dated: August 13, 2016)

SI. ADDITIONAL RESULTS

A. Free MD simulations of dyclonine

1. *Insertion of ionized dyclonine into the lipid bilayer*

All eight ionized dyclonines in the simulation $8D^+$ insert tail-first into the lipid leaflet. Snapshots of one such insertion event are shown in Fig. S1. There, the inserting dyclonine molecule is highlighted in brighter colors and the phosphorus atoms of the lipid phosphates are shown as brown balls.

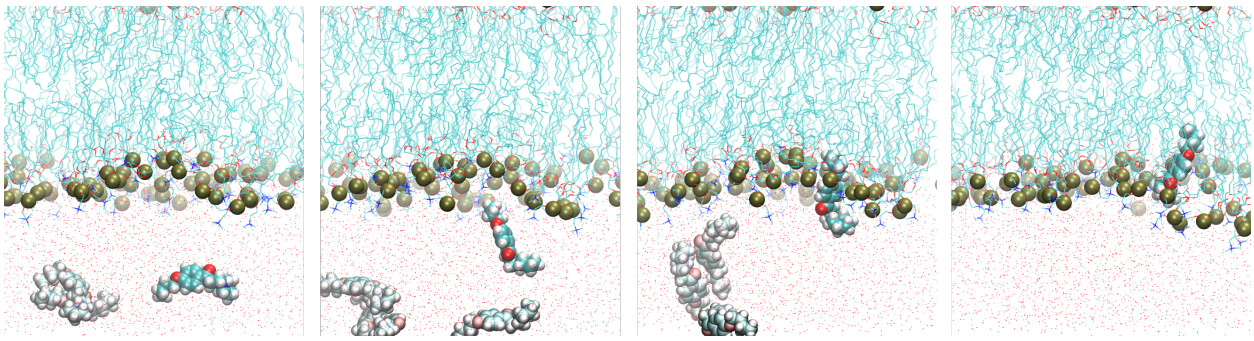


FIG. S1. Snapshots from the simulation $8D^+$ showing the progressive (from left to right) insertion of a protonated dyclonine into the lipid bilayer.

Just before the 50^{th} ns of the simulation $8D^+$ one drug molecule inserts its charged head group below the phosphates of the upper bilayer leaflet but quickly comes off into the water phase just after the 50^{th} ns because its tail fails to insert (red trace in Fig. 2B of main text). Successive snapshots of this failed insertion attempt are shown in Fig. S2. After leaving the bilayer, this molecule eventually inserts successfully at about the 80^{th} ns (red trace in Fig. 2B of main text).

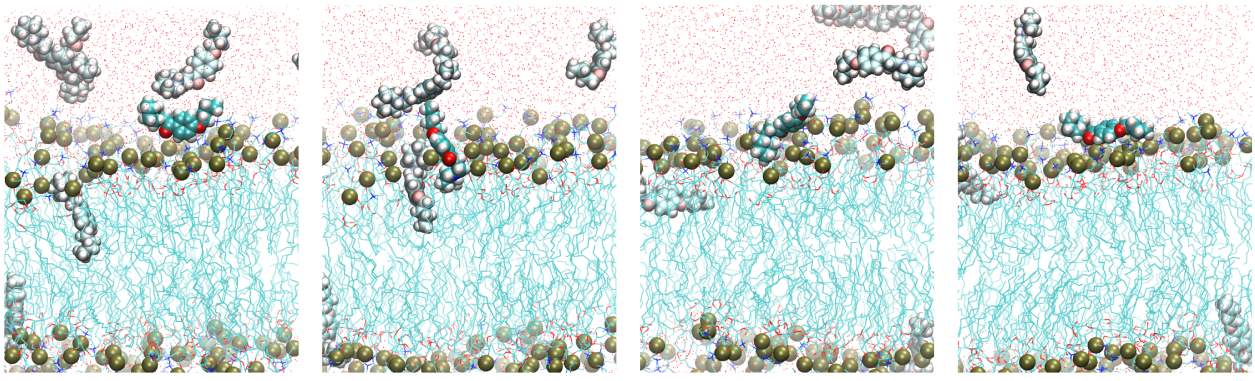


FIG. S2. Snapshots from the simulation $8D^+$ showing the unsuccessful insertion attempt of a protonated dyclonine.

2. *Insertion of neutral dyclonine into the lipid bilayer*

Eight neutral dyclonine molecules were also placed in water outside the pre-equilibrated lipid bilayer to form the system $8D^0$. Initial and final views of the simulation box are shown in Fig. S3A. The positions of the eight drug molecules along the z axis as a function of time are shown in Fig. S3B. During the simulation time of 250 ns only one of the neutral dyclonine molecules is seen to have inserted into the lipid bilayer. The remaining seven drug molecules are observed to have clustered in water (Fig. S3A, right). The neutral dyclonine that has managed to enter into the lipid bilayer is seen to switch between its two leaflets several times. The locations of the drug head groups, the phosphorous atoms of the lipid head groups, the carbon atoms of the lipid tails, and water molecules are given in Fig. S3C.

Snapshots from the insertion of the neutral dyclonine into the lipid bilayer directly from the hydrophobic cluster formed by the other drugs are shown Fig. S4. Interestingly, the drug also inserts its tail first.

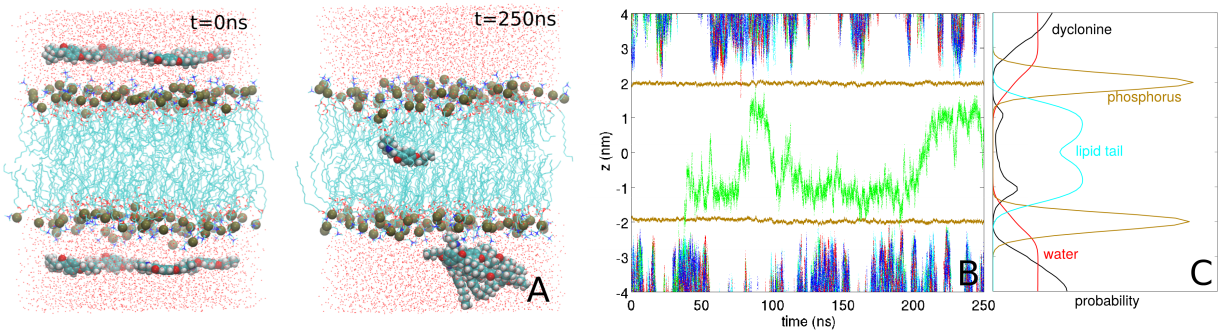


FIG. S3. (A) Initial and final views of the lipid bilayer system with 8 neutral dyclonine molecules ($8D^0$). (B) Positions of the amine groups of the drugs with respect to the center of mass of the lipid phosphates (brown lines) during the MD simulations. (C) Histograms of the drug head groups (black), the phosphorus atoms of the lipids (brown), the carbon atoms of the lipid tails (cyan), and water (red) as extracted from the MD simulations.

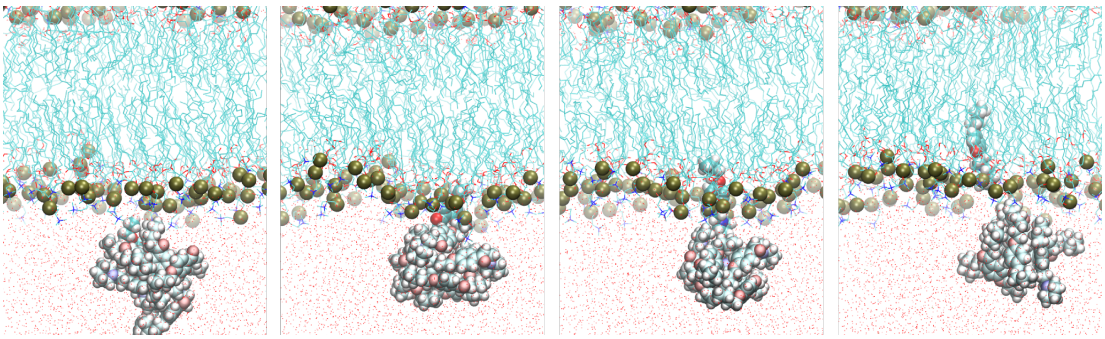


FIG. S4. Snapshots from the simulation $8D^0$ showing the progressive dissociation of a neutral dyclonine (highlighted in brighter colors) from the cluster of drugs and its simultaneous insertion into the lipid bilayer. Similar to ionized dyclonines, the drug tail was first to insert.

3. *Clustering in water*

The size of the largest cluster formed by the 8 ionized and 8 neutral dyclonine molecules was analyzed for the two simulated systems $8D^+$ and $8D^0$ (Fig. S5). When the drug molecules are neutral a cluster containing all of them is seen to form very rapidly (< 5 ns) after initiating the MD simulation (Fig. S5, red solid line). At about the 35^{th} ns one of the dyclonine molecules separates from the cluster and simultaneously inserts into the lipid bilayer, at which point the cluster size drops to 7 molecules. Around the 80^{th} ns, for almost 10 ns the cluster includes 5 dyclonines although 7 molecules are still outside the bilayer. Similar reductions of the cluster size, lasting for less than 1 ns, are observed at about the 135^{th} and 190^{th} ns. This behavior of the neutral drug molecules should be contrasted with

the clustering tendency in $8D^+$ (Fig. S5, black lines). Here, the size of the largest cluster fluctuates rapidly between 1 and 5 molecules. Naturally, the fluctuations decrease when more and more charged dyclonines insert into the lipid bilayer (dashed black line). The cluster size drops to zero when all the ionized molecules enter into the lipid bilayer at about the 140^{th} ns.

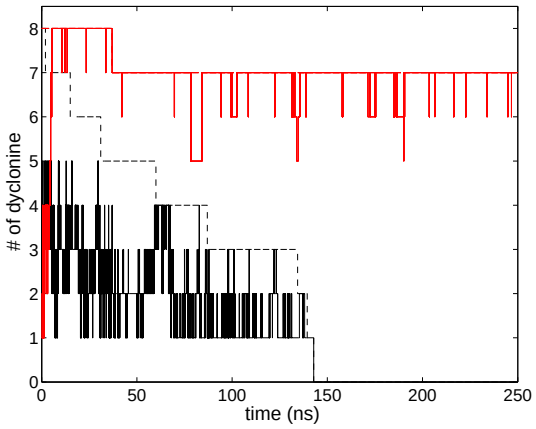


FIG. S5. Cluster sizes of charged (black) and neutral (red) dyclonine molecules in water from simulations $8D^+$ and $8D^0$. The total number of dyclonine molecules outside the lipid bilayer (dashed) is compared with the number of drugs forming a cluster in water (solid).

4. Orientational order of dyclonine in the lipid bilayer

To monitor the orientation of the dyclonine molecule in the lipid bilyaer, a vector pointing from the center of mass of the last four carbon atoms of the drug tail to the center of mass of the drug head group was defined as depicted at the top of Fig. 4A in the main text. The angle between this vector and the positive z direction, denoted by θ in the same figure, was extracted together with the z position of the drug head group for the eight charged dyclonine molecules in the simulation $8D^+$. Dividing the resulting histograms by $\sin\theta$ led to the distributions shown in Fig. S6. Clearly, when in the lipid bilayer the ionized head groups of the drugs stay at $|z| \approx 2$ nm with most likely orientation relative to the bilayer normal of $\theta \approx 0^\circ$ ($z > 0$) or $\theta \approx 180^\circ$ ($z < 0$). In water, the drug molecules freely rotate in all directions.

The θ-z distributions, similarly normalized by $\sin\theta$, are shown in Fig. S7 for each of the four neutral dyclonine molecules in the simulation $4D^0$. Very similar distributions are obtained for the first three drug molecules, while the fourth molecule appears to have sampled only half of the θ-z space accessed by the other drugs. Indeed, the red trace in Fig. 3B of the main text shows that this last drug molecule has never properly transitioned to the lower leaflet of its lipid bilayer. The average of the four distributions in Fig. S7 is shown in Fig. 4A of the main text, where the paths in the θ-z plane that are followed during drug translocation are depicted.

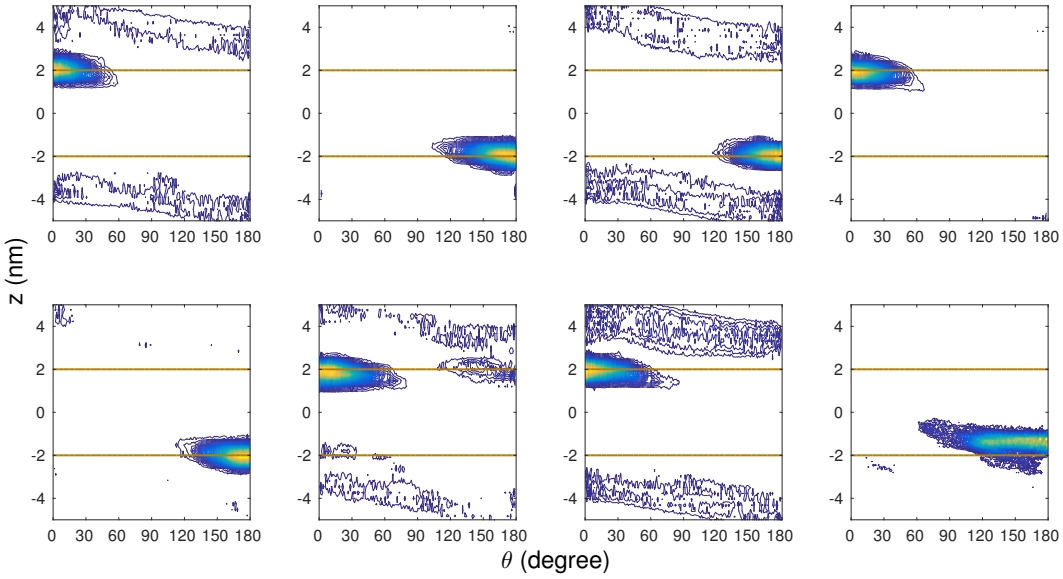


FIG. S6. Distributions of tilt angle, θ, against the position in the bilayer, z, for each of the 8 ionized dyclonine molecules in the simulation $8D^{+}$. The horizontal brown lines at $|z| = 2$ nm depict the centers of mass of the lipid phosphates in the two leaflets of the bilayer. The presence/absense of density for the molecules in water depends on the cut-off selected for the visualization as well as on the relative time each drug spends in water and in the lipid. The density seen for the four molecules that penetrate last into the lipid bilayer (depicted first, third, sixth, and seventh) shows that in water drugs take any orientation with equal probability. In contrast, after inserting into the lipid bilayer the drugs are seen to preferentially align their tails with the normal to the bilayer plane, i.e., most likely θ is 0° ($z > 0$) and 180° ($z < 0$), similar to the phospholipids making up the lipid bilayer.

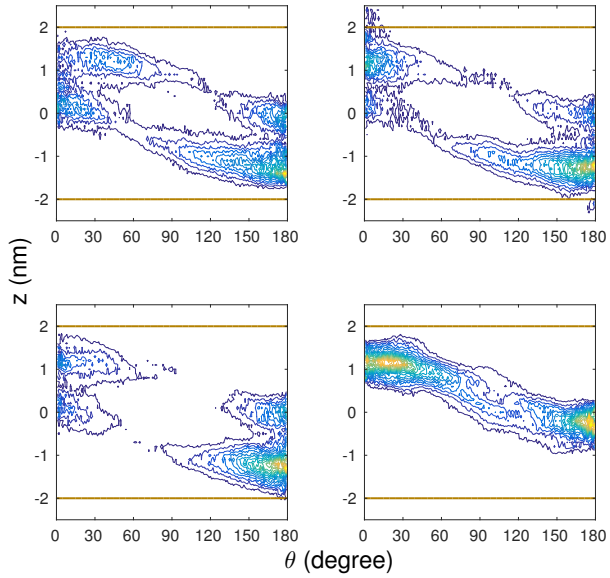


FIG. S7. Same as Fig. S6 for the four neutral dyclonines in the simulation $4D^{0}$.

5. *Effect of simulation time on free energy estimated from unrestrained simulations*

When the free energy of ionized dyclonine is estimated from the logarithm of its probability according to expression (18) in the main text, the value of the free energy in water ($z = 4$ nm) relative to that in the lipid ($z = 1.9$ nm) changes upon increase of the total simulation time. Specifically, two estimates from either the entire 250 ns or only the first 100 ns of the simulation $8D^+$ are shown in Fig. S8. For increasing simulation time, the estimated free energy changes in a way that brings it closer to the estimates from the restrained umbrella sampling (US) simulations also shown in Fig. S8. The remaining discrepancy, however, is still huge since no drug was observed to dissociate from the lipid bilayer during the entire simulation time. The free energies obtained from restrained US simulations and from the logarithm of the probability density in the unrestrained simulations are expected to agree only when the latter contain several drug dissociation events.

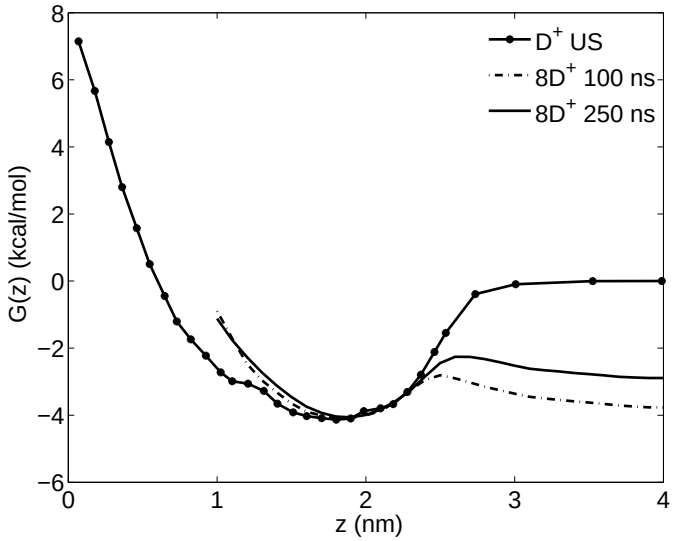


FIG. S8. Free energy profiles estimated from the unrestrained simulation $8D^+$ using the drug positions during either the first 100 ns (dashed line) or all 250 ns (solid line). The estimated free energy difference between the drug in lipid and in water is dictated by the simulation time since no dissociation of the drug from the lipid bilayer was observed during the MD simulation. The free energy profile from a long simulation containing many dissociation events is expected to be similar to the umbrella sampling (US) estimate (symblos).

B. Free MD simulations of 4-phenylbutyrate

1. *Insertion and orientation of ionized phenylbutyrate*

All the eight charged phenylbutyrates insert into and dissociate from the membrane during the 500 ns-long simulation $8F^-$. An example of one insertion event is shown in Fig. S9 which is comprised of 4 snapshots taken between the 174^{th} and 175^{th} ns of the simulation. The snapshots show that the phenyl ring inserts first and then the rest of the molecule follows. Such tail-first insertion applies for all the other events in the MD simulation $8F^-$.

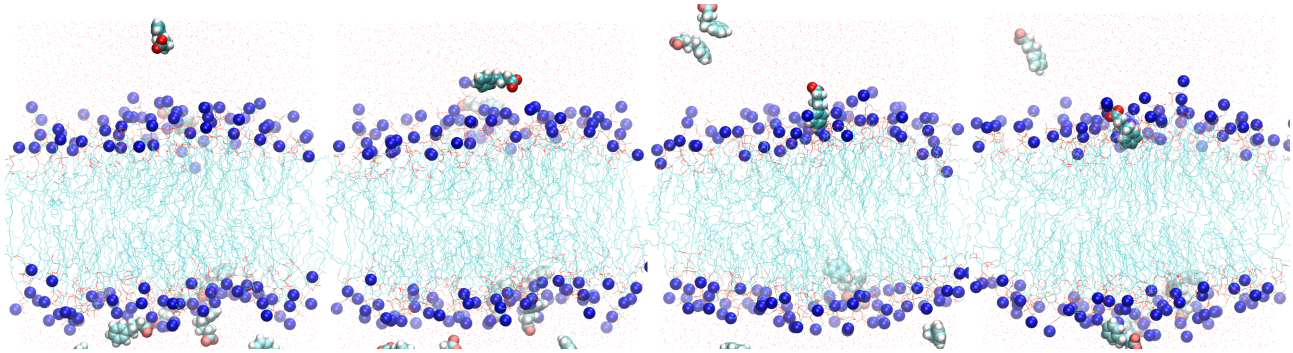


FIG. S9. Snapshots from the simulation $8F^-$ showing the progressive (from left to right) insertion of an ionized phenylbutyrate into the lipid bilayer.

In the lipid bilayer, the ionized phenylbutyrates show preferential alignment with the bilayer normal, similar to that exhibited by ionized dyclonines. This can be seen in the θ-z distributions shown in Fig. S10. To extract the tilt angle θ from the MD trajectories, a vector joining the center of mass of the phenyl ring to the center of mass of the carboxyl group was defined as shown in an inset in Fig. S11. θ is the angle between this vector and the normal of the bilayer. The distributions in Fig. S10 demonstrate that there is no orientational preference in water, while in the lipid bilayer the ionized drug molecules tend to be parallel to the bilayer with the phenyl ring pointing toward the bilayer center and the carboxyl group localized at the interface between the lipid head groups and water.

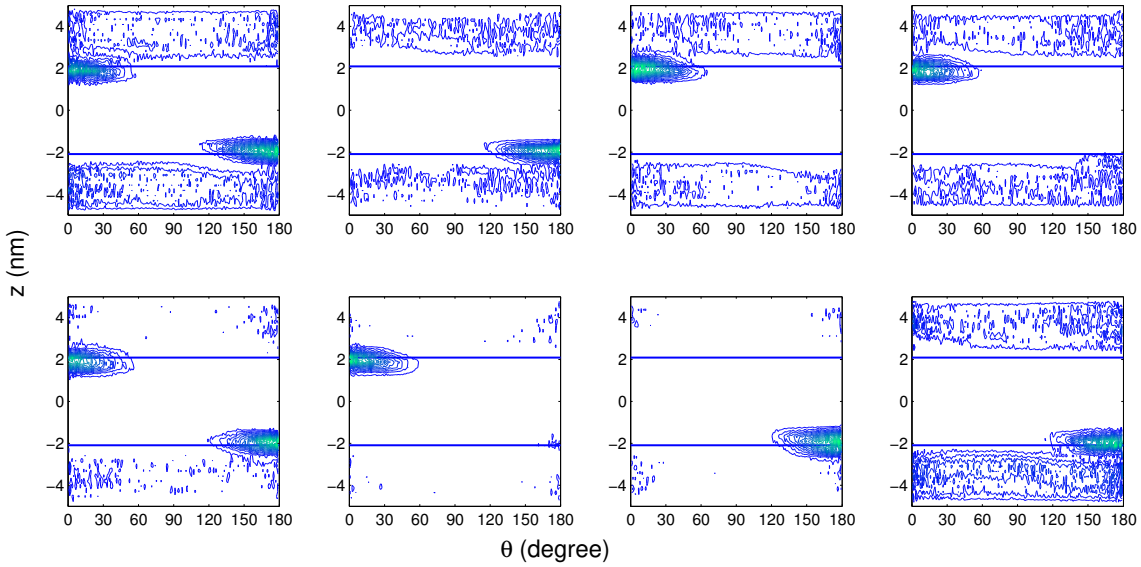


FIG. S10. Tilt angle with respect to the bilayer normal against the position in the lipid bilayer for the eight charged phenylbutyrates in the simulation $8F^-$. The horizontal blue lines at $|z| = 2.1$ nm indicate the most likely location of the positively charged choline groups of the lipids (*cf.* Figs. 7C and 8C in main text).

2. *Neutral phenylbutyrate in lipid: Orientational order and hydrogen bonding*

In the simulation $4F^0$, the internalized neutral phenylbutyrates also show a strong orientational order, as evident from the tilt angle distributions in Fig. S11. While the drug molecules again tend to be parallel to the bilayer normal, in this case the neutral head groups are located deeper in the bilayer well below the choline groups.

The θ-z distributions of neutral dyclonine were informative about the pathways followed by the drugs during translocation between bilayer leaflets (Fig. S7). For neutral phenylbutyrate, the translocation events are rare and relatively fast so that the populations of the tilt angles in the intermediate region are not visible in Fig. S11. In Fig. 9 of the main text, therefore, we showed the entire transition paths with the θ-z distribution of the translocating drug in the background.

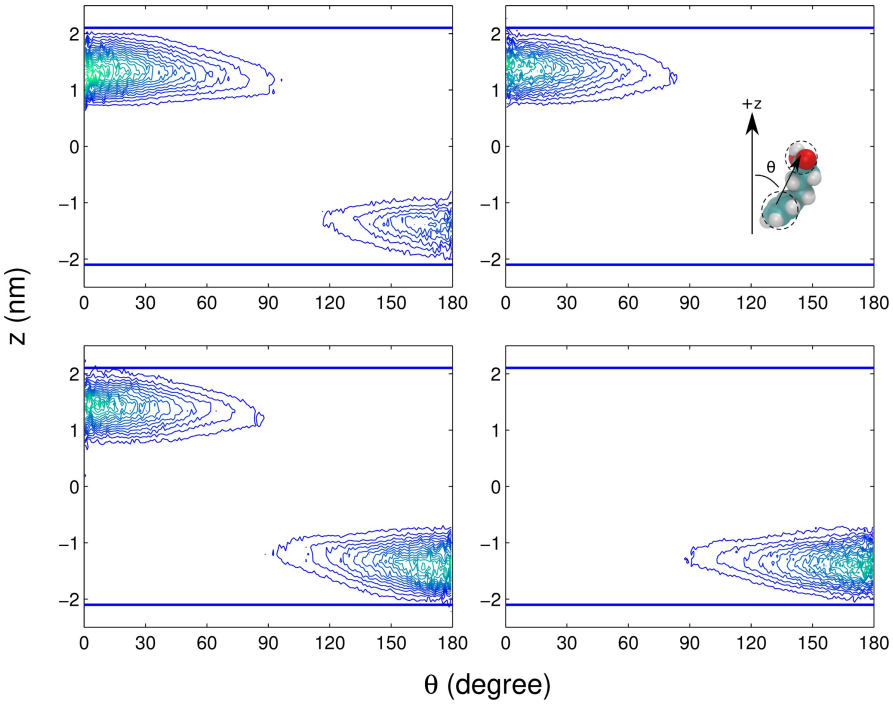
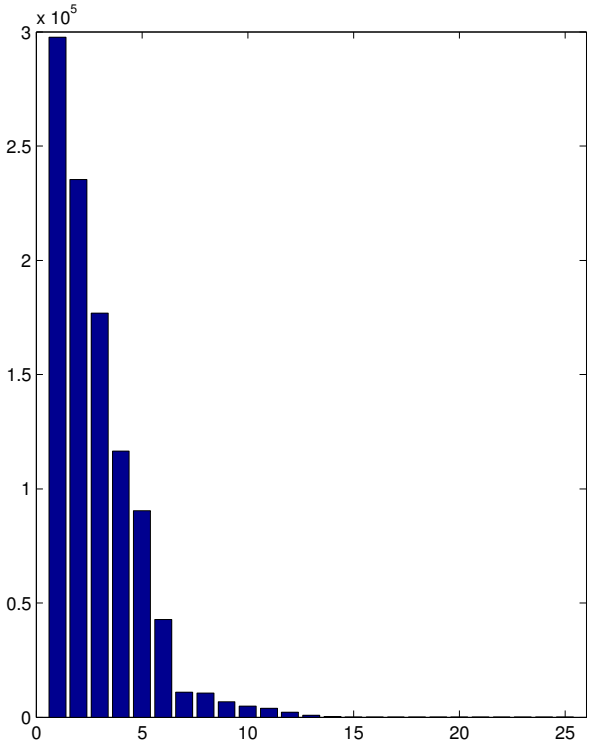


FIG. S11. Tilt angle with respect to the bilayer normal against the position in the bilayer for the four neutral phenylbutyrates in the simulation $4F^0$. The definition of θ is depicted in upper right part of the figure. The horizontal lines correspond to the average position of the lipid choline groups.

Having one hydrogen bond donor (OH) and one acceptor (O), the carboxyl group of an internalized neutral phenylbutyrate can form hydrogen bonds with interfacial waters (W) and with lipid carbonyl groups (C=O). With water, two types of hydrogen bonding are possible: the drug could be either a hydrogen donor (W$\cdots$OH) or acceptor (W$\cdots$O). When hydrogen bonding to a lipid carbonyl (C=O$\cdots$OH) the drug can only be a donor. The 25 different possibilities that emerge when considering simultaneous hydrogen bonding are listed in the table on the right of Fig. S12. The numerical entries in the table indicate the number of hydrogen bonds of a given type that a single drug is engaged in. The possibilities range from complete absence of hydrogen bonding (row number 2 in the table) to five simultaneous hydrogen bonds (rows number 23, 24 and 25).



	W-OH	W-O	C=O-OH
1	1	0	0
2	0	0	0
3	1	1	0
4	0	1	0
5	0	0	1
6	0	1	1
7	2	0	0
8	1	2	0
9	0	2	0
10	2	1	0
11	1	0	1
12	0	2	1
13	1	1	1
14	2	2	0
15	1	3	0
16	3	0	0
17	0	3	0
18	1	2	1
19	2	0	1
20	3	1	0
21	0	3	1
22	2	1	1
23	2	3	0
24	3	2	0
25	1	3	1

FIG. S12. Analysis of type (right) and frequency (left) of hydrogen bonds formed by a neutral phenylbutyrate drug from the MD simulation $4F^0$. The number of occurrences of the 25 different possibilities for hydrogen bonding, given along the vertical axis of the histogram, add up to a total of 10^6 cases, which result form 25×10^4 MD snapshots (coordinates saved every 2 ps) with each snapshot containing 4 drug molecules.

We investigated the hydrogen bonds formed by the four drug molecules in the entire simulation $4F^0$ using VMD.[1] A geometric criterion—distance between heavy atoms of less than 0.35 nm and an angle of less than 30°— was employed to identify the presence of a hydrogen bond. The statistics for all the 25 possibilities of simultaneous bond formation are shown as a histogram on the left of Fig. S12. Most likely (occurring almost 30% of the time) is hydrogen bonding with a single water in which the drug is a hydrogen donor. The absence of any hydrogen bond is second, occurring 24% of the time. This is followed by simultaneous hydrogen bonding to two waters (18%). In fact, the different possibilities in the table on the right of Fig. S12 were numbered according to their frequency of occurrence in the MD simulation. Thus, fourth (12%) is hydrogen bonding to single water where the drug is now an acceptor, fifth (9%) is hydrogen bonding to a single lipid carbonyl only, and sixth (4%) is hydrogen bonding to one lipid carbonyl (drug as donor) and one water (drug as acceptor). The remaining possibilities occur with progressively smaller frequencies. Although much less likely, simultaneous hydrogen bonding to 4 waters and one lipid carbonyl was also observed (numbered 25 in the table on the right of Fig. S12).

C. Umbrella Sampling Simulations

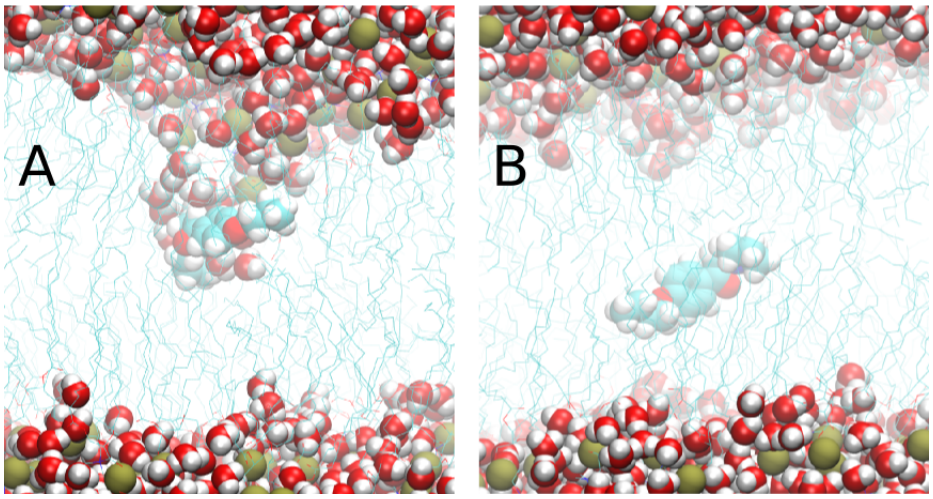


FIG. S13. Snapshots from the umbrella sampling simulations with (A) ionized and (B) neutral dyclonine molecules harmonically restricted to the middle of the lipid bilayer ($z = 0$). In (A), the positively charged dyclonine drags negatively charged phosphate groups (brown balls) of lipid head groups and water molecules into the membrane core, thus deforming the membrane. In contrast, the neutral dyclonine in (B) does not deform the membrane.

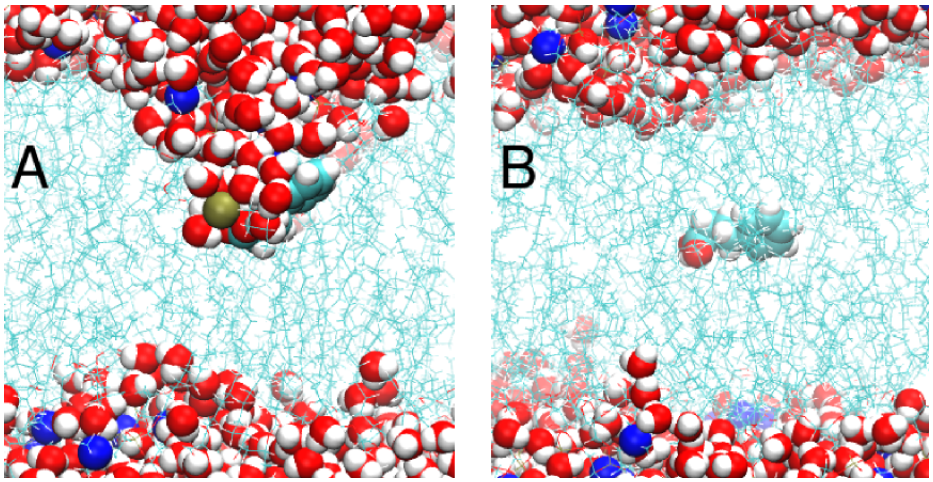


FIG. S14. Snapshots from the umbrella sampling simulations with (A) ionized and (B) neutral phenylbutyrate molecules harmonically restricted to $z = 0$. In (A), the negatively charged group of phenylbutyrate drags a positively charged potassium ion (brown ball) and water molecules into the membrane core resulting in a membrane deformation. Neutral phenylbutyrate in (B) does not deform the membrane in spite of the hydrogen bonding ability of its carboxyl end. The nitrogen atoms of the lipid head groups are shown as blue balls.

D. Average free energy profile in the case of fast exchange

Average free energy profiles, relevant under the assumption of extremely fast kinetics of protonation/deprotonation, were calculated according to equation (34) in the main text. Such profiles are shown in Figs. S15 and S16 for the choice of $r_0 = 1/64$, which corresponds to pH 6.6 for both dyclonine (pK_a of 8.4) and phenylbutyrate (pK_a of 4.8).

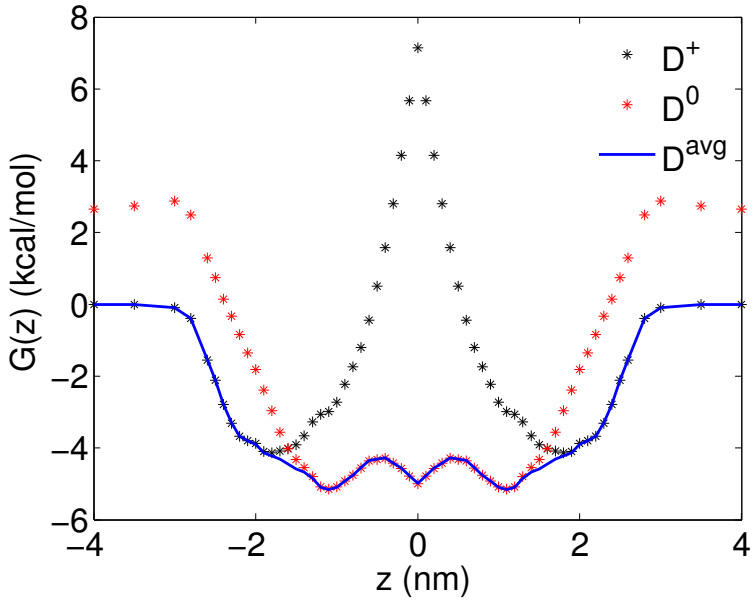


FIG. S15. The average free energy profile of dyclonine (blue line) at pH 6.6 calculated from the US profiles of the neutral (red symbols) and ionized (black symbols) forms of the drug.

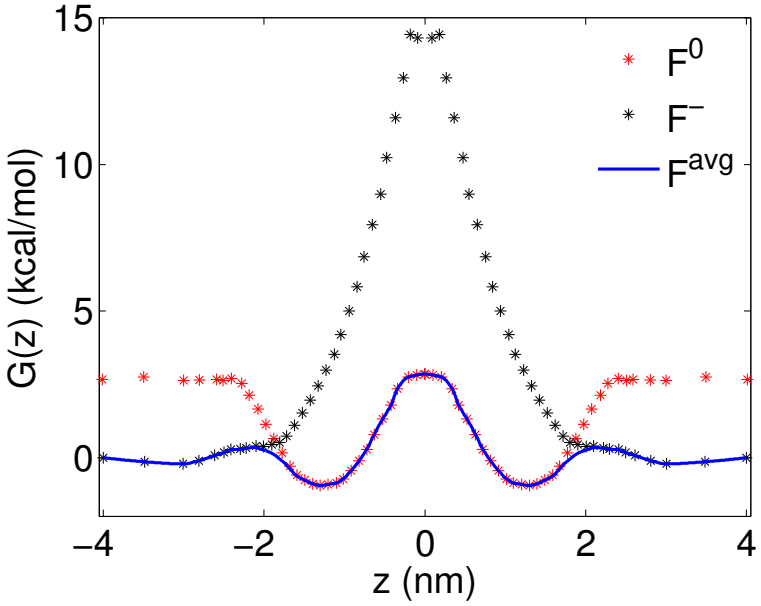


FIG. S16. The average free energy profile of phenylbutyrate (blue line) at pH 6.6 calculated from the US profiles of the neutral (red symbols) and ionized (black symbols) forms of the drug.

SII. SIMULATION METHODS

A. Unrestrained MD simulations

All simulations were performed with NAMD[2] using the CHARMM36 force field for lipids.[3,4] Electrostatic interactions were treated with the particle-mesh Ewald method and covalent bonds to hydrogen atoms were constrained with the SHAKE algorithm as implemented in NAMD. A Langevin thermostat with fiction of 1 ps^{-1} was used to maintain an average temperature of 323.15 K. An average pressure of 101325 Pa (1 atm) was maintained with a Langevin barostat with parameters of piston period 50 fs and piston decay 25 fs.

1. Insertion from water into lipid

A simulation box consisting of a lipid bilayer composed of 128 DPPC molecules, 7131 TIP3P waters, 22 K^+ and 22 Cl^- ions (0.15 mM KCl) was constructed with CHARMM-GUI.[5] The unit cell was a hexagonal prism with a height of 9.5 nm and a side of the hexagon of 6.8 nm. This pure lipid bilayer system was simulated for 13 ns after which eight drug molecules were introduced in water well outside the lipid bilayer.

When introducing the positively charged dyclonine molecules, 4 K^+ ions were removed from and 4 Cl^- ions were added to the 22 ions of each type that were already present, thus ensuring charge neutrality. Water molecules overlapping with the introduced drugs were removed. The resulting system, referred to as $8D^+$ and shown on the left of Fig. 2A in the main text, included 128 DPPC lipids, 18 K^+ and 26 Cl^- ions, 7003 waters, and 8 ionized dyclonine molecules.

The number of ions was left unchanged when inserting eight neutral dyclonines. This system, referred to as $8D^0$ and shown on the left of Fig. S3A, was composed of 128 DPPC lipids, 22 K^+ and 22 Cl^- ions, 7006 waters, and 8 deprotonated dyclonine molecules.

When introducing eight negatively charged phenylbutyrates to construct the system $8F^-$, charge neutrality was preserved by adding 4 K^+ ions to and removing 4 Cl^- ions from the already present ones. After deleting the overlapping water molecules, the resulting system contained 128 DPPC lipids, 26 K^+ and 18 Cl^- ions, 7047 waters, and 8 ionized phenylbutyrates. It is shown on the left of Fig. 7A in the main text.

The systems $8D^+$, $8F^-$, and $8D^0$ were simulated for 250 ns (Figs. 2B and 7B in main text, and Fig. S3B).

2. Neutral drugs inside the lipid bilayer

To study the behaviour of neutral drugs in the interior of the lipid bilayer a simulation box in the form of a hexagonal prism that contained two stacked lipid bilayers was constructed. To this end, a new bilayer system consisting of 128 DPPC lipids, 2560 waters, 4 K^+ and 4 Cl^- ions (0.15 mM) was built using CHARMM-GUI.[5] After equilibration, the system dimensions were 6.8 nm (side of the hexagon) and 5.8 nm (height of the prism). The atomic coordinates were then duplicated and shifted by 5.8 nm in the z direction. A double bilayer system was formed by assembling the two newly generated bilayers in a hexagonal prism of height 11.6 nm. At the end, two neutral drugs were inserted into the center of each lipid bilayer for both dyclonine and phenylbutyrate, as shown on the left hand sides of Figs. 3A and 8A in the main text. The resulting systems, referred to as $4D^0$ and $4F^0$ respectively, were simulated for 500 ns (Figs. 3B and 8B in the main text).

B. Umbrella sampling simulations

In the umbrella sampling (US) simulations, in addition to the interactions described by the molecular force field, harmonic interaction potential was introduced between the center of mass of the drug head group, as defined in Fig. 1 of the main text, and the center of mass of all lipids forming the lipid bilayer. (Hydrogen atoms were not included in the calculations of center of mass.) We aimed to obtain the free energy along this reaction coordinate from $z = 4$ nm (drug in water) to $z = 0$ nm (drug in the middle of the lipid bilayer).

The pre-equilibrated pure bilayer system described in the first paragraph of Sec. Sii A 1 was used as a starting point for the US simulations. To reduce sampling time, two molecules of the same drug—one neutral and one charged—were introduced simultaneously in water away from the lipid bilayer. Charge neutrality was maintained by removing either one K^+ ion (in the case of dyclonine) or one Cl^- ion (in the case of phenylbutyrate). Waters overlapping with the drug molecules were also deleted. For the US simulation with restraining potential at $|z| = 4$ nm, for example, the ionized

form of the drug was restrained to $z = +4$ nm while the neutral form was restrained at $z = -4$ nm. A snapshot from this US simulation containing one ionized (above the lipid bilayer) and one neutral (below the lipid bilayer) dyclonine molecule is shown on the left of Fig. S17. Similar snapshots from the US simulations with dyclonines restrained to $z = \pm 2.1$ nm and $z = \pm 1.1$ nm are shown in the middle and on the right of the same figure.

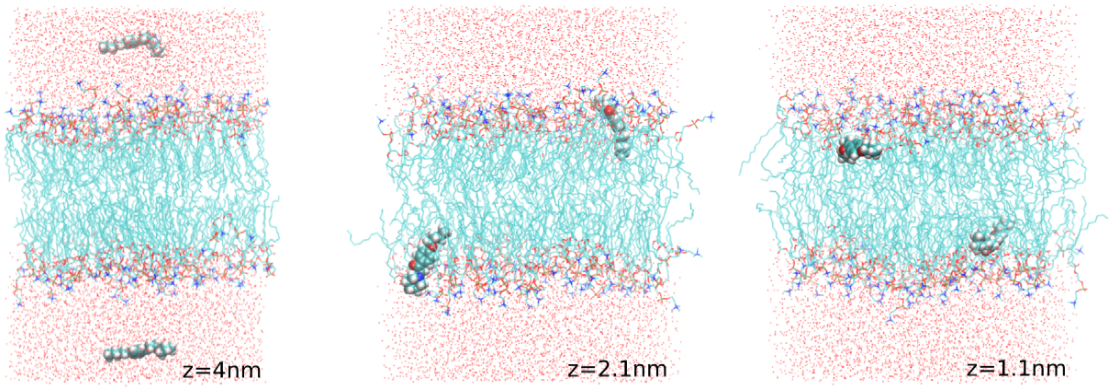


FIG. S17. Both charged (up) and neutral dyclonine (down) molecules were inserted into the same system. Initial views of the simulation boxes are shown for umbrellas centered at ±4 nm, ±2.1 nm, and ±1.1 nm.

US simulations with umbrella potentials separated by 0.5 nm were performed for the drugs in water (3 nm $\leq |z| \leq$ 4 nm). To ensure that values of z sampled in two neighbouring US simulations had sufficient overlap, relatively soft spring constants ($k \leq 0.5$ kcal mol^{-1}Å^{-2}) were selected for these umbrellas. As the drugs approached the lipid bilayer, the interval between two neighbouring umbrellas was first reduced to 0.2 nm (2.6 nm $\leq |z| \leq 3$ nm) and eventually to 0.1 nm ($|z| \leq 2.6$ nm). At the same time, the spring constants were increased, and kept at $k = 2$ kcal mol^{-1}Å^{-2} for the umbrellas with $|z| < 2.6$ nm. Information about the positions of the restraining potentials, their harmonic spring constants, and the total simulation time of all the performed US simulations is compiled in Table S1.

To eliminate direct and indirect interactions between the drugs, separate simulation boxes containing only the charged or the neutral form of the drug were constructed and simulated for small values of z. The initial snapshots from the US simulations in which the drug is restrained to $|z| = 0.6$ nm are given in Fig. S18. The protonated and deprotonated forms of dyclonine were simulated separately in all US simulations with potentials centered at $|z| \leq 1$ nm. In the case of phenylbutyrate, the protonated and deprotonated forms were simulated separately in all US simulations with potentials centered at $|z| \leq 0.9$ nm. This is reflected in Table S1 where separate information is provided for each separate US simulation.

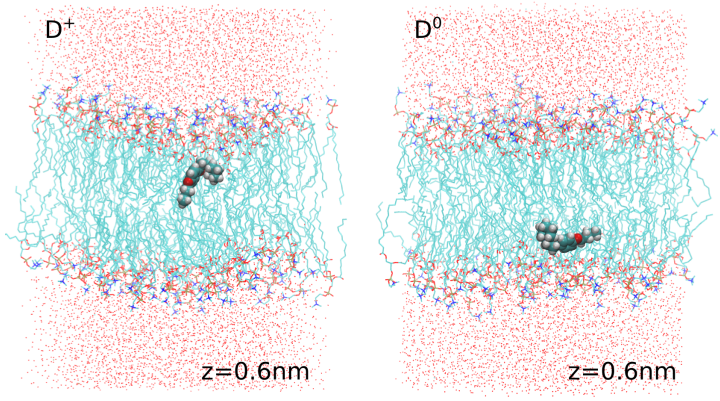


FIG. S18. Initial snapshots from the US simulations of ionized (left) and neutral (right) dyclonine with umbrella potentials centered at $|z| = 0.6$ nm. The ionized and neutral drugs were separated in the simulations with umbrellas held at $|z| \leq 1$ nm.

Given the ease with which a neutral dyclonine switched back and forth between the two leaflets of the lipid bilayer (Fig. 3B in main text), in the US simulations of neutral dyclonine in the hydrophobic core of the lipid ($|z| \leq 1$ nm) umbrella potentials with comparatively softer springs ($k = 0.8$ kcal mol^{-1}Å^{-2}) were spaced by 0.2 nm. For all other

TABLE S1. Force constants (in kcal mol^{-1}Å^{-2}) applied to harmonically restrain dyclonine (D) and phenylbutyrate (F) to distance z from the center of mass of the lipid bilayer. The total simulation times of the umbrella sampling (US) simulations (in ns) are given in parenthesis. One column is used for the ionized and neutral forms of a drug when the two drug forms were simultaneously present in the simulation box. Separate columns for the neutral and charged forms of a given drug indicate that only one drug molecule was present in the corresponding US simulation.

$\|z\|$/nm	D^+	D^0	F^-	F^0
4.0	0.5 (20)		0.5 (15)	
3.5	0.15 (20)		0.15 (15)	
3.0	0.5 (15)		0.5 (15)	
2.8	0.8 (15)		0.8 (15)	
2.6	1.0 (15)		1.0 (15)	
2.5, 2.4, 2.3, 2.1	2.0 (15)		2.0 (30)	
2.0	2.0 (20)		2.0 (15)	
1.9, 1.8, 1.7, 1.6, 1.5, 1.4	2.0 (15)		2.0 (15)	
1.3	2.0 (20)		2.0 (15)	
1.2, 1.1	2.0 (15)		2.0 (15)	
1.0	2.0 (15)	1.0 (15)	2.0 (15)	
0.9	2.0 (15)	-	2.0 (15)	2.0 (15)
0.8	2.0 (15)	0.8 (20)	2.0 (15)	2.0 (15)
0.7	2.0 (15)	-	2.0 (15)	2.0 (15)
0.6	2.0 (15)	0.8 (20)	2.0 (15)	2.0 (15)
0.5	2.0 (15)	-	2.0 (15)	2.0 (15)
0.4	2.0 (15)	0.8 (20)	2.0 (15)	2.0 (15)
0.3	2.0 (15)	-	2.0 (15)	2.0 (15)
0.2	2.0 (15)	0.8 (20)	2.0 (15)	2.0 (15)
0.1	2.0 (15)	-	2.0 (15)	2.0 (15)
0	2.0 (15)	0.8 (15)	2.0 (15)	2.0 (15)

drug species, including neutral phenylbutyrate, the spacing between the umbrella potentials in the interior of the lipid bilayer was 0.1 nm (Table S1). Histograms of the values of z sampled in the US simulations of neutral and charged dyclonine are shown in Fig. S19. The same histograms for neutral and charged phenylbutyrate are given in Fig. S20. The information contained in these histograms was converted to the free energy profiles reported in Figs. 5 (dyclonine) and 10 (phenylbutyrate) of the main text using the Weighted Histogram Analysis Method (WHAM).[6]

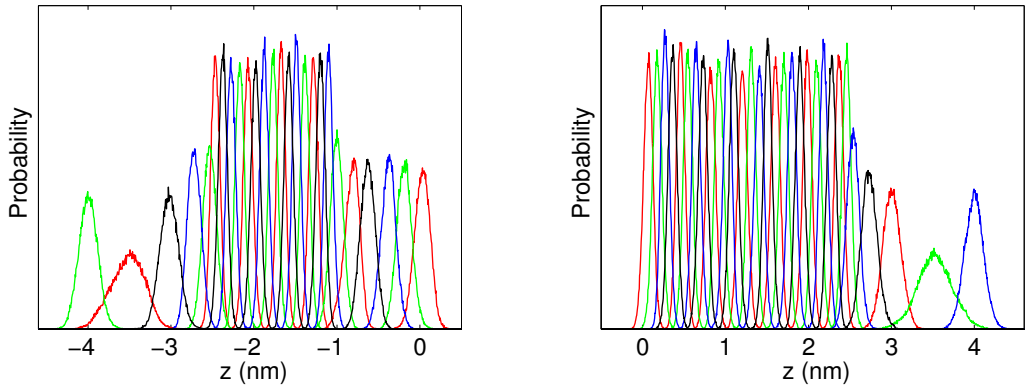


FIG. S19. Position distributions of the neutral (left) and charged (right) dyclonine molecules from the US simulations.

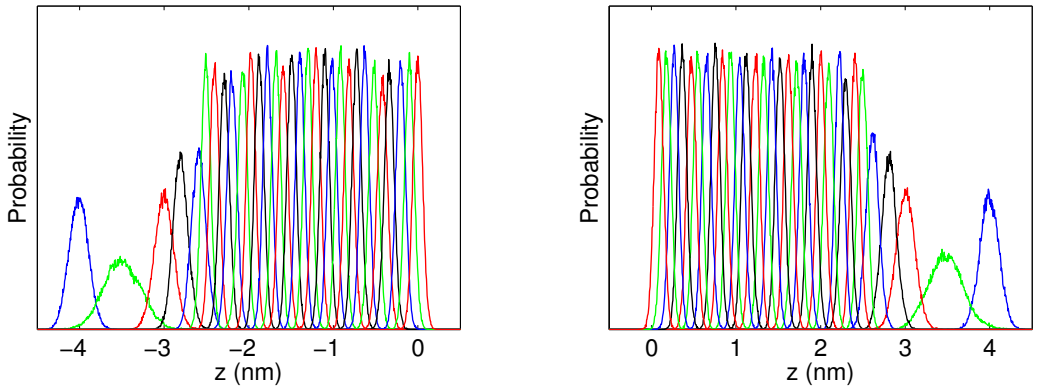


FIG. S20. Position distributions from the US simulations of the neutral (left) and charged (right) phenylbutyrate molecules.

SIII. ANALYSIS METHODS

A. Local estimate of diffusion from umbrella sampling simulations

1. Theoretical background

A strategy commonly followed to determine position-dependent diffusion coefficient is to estimate $D(Z)$ locally from the US simulation centered at $z = Z$. One possibility, which we do not consider here, is to use the time correlation function (TCF) of the restraining force.[7] Another possibility, due to Hummer,[8] is to calculate the TCF of the position of the restrained degree of freedom. If z_Z denotes the z coordinate of the drug that is harmonically restrained at Z, and $\bar{z}_Z = \langle z_Z(t)\rangle_t$ is the average of z_Z along the US trajectory, the position TCF for that umbrella is

$$C_Z(t) = \langle \delta z_Z(\tau + t)\delta z_Z(\tau)\rangle_\tau, \tag{S1}$$

where $\delta z_Z(t) = z_Z(t) - \bar{z}_Z$ is the deviation of z_Z from its average value. Note that the value of this TCF at $t = 0$ gives the variance of z_Z, i.e., $C_Z(0) = \mathrm{Var}(z_Z)$. As shown by Hummer,[8] the local estimate of the diffusion coefficient in a given umbrella is

$$D(Z) = \frac{[C_Z(0)]^2}{\int_0^\infty \mathrm{d}t\, C_Z(t)}. \tag{S2}$$

This expression was derived assuming underdamped Langevin dynamics with a memory kernel in an external harmonic potential.[8–10] It should be emphasized that while the memory kernel is time-dependent, it is not a function of the position. For a drug molecule observed at sufficiently long times (a few ps), the dynamics is expected to be overdamped and memoryless, i.e., Brownian. Since such dynamics can be viewed as a limit of the more general Langevin dynamics assumed in the derivation of (S2), this expression should also apply to Brownian dynamics. Below, we show that (S2) holds for a Brownian particle moving in a harmonic potential with position-independent diffusion constant.

Assuming that the restrained MD motion of the center of mass of the drug head group can be modeled as overdamped diffusion in a harmonic potential, one can write the 1D Smoluchowski equation for a particle attached to a spring with spring constant k_Z held at $\bar{z}_Z$. (Note that on a tilted free energy surface the average $\bar{z}_Z$ is different from the position of the restraint specified in the US simulation and denoted here by Z.) Below, we drop the subscript Z with the understanding that the analysis is applied separately to each US simulation. For a Brownian particle in a harmonic potential the eigenfunctions of the Smoluchowski operator can be written in terms of the Hermit polynomials $H_n(\zeta)$, where $\zeta = \sqrt{\beta k/2}(z - \bar{z})$ and $\beta = 1/RT$.[11–13] Because a Hermit polynomial of order n is an eigenfunction, its TCF decays monoexponentially with a decay time τ_n that is related to the corresponding eigenvalue: $\tau_n = 1/n\beta kD$.[11] In the case of $H_1(\zeta) = 2\zeta = \sqrt{2\beta k}(z - \bar{z})$, for example, one finds $C_1(t) = \langle H_1(\zeta(\tau + t))H_1(\zeta(t))\rangle_\tau = 2\mathrm{e}^{-\beta kDt}$. Therefore, the TCF of $\delta z = (z - \bar{z})$ is

$$C(t) = \langle \delta z(\tau + t)\delta z(\tau)\rangle_\tau = \frac{1}{\beta k}\mathrm{e}^{-\beta kDt}. \tag{S3}$$

The value of this TCF at $t = 0$ is $C(0) = 1/\beta k$, whereas the area under the TCF is

$$\int_0^\infty \mathrm{d}t \frac{1}{\beta k}\mathrm{e}^{-\beta kDt} = \frac{1}{\beta k}\frac{1}{\beta kD}. \tag{S4}$$

Using these on the right hand side of (S2) we have

$$\frac{[C(0)]^2}{\int_0^\infty \mathrm{d}t\, C(t)} = \frac{\frac{1}{\beta^2k^2}}{\frac{1}{\beta^2k^2D}} = D, \tag{S5}$$

in agreement with Hummer's result. The presented derivation, however, comes with the additional requirement for monoexponential decay of the position TCF.

According to (S3), $\beta kC(t)$ should start from 1 at $t = 0$ and decay monoexponentially. In addition, (S3) suggests that $\tilde{t} = \beta kt$ is the natural variable to monitor the decay of the position TCF since

$$\beta kC(\tilde{t}) = \mathrm{e}^{-D\tilde{t}}. \tag{S6}$$

In the following analysis, therefore, we plot the rescaled position TCF, $\tilde{C}(\tilde{t}) = \beta kC(\tilde{t})$, against the rescaled time $\tilde{t}$.

As far as numbers go, $RT = 0.64$ kcal/mol at the temperature of the MD simulations ($T = 323$ K). For a "typical" spring constant $k = 1$ kcal/mol^{-1}Å^2, the scaling factor is $\beta k = 156$ nm^{-2}.

2. *Analysis*

The rescaled position TCFs obtained from the US simulations centered at $z = 4$ nm (ionized dyclonine) and $z = -4$ nm (neutral dyclonine) are drawn with solid lines on the left and on the right of Fig. S21. Mono-, bi-, and tri-exponential fits to the position TCF in the form

$$C^m(t) = \sum_{i=1}^{m} a_i^m \mathrm{e}^{-t/\tau_i^m} \tag{S7}$$

are also drawn in the figure with dashed or dotted lines. Clearly, the TCFs estimated from the US trajectories behave as expected from (S3). Namely, the decay is almost perfectly monoexponential and starts from approximately 1 at $t = 0$. Thus, away from the lipid bilayer the two forms of dyclonine undergo Brownian diffusion with estimated diffusion coefficients of $D = 1.51$ nm^2/ns (ionized) and $D = 1.38$ nm^2/ns (neutral).

The same observations hold true for phenylbutyrate in water whose TCFs calculated from the US simulations at $|z| = 4$ nm are shown in Fig. S22. Comparing the neutral drug forms in Figs. S21 and S22 we see faster decay of the TCFs of F^0, implying a diffusion constant larger than that of D^0. Indeed, the diffusion coefficients of phenylbutyrate estimated from the monoexponential fits to the TCFs are $D = 1.54$ nm^2/ns (ionized) and $D = 2.40$ nm^2/ns (neutral).

Monoexponential fits to the position TCFs from the US simulations at $|z| = 3.5$ nm and $|z| = 3.0$ nm were of satisfactory quality (not shown), leading to the estimates of diffusion coefficients shown in Table S2. These values are shown with filled circles in Figs. 6 and 11 in the main text.

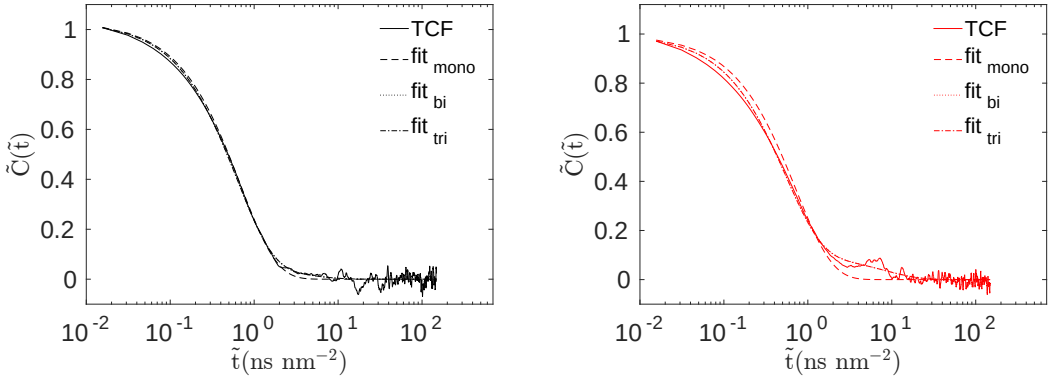


FIG. S21. TCFs calculated from the US simulations restrained at $|z| = 4$ nm (solid lines) and mono- (dashed), bi- (dotted), and tri- (dotted dash) exponential fits for ionized (left) and neutral (right) dyclonines. Both axes are rescaled with k/RT. The TCFs are seen to decay monoexponentially. The decay rates provide estimates of the local diffusion coefficients.

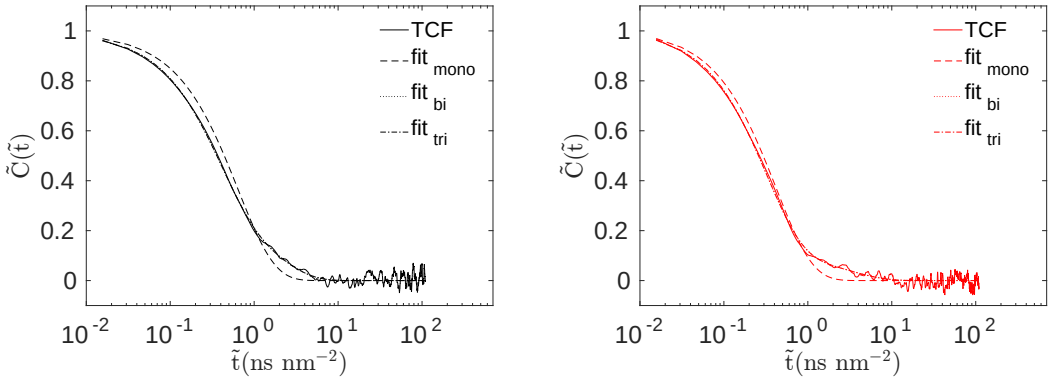


FIG. S22. Same as Fig. S21 for ionized (left) and neutral (right) phenylbutyrates restrained to $|z| = 4$ nm.

TABLE S2. Diffusion constants (in nm^2/ns) estimated from the US simulations of the drugs in water (3 nm $\leq |z| \leq$ 4 nm) by fitting the position TCFs to a single exponential.

$\|z\|$/nm	D^+	D^0	F^-	F^0
4.0	1.51	1.38	1.54	2.40
3.5	1.01	1.26	1.43	1.28
3.0	0.92	1.17	1.10	1.70

In contrast to the situation of drugs in water, the position TCFs calculated from the US simulations with restraining potentials at $|z| < 3$ nm could not be fit satisfactorily with a single exponential function. As an example, mono-, bi-, and tri-exponential fits to the position TCFs from the US simulations with potentials at $|z| = 0$ are shown in Figs. S23 (dyclonine) and S24 (phenylbutyrate). Clearly, decay on at least three different time scales is present. Most likely the observed relaxation on multiple time scales is due to the richer dynamics of the lipids surrounding the drug compared to waters. In any case, deviation of the TCF from monoexponential decay is an indication that overdamped Brownian diffusion is not a realistic model for the dynamics in the MD simulations.

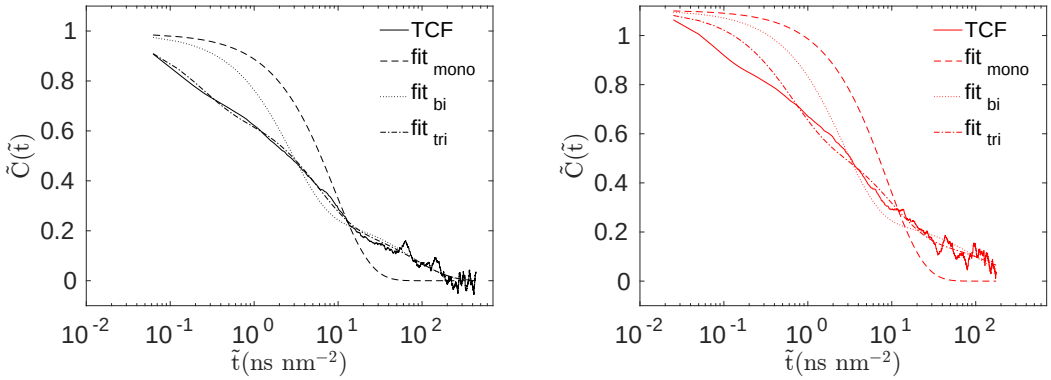


FIG. S23. Rescaled TCFs for umbrella potentials at $z = 0$ nm (solid lines) and fits by mono- (dashed), bi- (dotted) and tri- (dotted dash) exponential functions for charged (left) and neutral (right) dyclonine molecules.

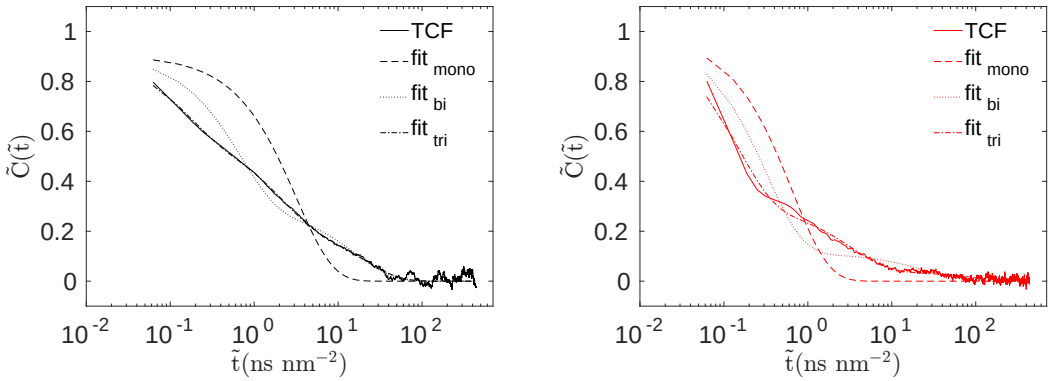


FIG. S24. Same as Fig. S23 for charged (left) and neutral (right) phenylbutyrate molecules.

We quantified the "goodness" of the fits by calculating the difference between the calculated TCF $C_Z(\tilde{t})$ and the best m-exponential fit $C_Z^m(\tilde{t})$ in the interval $t_\mathrm{i} = 0$ ns and $t_\mathrm{f} = 1.4$ ns $\times k/RT$:

$$E^m(Z) = \int_{t_\mathrm{i}}^{t_\mathrm{f}} |C_Z(\tilde{t}) - C_Z^m(\tilde{t})| \mathrm{d}\tilde{t}. \tag{S8}$$

When calculating the integral numerically, the rescaled time axis ($\tilde{t}$) was discretized on a logarithmic scale in order to increase the contribution of the points in the initial fast decaying part of the TCF. Calculated errors, $E^m(Z)$, for

dyclonine and phenylbutyrate are shown in Figs. S25 and S26. It is seen that while the errors of the monoexponential fit are fairly low in the aqueous environment, they increase substantially for the umbrella potentials at $|z| \leq 2.8$ nm (Figs. S25 and S26, dashed lines). In this region, fits to bi- and tri- exponential functions are better than fits to a single exponential decay.

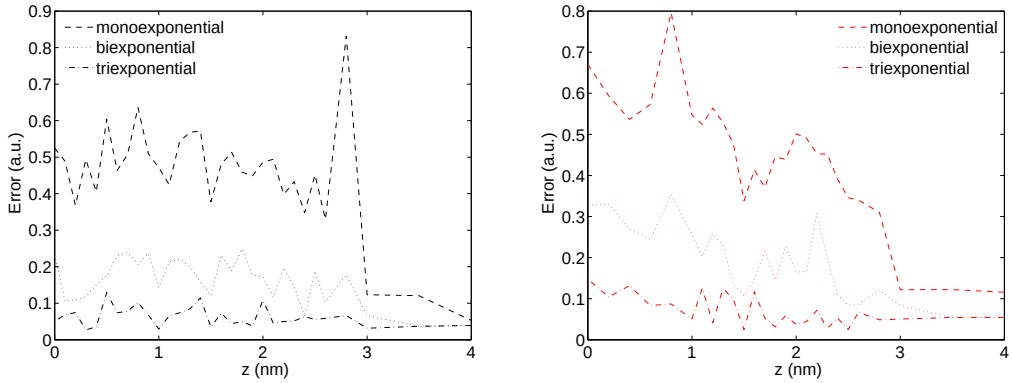


FIG. S25. Goodness-of-fit (error) of the mono- (dashed), bi- (dotted) and tri- (dotted dashed) exponential fits to the TCF obtained from the US simulations for charged (left) and neutral (right) dyclonine.

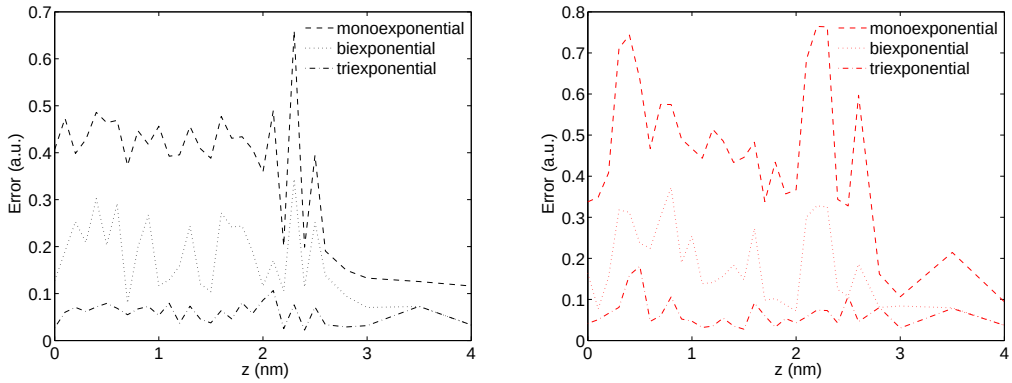


FIG. S26. Same as Fig. S25 for charged (left) and neutral (right) phenylbutyrate.

When the TCF is in the multiexponential form (S7), the estimate of the diffusion coefficient in (S2) becomes

$$D^m(Z) = \frac{\left(\sum_{i=1}^m a_i^m(Z)\right)^2}{\sum_{i=1}^m a_i^m(Z)\tau_i^m(Z)}, \tag{S9}$$

where we have explicitly shown the dependence of the fitting coefficients on Z. Diffusion coefficients calculated according to (S9) for all mono-, bi- and tri-exponential fits are show in Figs. S27 (dyclonine) and S28 (phenylbutyrate). In both figures, on the left are estimates for the ionized drug form (drawn in black) and on the right are estimates for the neutral form (in red). The diffusion coefficients estimated from monoexponential fits are systematically larger than the ones calculated from bi- and tri-exponential fits. On the other hand, diffusion coefficients estimated from bi- and tri-exponential fits are not much different from each other, in spite of the differences in their fitting errors (Figs. S25 and S26). Therefore, for the diffusion coefficients of the drugs in lipid ($|z| \leq 2.8$ nm) we take the numerical values calculated from the biexponential fits. These are indicated with empty circles in Figs. 6 and 11 of the main text showing the estimated diffusion coefficients.

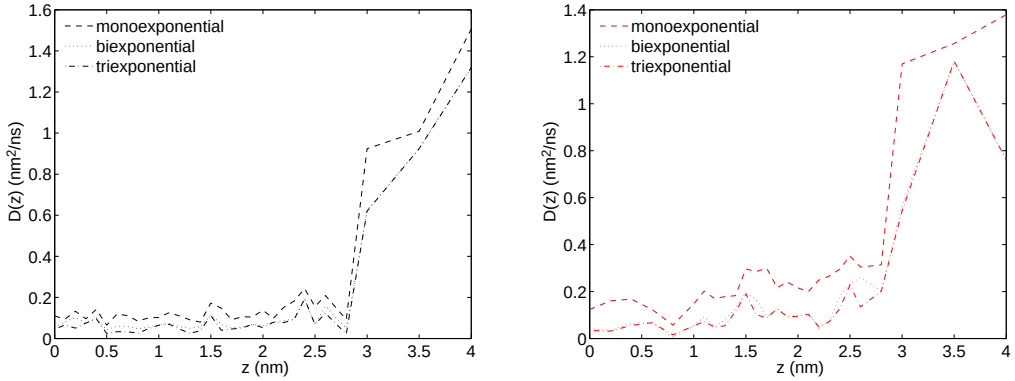


FIG. S27. Diffusion coefficients calculated from mono- (dashed), bi- (dotted) and tri- (dotted dashed) exponential functions for charged (black, left) and neutral (red, right) dyclonine are shown.

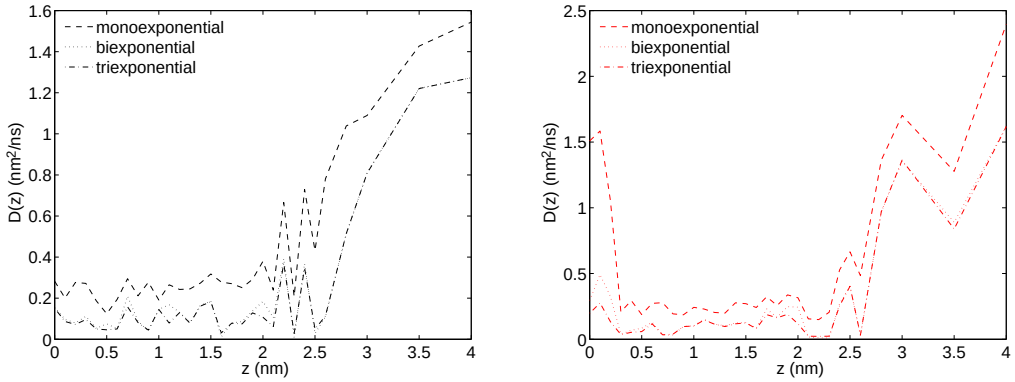


FIG. S28. Diffusion coefficients calculated from mono- (dashed), bi- (dotted) and tri- (dotted dashed) exponential functions for charged (black, left) and neutral (red, right) phenylbutyrate are shown.

B. Global estimate of diffusion from unrestrained simulations

1. *Theoretical background and notation*

a. Mean first passage times The mean first passage time (MFPT) $T_{z\to x}$ is the average time to reach a point x for the first time, starting a diffusive motion from point z. While by definition x is reached only once, the diffusing particle may cross the point z many times during its motion towards x. Various expressions for MFPTs with position-dependent diffusion in a potential are available depending on the nature (reflecting or absorbing) of the boundaries to which the diffusion process is restricted.[13]

For $a < z < x$ with a reflecting boundary at a and an absorbing boundary at x the MFPT is[13]

$$T_{z\to x}^{|a,x\rangle} = \int_z^x \mathrm{d}\zeta \, \frac{\mathrm{e}^{u(\zeta)}}{D(\zeta)} \int_a^\zeta \mathrm{d}\eta \, \mathrm{e}^{-u(\eta)}. \tag{S10}$$

[To keep track of the nature of the boundaries, we have written the interval (a, x) as $|a, x\rangle$, with the understanding that $|$ indicates a reflecting and $\rangle$ an absorbing boundary.] Note that, because this expression contains the product of $\mathrm{e}^{u(\zeta)}$ and $\mathrm{e}^{-u(\eta)}$, the arbitrary zero reference of the free energy $u(z)$ is immaterial, as long as the same reference point is used consistently. Similarly, for $x < z < b$ with an absorbing boundary at x and a reflecting boundary at b the MFPT is[13]

$$T_{x\leftarrow z}^{\langle x,b|} = \int_x^z \mathrm{d}\zeta \, \frac{\mathrm{e}^{u(\zeta)}}{D(\zeta)} \int_\zeta^b \mathrm{d}\eta \, \mathrm{e}^{-u(\eta)}. \tag{S11}$$

Netz and coworkers have introduced the round-trip time (RTT), which is the MFPT to go from z to x plus the

MFPT to go from x to z.[14] For $a < x < z < b$, with reflecting boundaries at a and b the RTT is

$$T_{x\leftrightharpoons}^{|a,b|}(z) = T_{x\leftarrow z}^{\langle x,b|} + T_{x\to z}^{|a,z\rangle} = Z_{ab}\int_x^z \mathrm{d}\zeta\, \frac{\mathrm{e}^{u(\zeta)}}{D(\zeta)}, \tag{S12}$$

where

$$Z_{ab} = \int_a^b \mathrm{d}\eta\, \mathrm{e}^{-u(\eta)}. \tag{S13}$$

Expression (S12) can be used to estimate the profile of the diffusion coefficient:[14]

$$D_{x\leftrightharpoons}^{|a,b|}(z) = \frac{Z_{ab}\mathrm{e}^{u(z)}}{\mathrm{d}T_{x\leftrightharpoons}^{|a,b|}(z)/\mathrm{d}z}. \tag{S14}$$

Similarly, when z is between a and x ($a < z < x < b$), the RTT is

$$T_{\rightleftharpoons x}^{|a,b|}(z) = T_{z\to x}^{|a,x\rangle} + T_{z\leftarrow x}^{\langle z,b|} = Z_{ab}\int_z^x \mathrm{d}\zeta\, \frac{\mathrm{e}^{u(\zeta)}}{D(\zeta)}, \tag{S15}$$

which can be used to obtain the following estimate of the diffusion profile:[14]

$$D_{\rightleftharpoons x}^{|a,b|}(z) = -\frac{Z_{ab}\mathrm{e}^{u(z)}}{\mathrm{d}T_{\rightleftharpoons x}^{|a,b|}(z)/\mathrm{d}z}. \tag{S16}$$

In our analysis, we used the drug trajectories in the unrestrained MD simulations to obtain estimates of the RTTs $T_{x\leftrightharpoons}^{|a,b|}(z)$ and $T_{\rightleftharpoons x}^{|a,b|}(z)$ for several values of $z \in (a, b)$. When estimating $T_{x\leftrightharpoons}^{|a,b|}(z)$ we chose $x = a$, whereas for $T_{\rightleftharpoons x}^{|a,b|}(z)$ we chose $x = b$. In other words, we took the point x to coincide with either the lower or the upper boundary of the interval (a, b). In this way, the RTTs depend on three variables only: a, b and z.

When plotting the MFPTs in Sec. Siii B 2, the boundaries of the interval are clear from the range of the horizontal axis of the plot. Therefore, to simplify the notation, we do not write the interval boundaries as a superscript. In particular, instead of $T_{a\leftarrow z}^{\langle a,b|}$ and $T_{a\to z}^{|a,z\rangle}$ that appear in (S12) we write $T_{a\leftarrow}(z)$ and $T_{a\to}(z)$, respectively. By writing z in parenthesis, rather than placing it at one end of the arrow, we focus on the dependence of the RTTs on z. With the same logic, instead of $T_{z\to b}^{|a,b\rangle}$ and $T_{z\leftarrow b}^{\langle z,b|}$ that appear in (S15) we write $T_{\to b}(z)$ and $T_{\leftarrow b}(z)$. Again the variable z that is written in parenthesis should be thought of as sitting at the empty side of the arrow.

In Sec. Siii B 2, the RTTs involving the lower boundary of the analysis interval is denoted by $T_{\leftarrow}$, i.e.,

$$T_{\leftarrow}(z) = T_{a\leftarrow}(z) + T_{a\to}(z). \tag{S17}$$

The diffusion profile calculated from this RTT according to (S14) is denoted by $D_{\leftarrow}(z)$. Similarly, the RTTs involving the upper boundary of the interval is denoted by $T_{\to}$:

$$T_{\to}(z) = T_{\to b}(z) + T_{\leftarrow b}(z). \tag{S18}$$

The diffusion profile calculated from this RTT according to (S16) is denoted by $D_{\to}(z)$. This notation for the diffusion profiles is used in Figs. 6 and 11 of the main text.

b. Transition path time Unlike the MFPT where the starting point can be re-crossed many times, the transition path time (TPT) $T_{a\to b}^{\mathrm{TP}}$ is the average time to reach b for the first time, since leaving a for the last time. (See Ref. 15 for a clear discussion and a nice drawing of a first-passage path and a transition path.) Assuming a Brownian motion with a position-independent diffusion constant, D, in a dimensionless potential, $u(z)$, where $z \in (a, b)$, the TPT is[15,16]

$$T_{a\to b}^{\mathrm{TP}} = \frac{1}{D}Q_{ab}\int_a^b \mathrm{d}z\, q(z)(1-q(z))\mathrm{e}^{-u(z)}. \tag{S19}$$

Here,

$$Q_{ab} = \int_a^b \mathrm{d}z\, \mathrm{e}^{u(z)} \tag{S20}$$

and

$$q(z) = \frac{1}{Q_{ab}} \int_z^b \mathrm{d}\zeta \, \mathrm{e}^{u(\zeta)}. \tag{S21}$$

Given a free energy profile and a TPT that is estimated from transitions occurring spontaneously in the unrestrained MD simulations, (S19) can be used to obtain a diffusion constant:

$$D = \frac{1}{T^{\mathrm{TP}}_{a\to b}} Q_{ab} \int_a^b \mathrm{d}z \, q(z)(1-q(z))\mathrm{e}^{-u(z)}. \tag{S22}$$

The MD simulations $4\mathrm{F}^0$ contained four independent processes of drug translocation from one lipid bilayer leaflet to the other (Fig. 8B in the main text). The corresponding transition paths, defined as the trajectory fragments connecting the points −1.3 nm and 1.3 nm, are depicted in Fig. 9 of the main text.

2. Analysis

Round trip times (RTTs) for a selected interval (a, b) were estimated by identifying first-passage events in the MD trajectories and accumulating the first passage times. To estimate error bars, the observed events were divided into three groups, with approximately equal number of events in each group. Three estimates of the MFPT were produced by averaging the first passage times in each group. A standard deviation was calculated from these three estimates. Finally, a single MFPT is obtained by averaging the three MFPT estimates. This procedure was first used to estimate the MFPTs $T_{a\leftarrow}(z)$ and $T_{a\rightarrow}(z)$ and their standard deviations $\sigma_{a\leftarrow}(z)$ and $\sigma_{a\rightarrow}(z)$. The RTT $T_{\leftarrow}(z)$ was estimated from (S17) and its standard deviation was obtained as

$$\sigma_{\leftarrow}(z) = \sqrt{(\sigma_{a\leftarrow}(z))^2 + (\sigma_{a\rightarrow}(z))^2}. \tag{S23}$$

Then, the same analysis was performed to obtain the MFPTs $T_{\rightarrow b}(z)$ and $T_{\leftarrow b}(z)$, and their standard deviations $\sigma_{\rightarrow b}(z)$ and $\sigma_{\leftarrow b}(z)$. The former were added to produce the RTT $T_{\rightarrow}(z)$ according to (S18), and the latter gave its standard deviation

$$\sigma_{\rightarrow}(z) = \sqrt{(\sigma_{\rightarrow b}(z))^2 + (\sigma_{\leftarrow b}(z))^2}. \tag{S24}$$

When estimating the diffusion profiles of dyclonine and phenylbutyrate from the RTTs, the interval $(-L, L)$ was divided into four separate regions as shown in Table S3. Benefitting from the symmetry around $z = 0$, for regions II, III and IV statistics obtained for the intervals (a, b) and $(-a, -b)$ were treated together, although only the interval (a, b) is written in Table S3. In the case of dyclonine, the drug trajectories in the simulation $4\mathrm{D}^0$ were used for region I, while those in $8\mathrm{D}^+$ were used for regions II, III and IV. RTT analysis was applied to all four regions, however the sampling for region III was poor and did not lead to a reliable estimate of the diffusion. For phenylbutyrate, RTT analysis of the simulation $8\mathrm{F}^-$ was used to estimate diffusion profiles in regions II, III and IV (Table S3). In region I, a diffusion constant was obtained using the transition path time (TPT) estimated from the four translocation events observed in the simulation $4\mathrm{F}^0$.

Mean first passage times (MFPTs), and subsequently RTTs, were estimated from the trajectories with a resolution of 0.05 nm.

TABLE S3. Information about the global estimate of diffusion profiles.

	Dyclonine			Phenylbutyrate		
region	interval	MD simulation	method	interval	MD simulation	method
I	$(-1.4, 1.4)$	$4\mathrm{D}^0$	RTT	$(-1.3, 1.3)$	$4\mathrm{F}^0$	TPT
II	$(1.4, 2.6)$	$8\mathrm{D}^+$	RTT	$(1.5, 2.5)$	$8\mathrm{F}^-$	RTT
III	$(2.6, 3.0)$	$8\mathrm{D}^+$	-	$(2.5, 4.0)$	$8\mathrm{F}^-$	RTT
IV	$(3.0, 4.0)$	$8\mathrm{D}^+$	RTT	$(3.0, 4.0)$	$8\mathrm{F}^-$	RTT

a. Dyclonine Figure S29 shows the MFPTs to reach the lower boundary of region I starting from z (left), and to reach back to z starting from the lower boundary (right). Overall, the MFPTs are monotonically increasing in both plots. This is actually required since longer time is needed to reach for the first time a point that is further away from the starting point.

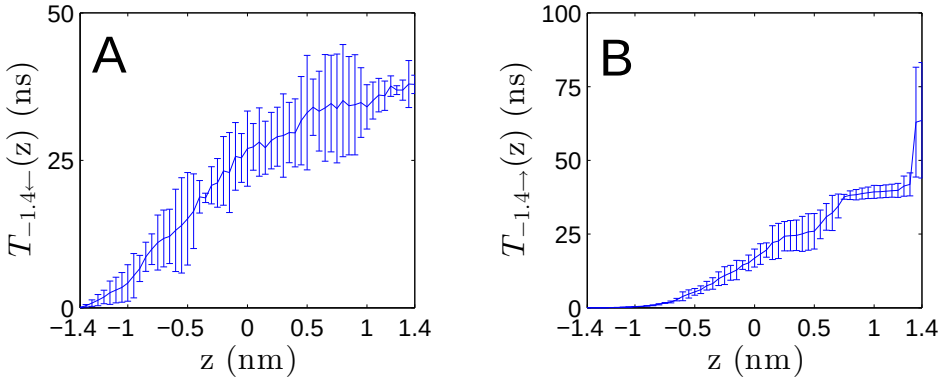


FIG. S29. (A) MFPTs $T_{-1.4\leftarrow}(z)$ to reach the lower boundary of the interval $(-1.4, 1.4)$ nm starting from z. (B) MFPTs $T_{-1.4\rightarrow}(z)$ to reach z starting from the lower boundary of the same interval. In (A) the upper boundary of the interval is treated as reflecting, whereas in (B) the lower boundary is treated as reflecting.

Two types of artifacts arising from the limited sampling of the drug dynamics are worth pointing out already at this stage. The first relates to the monotonic increase of the MFPT. Looking carefully at the estimates of $T_{-1.4\leftarrow}(z)$ on the left of Fig. S29 we see several cases where the value at a given z is slightly smaller than the neighbouring value at a smaller z. Such decrease of the estimated MFPT with increasing z is very problematic for the estimate of the diffusion constant, as will become apparent below. A second artifact, seen in the estimated $T_{-1.4\rightarrow}(z)$ on the right of Fig. S29, is the jump between $z = 1.3$ nm and $z = 1.4$ nm. Looking only at the estimated values of the MFPTs, it appears that it takes much longer for the drug to reach $z = 1.4$ nm for the first time starting from $a = -1.4$ nm, than it takes to reach $z = 1.3$ nm. However, the comparatively large error bar at $z = 1.4$ nm indicates that very few events have been observed in the former case. This particular sampling artifact results in a large slope of the MFPT, which the analysis interprets as being due to a very small diffusion constant.

The sum of $T_{-1.4\leftarrow}(z)$ and $T_{-1.4\rightarrow}(z)$ produced the RTTs given on the left of Fig. S30. Similar to the situation with MFPTs, some RTT values are smaller than the preceding ones, which translates into a negative slope of the RTTs. Such negative slope is not realistic since it predicts a negative diffusion coefficient. To avoid this behaviour, the estimated RTTs were smoothed by averaging over five (blue line) or seven (black line) successive points. The local diffusion coefficients calculated from (S14) are plotted on the right of Fig. S30. The estimated diffusivity is large where the slope of the RTT is small. Such regions appear as spikes in the diffusion profile. In some cases, increasing the averaging window reduces the height of the spikes. It is frustrating that almost identical RTTs, exemplified by the lines 's5' and 's7' on the right of Fig. S30, may lead to pronounced differences in the diffusion profile.

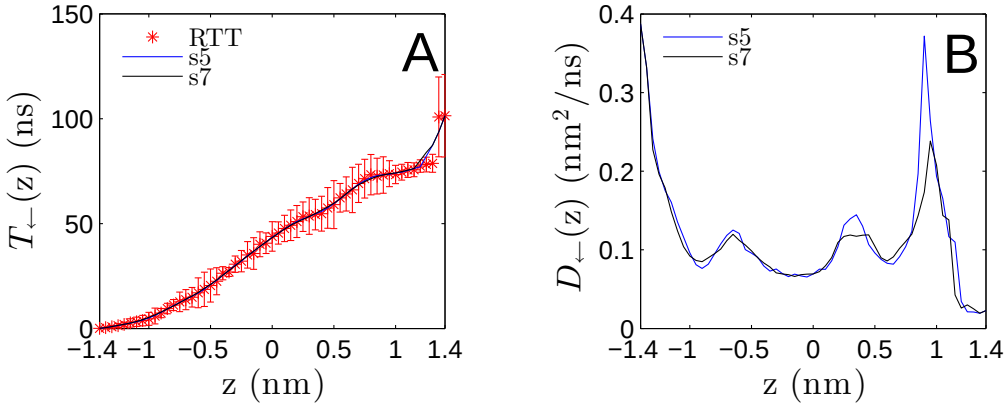


FIG. S30. (A) The round trip time $T_{\leftarrow}(z)$ (symbols) and its smoothed versions obtain by averaging over 5 (blue line) and 7 (black line) neighbouring points. (B) Diffusion profiles obtained from the smoothed RTTs according to (S14).

The drug dynamics in the same interval (region I in Table S3) was also analyzed in terms of the RTTs involving the upper boundary of the region. The MFPTs estimated from the MD trajectories are given in Fig. S31. Again the estimated mean time to reach the boundary $b = 1.4$ nm for the first time starting from z is shown on the left, and the MFPT to reach z starting from b is on the right. While overall the MFPTs decrease monotonically with increasing z, we see the same types of artifacts where the slope is inverted. The sharp decrease of the estimated $T_{\leftarrow 1.4}(z)$ around $z = -1.4$ is particularly troublesome.

The RTTs calculated from the MFPTs in Fig. S31 are shown on the left of Fig. S32. Clearly, smoothing the RTTs by averaging over 5 (s5) or 7 (s7) successive points does not help to remove the sharp turn at the lower end of the interval. As a result, the diffusion profile obtained from the slopes of these smoothed curves using (S16) exhibits negative values in this part of the analysis interval (Fig. S31, right panel).

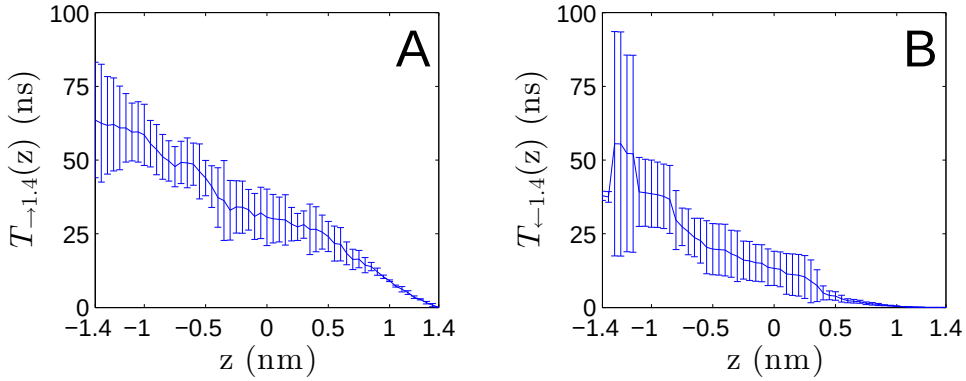


FIG. S31. (A) MFPTs $T_{\rightarrow 1.4}(z)$ to reach the upper boundary of $(-1.4, 1.4)$ nm starting from z. (B) MFPTs $T_{\leftarrow 1.4}(z)$ to reach z starting from the upper boundary of the same interval. In (A) the lower boundary of the interval is treated as reflecting, whereas in (B) the upper boundary is treated as reflecting. The calculation of the error bars, which correspond to one standard deviation, is described in the text above.

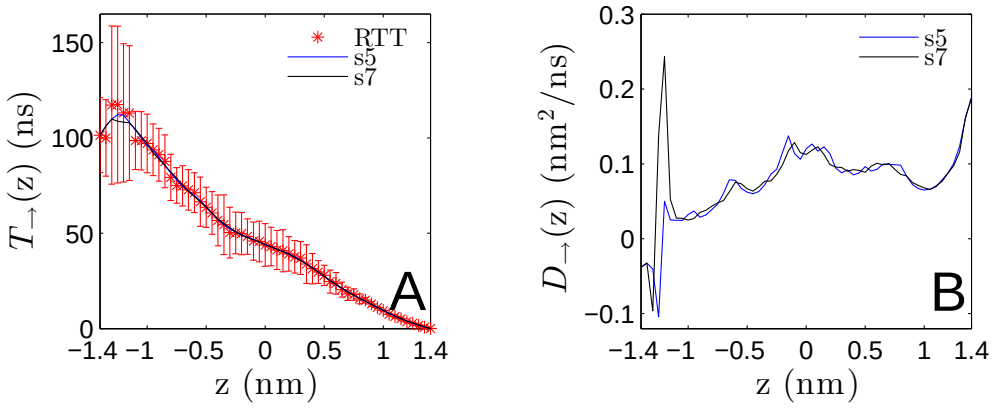


FIG. S32. (A) The round trip time $T_{\rightarrow}(z)$ (symbols), together with its 5-point (s5) and 7-point (s7) smoothed forms. (B) Diffusion profiles obtained from the smoothed RTTs according to (S16).

In Fig. S33 we plot the diffusivity profiles $D_{\leftarrow}$ (Fig. S30, right) and $D_{\rightarrow}$ (Fig. S32, right), obtained from the 7-point smoothing, with dashed and solid black lines, respectively. The average of the $z < 0$ part of $D_{\leftarrow}$ and the $z > 0$ part of $D_{\leftarrow}$, i.e., $(D_{\leftarrow}(z) + D_{\leftarrow}(-z))/2$, is plotted as $D_{\leftarrow}$ in Fig. 6 of the main text. Similarly, $(D_{\rightarrow}(z) + D_{\rightarrow}(-z))/2$ is plotted as $D_{\rightarrow}$ in the same figure.

To assess the sensitivity of the analysis on the choice of the interval, we repeated the same steps for the interval $(-1, 1)$. The deduced diffusivity profiles $D_{\leftarrow}$ and $D_{\rightarrow}$ are shown in Fig. S33 with dashed and solid red lines, respectively. It is seen that all the obtained diffusion profiles are consistent with each other.

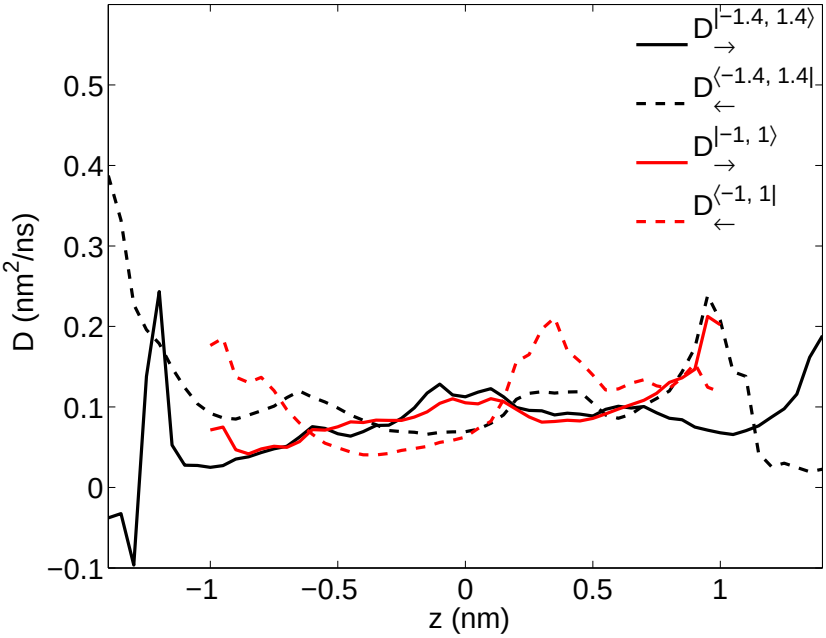


FIG. S33. Diffusion profiles estimated from RTT analysis in the intervals $(-1.4, 1.4)$ nm (black) and $(-1, 1)$ nm (red).

The same analysis was repeated for region II in Table S3. Remember that this involved collecting first-passage events form the MD trajectories simultaneously for the interval $(1.4, 2.6)$ nm as well as $(-1.4, -2.6)$ nm. The results are given in Fig. S34 (lower boundary) and Fig. S35 (upper boundary).

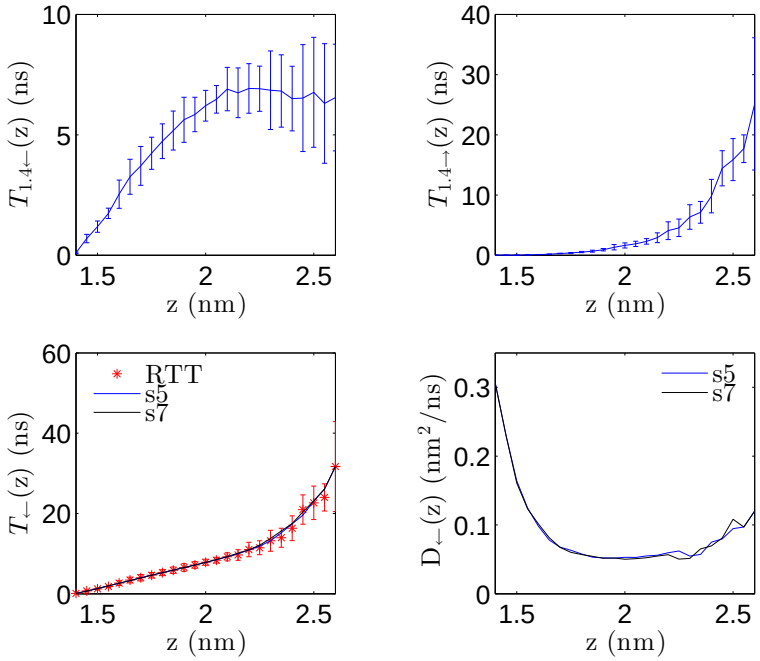


FIG. S34. MFPTs to and from the lower boundary of the interval (top), RTTs (bottom, left), and estimated diffusion profiles (bottom, right) for region II. The profile $D_{\leftarrow}$ calculated from 's5' is shown in Fig. 6 of the main text.

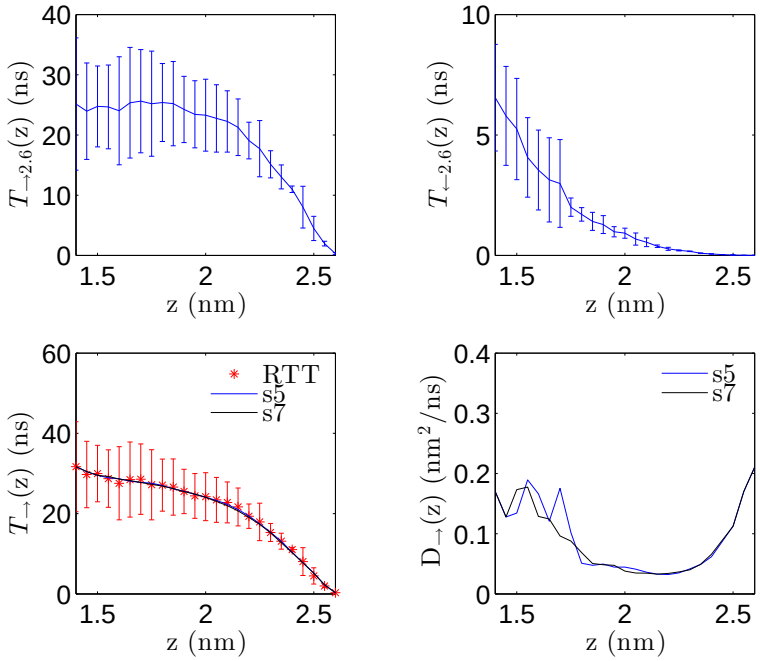


FIG. S35. MFPTs to and from the upper boundary of the interval (top), RTTs (bottom, left), and estimated diffusion profiles (bottom, right) for region II. The profile $D_{\rightarrow}$ calculated from 's5' is shown in Fig. 6 of the main text.

We could not determine a diffusion profile for region III since we did not have enough sampling to determine reliable RTTs in this interval. For the estimation of RTTs we need to have events in which molecules move in both directions

along bilayer normal. However as we did not see any dissociation events in the simulation $8D^+$, there was no molecule moving in both directions within this interval.

Lastly, the RTT analysis was applied to IV. The results are given in Fig. S36 (lower boundary) and Fig. S37 (upper boundary).

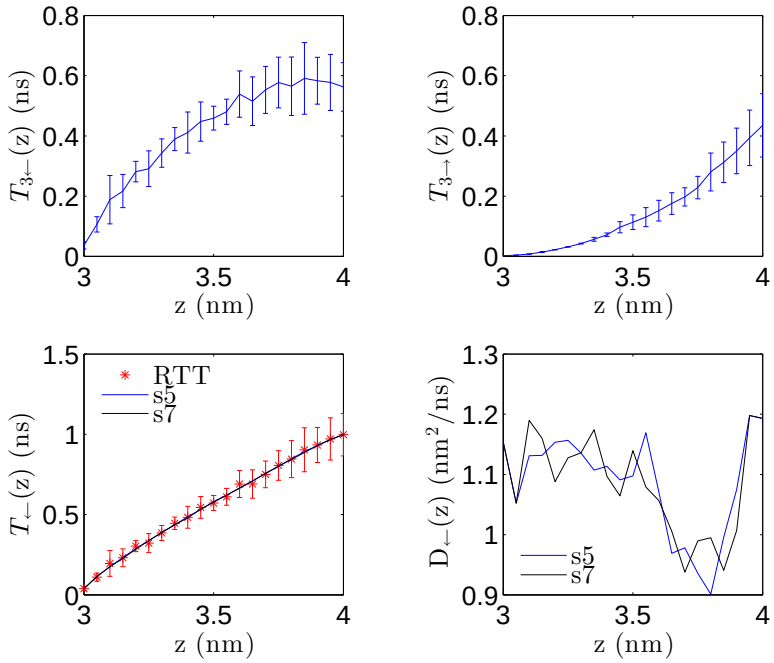


FIG. S36. MFPTs to and from the lower boundary of the interval (top), RTTs (bottom, left), and estimated diffusion profiles (bottom, right) for region IV. The profile $D_{\rightarrow}$ calculated from 's5' is shown in Fig. 6 of the main text.

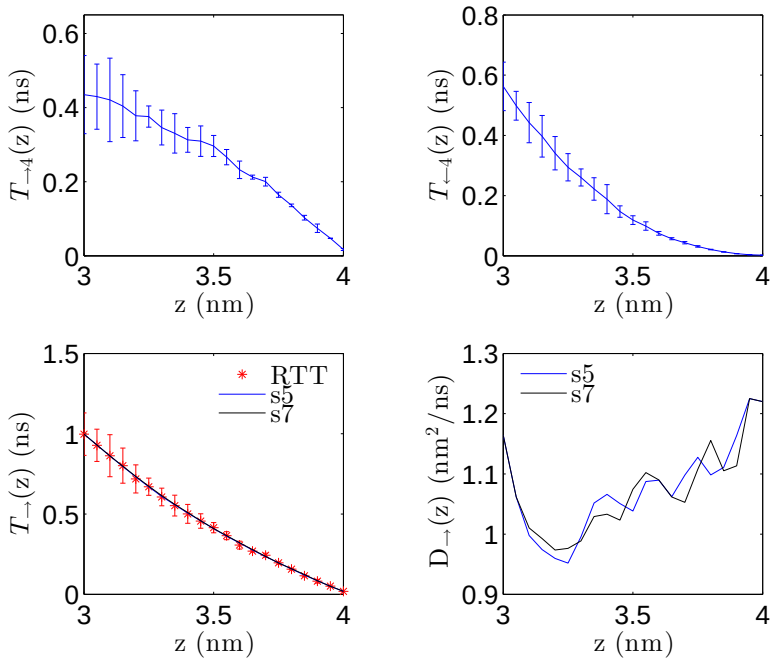


FIG. S37. MFPTs to and from the upper boundary of the interval (top), RTTs (bottom, left), and estimated diffusion profiles (bottom, right) for region IV. The profile $D_{\rightarrow}$ calculated from 's5' is shown in Fig. 6 of the main text.

b. Phenylbutyrate RTT analysis was also used to determine the diffusivity profile of phenylbutyrate for $|z| > 1.5$ nm. The results for region II (Table S3) are given in Fig. S38 (lower boundary at 1.5 nm) and Fig. S39 (upper boundary at 2.5 nm).

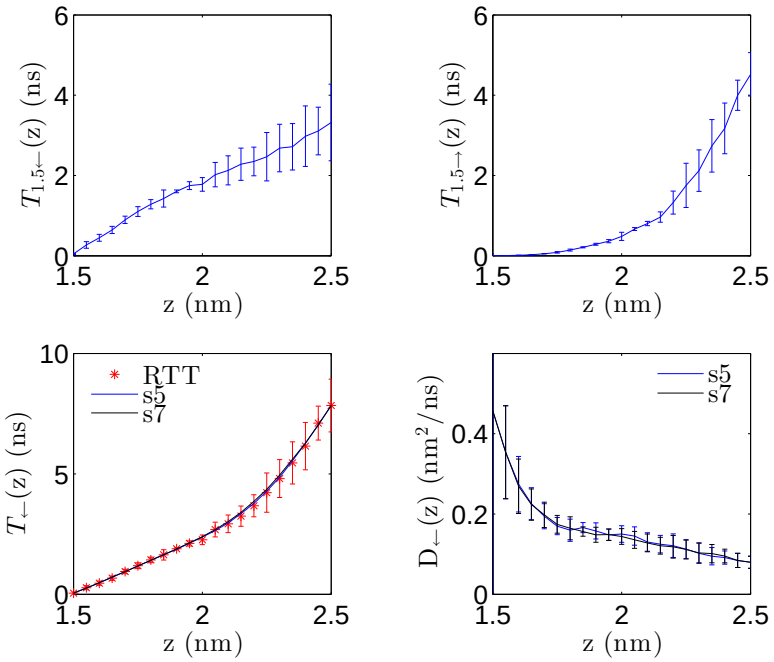


FIG. S38. MFPTs to and from the lower boundary of the interval (top), RTTs (bottom, left), and estimated diffusion profiles (bottom, right) for region II of phenylbutyrate. The profile $D_{\rightarrow}$ calculated from 's5' is shown in Fig. 11 of the main text.

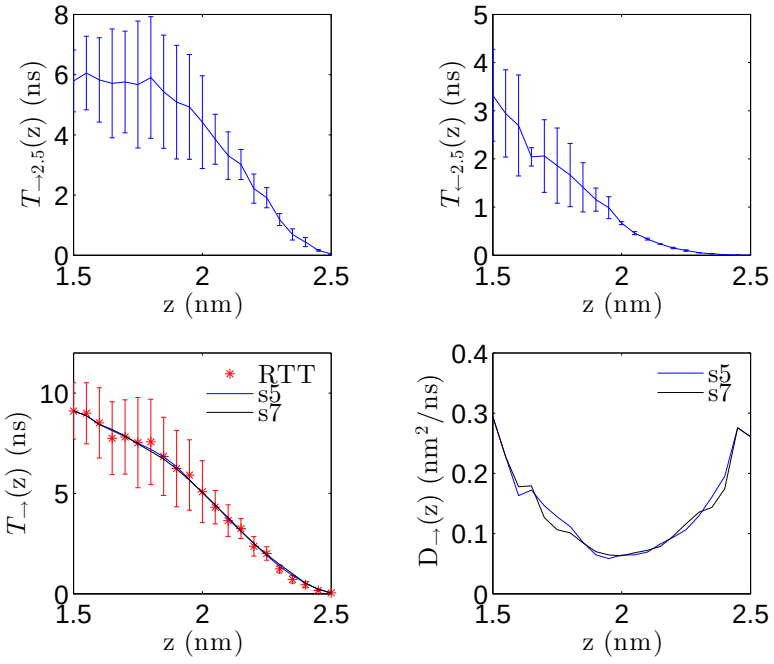


FIG. S39. MFPTs to and from the upper boundary of the interval (top), RTTs (bottom, left), and estimated diffusion profiles (bottom, right) for region II. The profile $D_{\rightarrow}$ calculated from 's5' is shown in Fig. 11 of the main text.

The results for region III (Table S3) are given in Fig. S40 (lower boundary at 2.5 nm) and Fig. S41 (upper boundary at 4 nm). Note that, in contrast to dyclonine, we could determine the diffusion coefficients in this range since phenylbutyrates dissociated from the lipid bilayer in the unrestrained MD simulations $8F^-$.

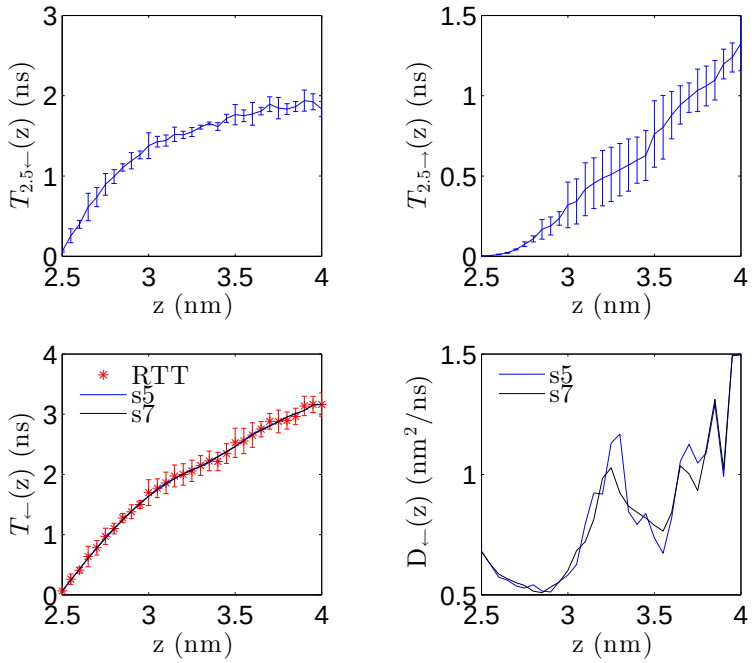


FIG. S40. MFPTs to and from the lower boundary of the interval (top), RTTs (bottom, left), and estimated diffusion profiles (bottom, right) for region III. The profile $D_{\rightarrow}$ calculated from 's5' is shown in Fig. 11 of the main text.

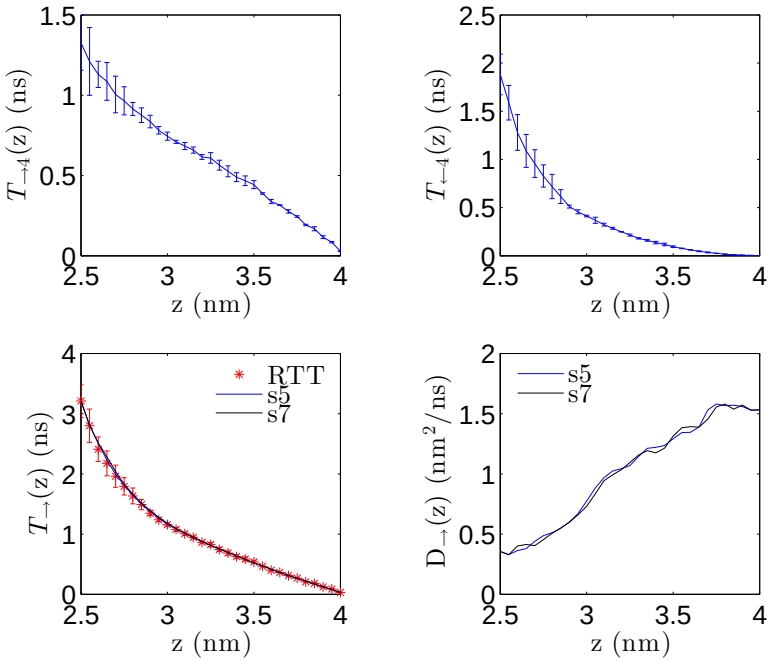


FIG. S41. MFPTs to and from the upper boundary of the interval (top), RTTs (bottom, left), and estimated diffusion profiles (bottom, right) for region III. The profile $D_{\rightarrow}$ calculated from 's5' is shown in Fig. 11 of the main text.

We also analyzed drug dynamics in region IV, which is a subregion of region III (Table S3). The results are given in Fig. S42 (lower boundary) and Fig. S43 (upper boundary). Apart from statistical fluctuations, the deduced diffusion profiles are consistent with those obtained for region III. Thus, the exact positions of the interval boundaries do not change the estimated diffusion coefficients.

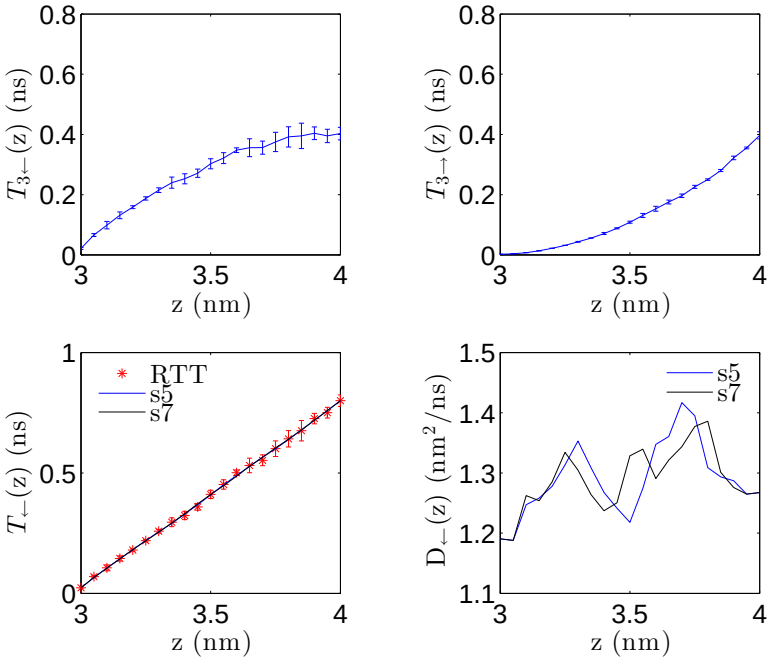


FIG. S42. MFPTs to and from the lower boundary of the interval (top), RTTs (bottom, left), and estimated diffusion profiles (bottom, right) for region IV. The profile $D_{\rightarrow}$ calculated from 's5' is shown in Fig. 11 of the main text.

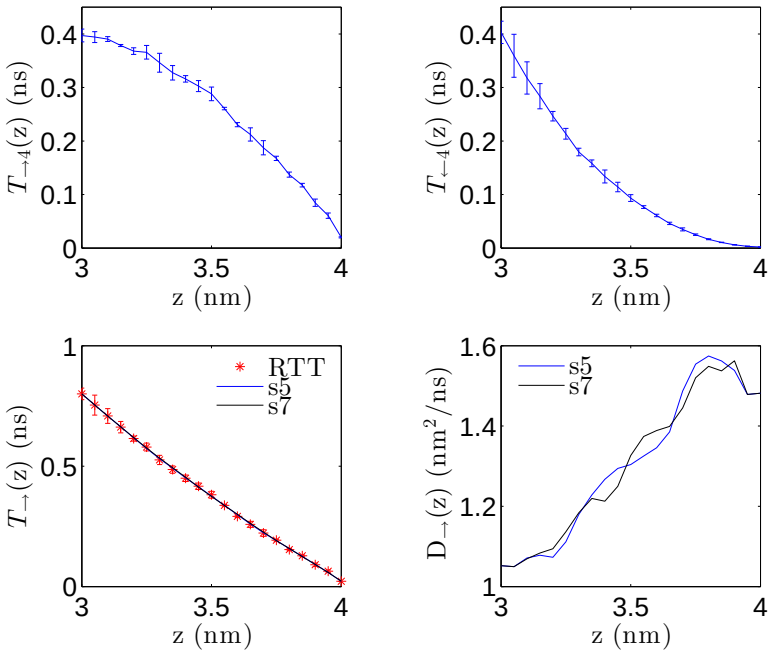


FIG. S43. MFPTs to and from the upper boundary of the interval (top), RTTs (bottom, left), and estimated diffusion profiles (bottom, right) for region IV. The profile $D_{\rightarrow}$ calculated from 's5' is shown in Fig. 11 of the main text.

C. Analytical profiles of partition coefficient

1. Translocation

By definition, the partition coefficient

$$K_{\mathrm{T}}(z) = \mathrm{e}^{-\frac{G(z)-G(-h)}{RT}} \tag{S25}$$

should satisfy the boundary conditions $K_{\mathrm{T}}(\pm h) = 1$. Requiring a profile that is symmetric about $z = 0$, for $z > 0$ we write

$$K_{\mathrm{T}}(z) = \begin{cases} 1, & z_1 \le z < h \\ K^{\ddagger} + m(z - z_2), & z_2 \le z < z_1 \\ K^{\ddagger}, & 0 \le z < z_2 \end{cases} \tag{S26}$$

In (S26), $z_1 = (1-\alpha)h$ and $z_2 = \beta h$ are the points where the slope of $K_{\mathrm{T}}(z)$ changes from zero to $m = (1-K^{\ddagger})/(\gamma h)$, with $\gamma = 1-\alpha-\beta$. The parameters α and β are positive numbers between 0 and 1. They correspond to the fractions of the interval $(-h, h)$ in which K_{T} is equal to, respectively, 1 and $K^{\ddagger}$ (see Fig. 12B in main text). Similarly, γ is the fraction of this interval where K_{T} increases or decreases linearly.

2. Insertion

In this case we are interested in the partition coefficient

$$K_{\mathrm{I}}(z) = \mathrm{e}^{-\frac{G(z)-G(-L)}{RT}} \tag{S27}$$

in the range $(-L, -h)$. The piecewise linear form shown in Fig. 12B of the main text has the form

$$K_{\mathrm{I}}(z) = \begin{cases} 1, & -L \le z < z_3 \\ 1 + n(z - z_3), & z_3 \le z < z_4 \\ K, & z_4 \le z < -h \end{cases} \tag{S28}$$

In terms of the length of the interval, $\ell = L - h$, the points where $K_{\mathrm{I}}(z)$ changes slope are $z_3 = -L + \delta\ell$ and $z_4 = -h - \epsilon\ell$. As before, the non-zero slope is $n = (K-1)/(\zeta\ell)$ with $\zeta = 1-\delta-\epsilon$, and δ and ϵ are the interval fractions where $K_{\mathrm{I}}(z)$ is equal to, respectively, 1 and K (Fig. 12B).

---

* dsezer@sabanciuniv.edu

[1] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics **14**, 33 (1996).
[2] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, J. Comp. Chem. **26**, 1781 (2005).
[3] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, and R. W. Pastor, J. Phys. Chem. B **114**, 7830 (2010).
[4] J. B. Klauda, V. Monje, T. Kim, and W. Im, J. Phys. Chem. B **116**, 9424 (2012).
[5] S. Jo, T. Kim, V. G. Iyer, and W. Im, J. Comp. Chem. **29**, 1859 (2008).
[6] M. Souaille and B. Roux, Computer Physics Communications **135**, 40 (2001).
[7] A. E. Cardenas, R. Shrestha, L. J. Webb, and R. Elber, J. Phys. Chem. B **119**, 6412 (2015).
[8] G. Hummer, New J. Phys. **7**, 34 (2005).
[9] J. E. Straub, M. Borkovec, and B. J. Berne, J. Phys. Chem. **91**, 4995 (1987).
[10] T. B. Woolf and B. Roux, J. Am. Chem. Soc. **116**, 5916 (1994).
[11] K. Schulten and I. Kosztin, "Lectures in theoretical biophysics," (2000).
[12] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, 2nd ed. (Springer, 1996).
[13] C. W. Gardiner, *Handbook of Stochastic Mathods for Physics, Chemistry and the Natural Sciences*, 3rd ed. (Springer, 2004).
[14] M. Hinczewski, Y. Von Hansen, J. Dzubiella, and R. R. Netz, J. Chem. Phys. **132**, 245103 (2010).
[15] J. O. Daldrop, W. K. Kim, and R. R. Netz, EPL (Europhysics Letters) **113**, 18004 (2016).
[16] G. Hummer, J. Chem. Phys. **120**, 516 (2004).