Supporting information for "From dioxin to dioxin congeners: understanding the differences in hydrophobic aggregation in water and absorption into lipid membranes by means of atomistic simulations."

Mosé Casalegno[a], Guido Raos [a], and Guido Sello[b,*]

[a] Dipartimento di Chimica, Materiali e Ingegneria Chimica "G. Natta", Politecnico di Milano, Via L. Mancinelli 7, 20131 Milano, Italy.
[b] Dipartimento di Chimica, Università degli Studi di Milano, Via C. Golgi 19, 20133 Milano, Italy.
* email: guido.sello@unimi.it.

Details of molecular dynamics simulations

All simulations were carried out using the GROMACS-5.0 program suite[1]. Pressure and temperature were maintained at 1 atm and 325 K using a weak-coupling scheme[2]. The equations of motion were integrated using the leap-frog algorithm with a time step of 1 fs. Periodic boundary conditions were applied along all axes. The system temperature was kept constant via velocity rescaling[3], with a time constant of 0.1 ps. The pressure was controlled via semi-isotropic coupling to Berendsen barostats, with time constant of 3 ps and isothermal compressibilityof $4.6{\cdot}10^{-5}$ atm^{-1}. All non-bonded cut-off radii were set to 1.2 nm. Electrostatic interactions were treated via the Particle-Mesh Ewald method with a Fourier grid spacing of 0.12 nm. Hydrogen atoms were modeled explicitly, and no constraints were applied to bonds or angles.

To model the DPPC molecules we used the force field described by Ulmschneider et al.[4]. The TIP3P force field was used for water molecules[5]. The force field for solute molecules was developed starting from suitable OPLS-AA parameters[6].

Conventional MD simulations were carried out for all systems, with one (N1) and ten (N10) solute molecules. In all cases, the MD input was assembled starting from a pre-equilibrated bilayer containing 128 DPPC molecules in water[4]. The simulation box, with the bilayer oriented orthogonally to the z-axis, originally comprised 3655 water molecules. In order to give more room to the solute molecules, and prevent the bilayer to interact with its own top and bottom periodic images, the number of water molecules was doubled. Na^+ and Cl^- ions were also added to model a physiological solution. This input structure was then equilibrated via a 20-ns NPT simulation in order to adjust the box size and correct the density. This was necessary to avoid successive equilibration steps, and to obtain a starting point suitable for solute insertion. The final box size was approximately $6.5\times6.5\times11.0$ nm^3 in all cases.

A program, written by our group, was then used to insert the contaminants in the aqueous phase. The insertion was accomplished removing the minimum number of water molecule surrounding the solute ones. Regardless the species, the final number of water molecules was about 7000 in all N1 systems. After this step, NPT production runs were performed. For N1 systems, 3 production runs were carried out to generate the starting configurations for z-constrained MD simulations (see below for details). Each run consisted of more sequential simulations, each with a duration of 20-50 ns. For each N10 system, we carried out 3 independent runs to assess the dynamics of compound absorption.

Beside conventional MD simulations, we performed z-constrained calculations on all the N1 systems, in order to evaluate the free energy of transfer ($\Delta G(z)$), the local diffusion coefficients ($D(z)$), solute resistance profiles ($R(z)$), the permeability coefficients (P), and the translocation times (τ). As mentioned above, the initial configurations in these simulations were drawn from unconstrained NPT trajectories. This strategy was shown to give more reliable results compared to the generation of starting configurations from pulling simulations.[8] For each species, we assembled five independent sets made up by twenty-one equally spaced points

(from 0 to 4 nm) along the z-coordinate separating the solute and bilayer center-of-mass. Only the NPT configurations within 0.05 nm from the correct positions along the constrained coordinate were selected for set building. This stage required about 100 ns of NPT runs for each molecule.

For each position within each set a 20-ns NPT run was performed. As detailed in the next paragraph, for some positions, the overall simulation time was extended to 40 ns. The instantaneous forces acting on the solute molecules, $F(z,t)$, were saved every 10 MD steps, corresponding to 0.01 ps. The mean force, $\bar{F}(z) \equiv \langle F(z,t) \rangle$, representing the average of the instantaneous force over the simulation time t_{md}, was calculated discarding the first 10 ns of simulation time as equilibration (see below).

The free energy difference betwee the bulk water phase (z = 4 nm) was obtained as potential of the mean force (PMF):[7]

$$\Delta G(z) = -\int_{z}^{+4} \bar{F}(z')\, dz' \tag{S1}$$

The local diffusion coefficient, $D(z)$, was evaluated according to the force autocorrelation function method:[7–9]

$$D(z) = \frac{(RT)^2}{\displaystyle\int_{t_0}^{t_{md}} \langle \Delta F(z,t)\, \Delta F(z,t_0) \rangle \; dt}, \tag{S2}$$

where $\Delta\, F(z,t) = F(z,t) - \bar{F}(z)$ is the deviation of the instantaneous force from the average force acting on the solute, while t_0 the time origin. We found the autocorrelation function to quickly approach zero. Therefore, the evaluation of the above integral was performed over a time interval of 10 ps. Following the procedure described by Allen et al.,[10] Eq. (S1) was evaluated starting from different time origins t_0 — taking into account for the equilibration time —, and the final result obtained as the average of ten individual estimates.

In order to improve the integral evaluation of the free energy, the mean force was interpolated via cubic Bèzier splines.[11] The same procedure was applied to the local diffusion coefficients while calculating the values of $R(z)$ (see below). Thousand sample points were adopted for both $\Delta G(z)$ and $D(z)$.

For each set, the solute resistance profiles, $R(z)$, and the permeability coefficients (P) were finally obtained according to the inhomogeneous solubility-diffusion model, integrating the solute resistance profile over the entire bilayer, from z = 4 to z = -4 nm:[7–9]

$$P = 1 \,/ \int_{-4}^{+4} R(z) dz = 1 / \int_{-4}^{+4} \frac{e^{\Delta G(z)/k_B T}}{D(z)} dz. \tag{S3}$$

The calculation was performed assuming the solute resistance symmetrical with respect to the bilayer center, e.g.: $R(-z)$=$R(z)$.

In addition to permeability coefficients, we also estimated the average time required for the solute molecule to cross the membrane, e.g. the translocation time (τ):[12]

$$\tau = \frac{1}{\bar{D}} \int_{-4}^{+4} e^{\bar{F}(y)/k_B T} dy \int_{-4}^{y} e^{-\bar{F}(z)/k_B T} dz, \tag{S4}$$

where $\bar{D}$ represents the value of the diffusion coefficient calculated averaging $D(z)$ over the interval from 1.0 to 1.1 nm, where the molecules where mostly located after the translocation process. The final values of free energies of transfer, local diffusion coefficients, solute resistances, reported in the main text, were obtained, for each solute molecule, averaging over the five independent sets. Similarly for permeability coefficients and translocation times.

Calculation of the structural and orientational parameters in N1 simulations

In order to quantify the effect of solute absorption on the DPPC membrane, we calculated four structural and orientational parameters: (1) the membrane thickness, (2) the area-per-lipid, (3) the solute orientation with respect to the membrane normal (e.g., the tilt angle θ in the main text), (4) the distance between the solute and the bilayer center-of-mass.

The solute orientation was calculated as the tilt angle between a molecular vector and the normal to the bilayer surface. For ANTH, THDD, TCDD and THDO, the molecular vector was chosen as that passing by the average positions of the outmost carbon atoms (2 for each side of the molecule). For TCBP, conversely, we only considered the vector connecting the two outmost carbon atoms. An angle of 0° indicates a solute molecule orthogonal to the membrane.

The figures below show the evolution of the above parameters over time in selected N1 simulations (one for each solute molecule). For better clarity, we reported 10-ns windows corresponding to solute absorption.

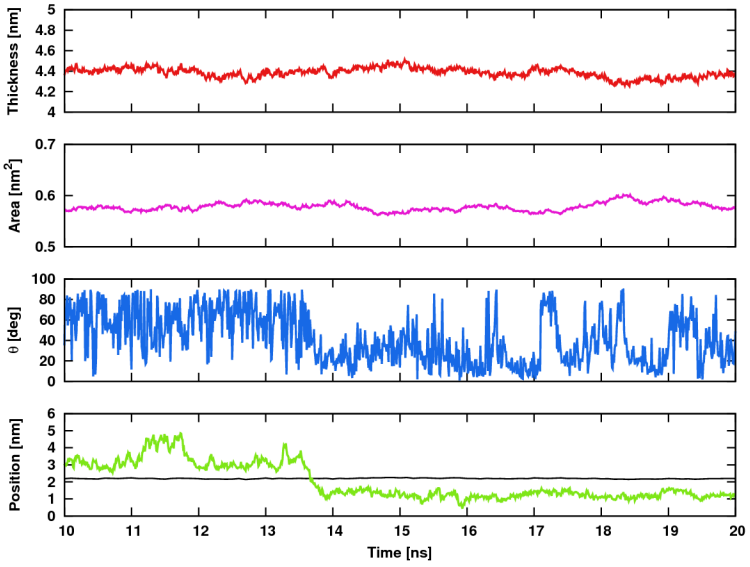

Fig. S 1: From the top: membrane thickness, area-per-lipid, tilt angle, and distance from bilayer center for ANTH.

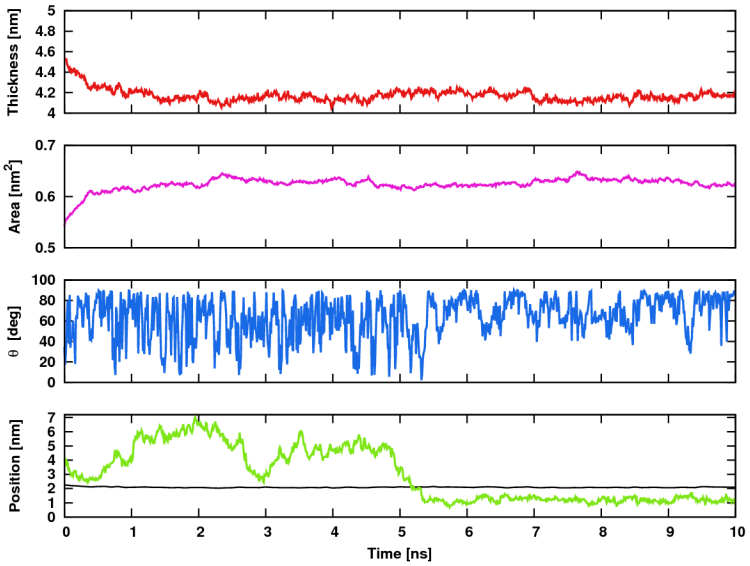

Fig. S 2: From the top: membrane thickness, area-per-lipid, tilt angle, and distance from bilayer center for TCDD.

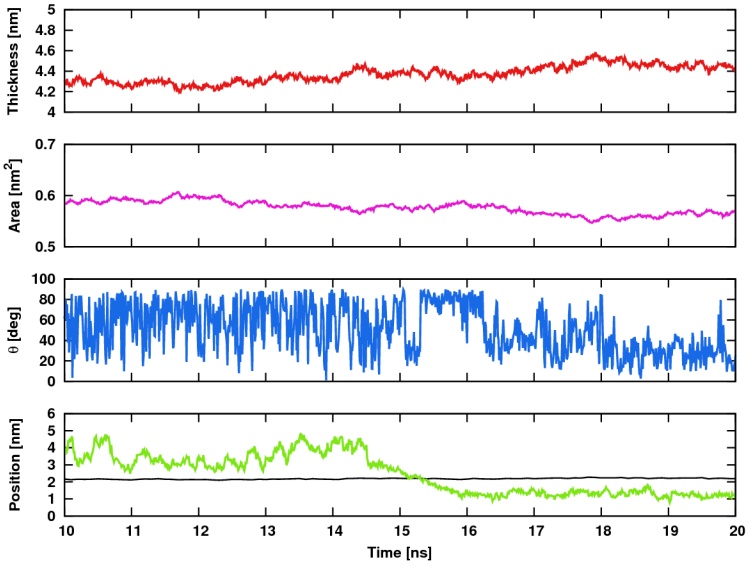

Fig. S 3: From the top: membrane thickness, area-per-lipid, tilt angle, and distance from bilayer center for THDD.

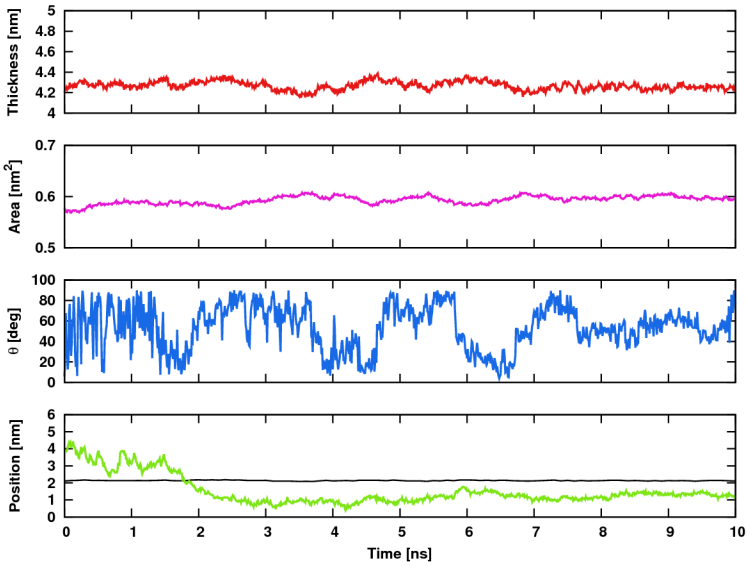

Fig. S 4: From the top: membrane thickness, area-per-lipid, tilt angle, and distance from bilayer center for TCBP.

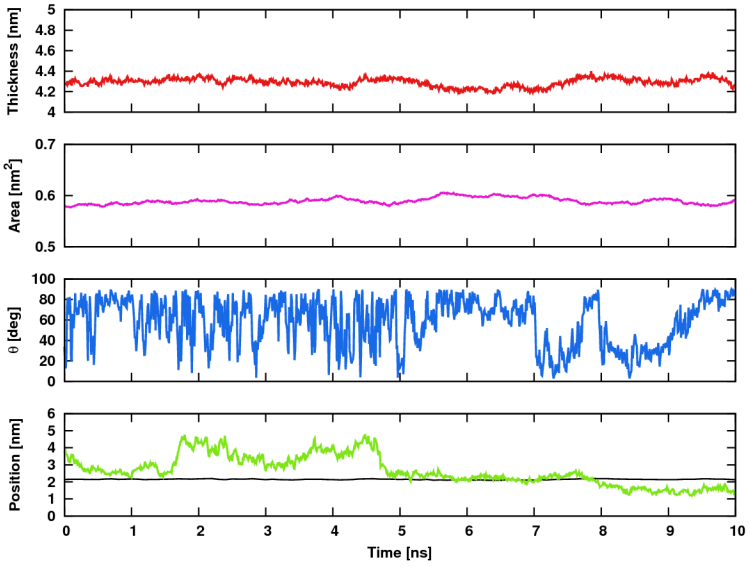

Fig. S 5: From the top: membrane thickness, area-per-lipid, tilt angle, and distance from bilayer center for THDO.

Statistical convergence of free energies of transfer

In this paragraph, we provide the results of some numerical tests aimed at assessing the statistical convergence of equilibrium properties of z-constrained simulations.[13] In particular, we evaluated the impact of the equilibration time (t_{eq}) on the average force and the free energy of tranfer. As we shall show below, this analysis provided a simple way to identify poorly sampled points along the constrained coordinate. The sampling at such points was then improved by extending the simulation time (t_{md}).

The analysis was performed on the results of NPT runs with overall duration of 20 ns for each position. Within each species, the five sets were tested independently. Hereafter, for the sake of brevity, we will focus on some illustrative examples. Fig. S6 show $\bar{F}(z)$ and $\Delta G(z)$ profiles for THDD (set 1). The numerical values of the force associated with simulation times shorter than t_{eq} were neglected in the calculation of the average force and the free energy. Both in bulk water and lipophilic tail phases, the change in t_{eq} had little effect on the values of $\bar{F}(z)$. Conversely, some deviations were observed in the headgroup region, between 2.4 and 1.8 nm from the bilayer center. As a consequence, the free energy profile had different behaviours below 2.4 nm. In particular we found the free energy to decrease by increasing t_{eq}.

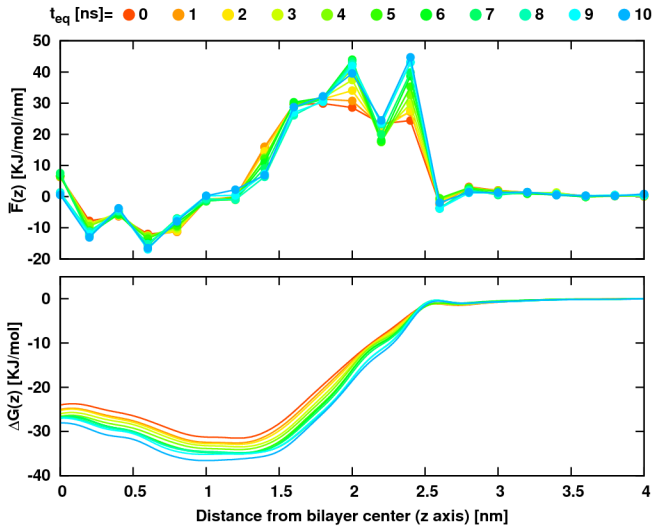

Fig. S 6: Profiles of $\bar{F}(z)$ and $\Delta G(z)$ for THDD (set 1) with increasing equilibration time (from 0 to 10 ns).

Fig. S7 shows the same profiles for TCBP (set 3). Also in this case, the largest fluctuations in $\bar{F}(z)$ were observed within the headgroup region. Entirely similar results were also obtained for other species. In order to minimize the numerical fluctuations in the free energy, whilst maintaining acceptable CPU costs, we decided to extend to 40 ns the simulation time of the most "critical" z-points, namely at z = 1.8, 2.0, 2.2, and 2.4 nm. Fig. S8 compares two free energy profiles for the THDD system (set 1), prior (left), and after (right) the extension. As one can see, extending the sampling decreased the differences in $\Delta G(z)$ across different values of t_{eq}, whereas the position of the free energy minima was unaffected.

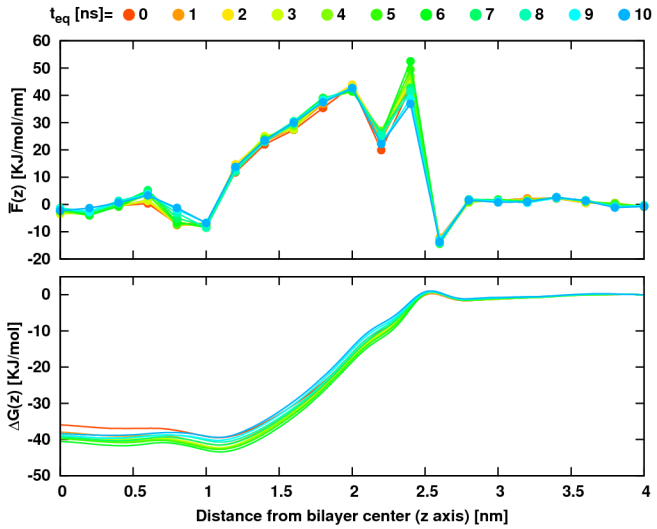

Fig. S 7: Profiles of $\bar{F}(z)$ and $\Delta G(z)$ for TCBP (set 3) with increasing equilibration time (from 0 to 10 ns).

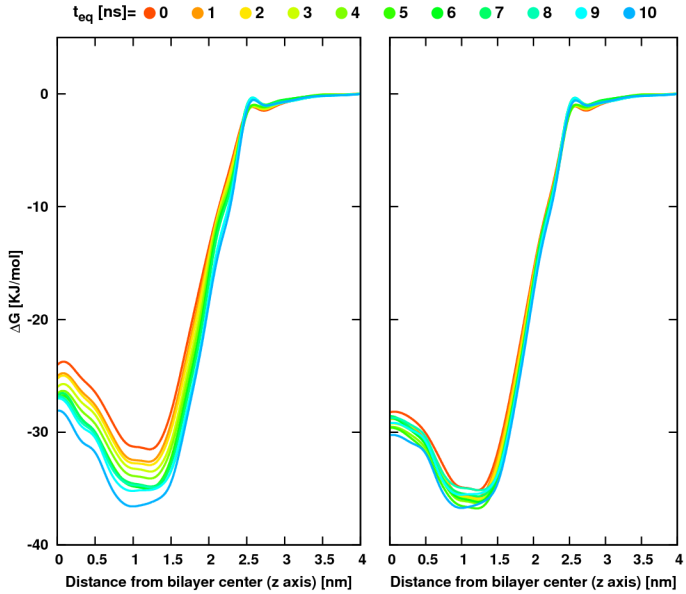

Fig. S 8: Comparison between $\Delta G(z)$ profiles for THDD (set 1) with increasing equilibration time (from 0 to 10 ns). Left panel: raw profiles (t_{md} = 20 ns). Right panel: the same profiles with extended sampling (t_{md} = 40 ns, for the critical z-points).

Fig. S9 shows a similar plot for ANTH (set 1). Also in this case, extending the sampling narrowed the difference between the $\Delta G(z)$ profiles. A similar outcome was also observed for the other species. We note that, the improvement in the free energy estimates required an additional CPU cost of about 60 hours per point, per set, for each solute molecule, corresponding to 6000 hours overall.

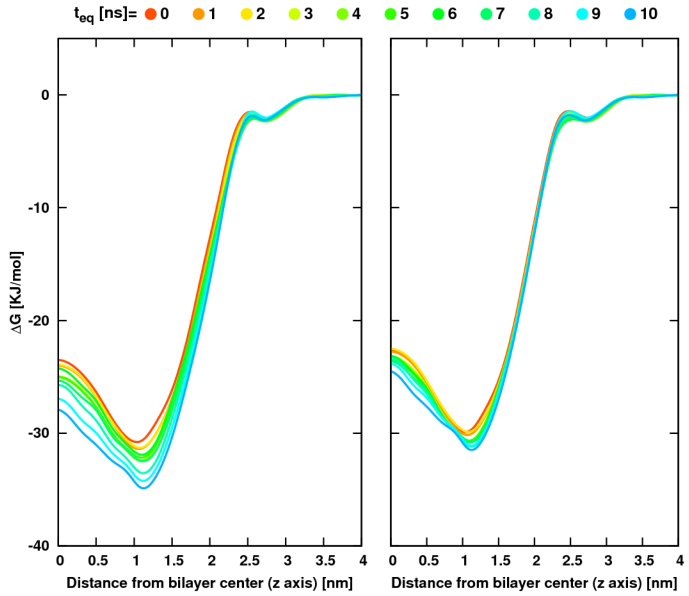

Fig. S 9: Comparison between $\Delta G(z)$ profiles for ANTH (set 1) with increasing equilibration time (from 0 to 10 ns). Left: panel: raw profiles (t_{md} = 20 ns). Right panel: the same profiles with extended sampling (t_{md} = 40 ns, for the critical z-points).

The z-constrained NPT simulations with extended sampling were then used to find an optimal value of t_{eq}. To this end, we evaluated the impact of t_{eq} on the convergence of the average free energy of each species. Fig. S10 shows the final (e.g. averaged over the five sets) free energy profiles for ANTH and TCDD. These were obtained by increasing the equilibration time from 0 to 12 ns. Equilibration times beyond this value were not considered, in order to preserve a statistically significant amount of data. For better clarity, we reported only the region around the free energy minima. Similarly to what observed in Figs. S8 and S9, increasing t_{eq} decreased $\Delta G(z)$ progressively. A smaller difference between consecutive profiles was observed for equilibration times equal or greater than 10 ns. Similar results were also obtained for the remaining molecules.

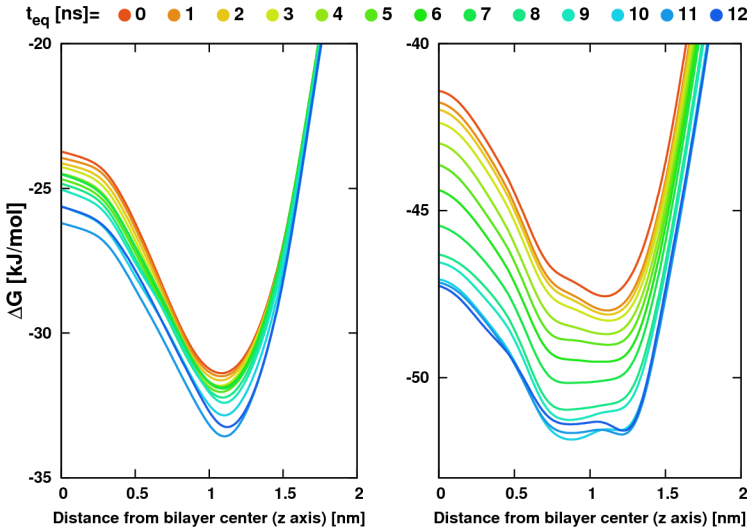

Fig. S 10: Free energy profiles obtained for ANTH (left) and TCDD (right) with increasing equilibration time (from 0 to 12 ns). For better clarity, only the region around the free energy minima is shown.

On the basis of these results, the free energy profiles corresponding to t_{eq} =10 ns were chosen as the most representative ones. We note that an entirely analogous result was achieved in our previous work.[14] Fig. S11 shows these profiles for all solute molecules together with the standard deviations calculated by averaging over the five sets. Note that the same profiles were reported in the main text, where, however the standard deviations were omitted for better clarity.

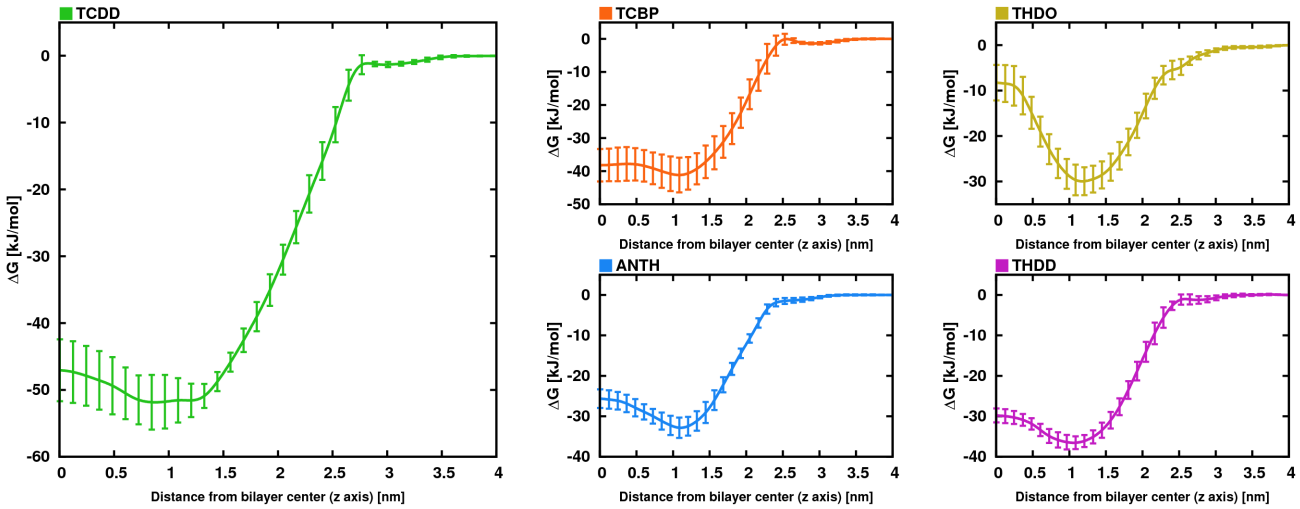

Fig. S 11: Free energy profiles for all solute molecules with standard deviations calculated over the five sets.

We performed one additional test aimed at checking the sampling efficiency of z-constrained runs.[13] More specifically, we verified that our simulations were not biased by the existence of poorly sampled regions along a coordinate orthogonal to the constrained one. As testing variable we chose the solute orientation with respect to the bilayer surface (θ), which is an important degree of freedom for the solute molecule.

The test was performed by extracting all θ values from the MD trajectories at any given constrained position. The values obtained over the five sets (thousand values for each z-point) were then grouped into bins of equal size (0.5°). Presence (P=1), or absence (P=0), of all angles within a given bin was then evaluated by means of a binary counter, and used as a measure of sampling exhaustiveness. For the sake of brevity, in the following, we shall focus on the results obtained for TCDD. Similar results were obtained for the other solute molecules.

Fig. S12 shows the sampling of the tilt angles at z = 0.2, 0.6, 1.0, and 1.6 nm. The values of z selected for this figure correspond to the inner regions of the lipid bilayer. As one can see, the tilt angles were well sampled over the whole range from 0° to 90°. The percentages of filled (P=1) bins, reported in Table 1, were in this case greater than 97%.

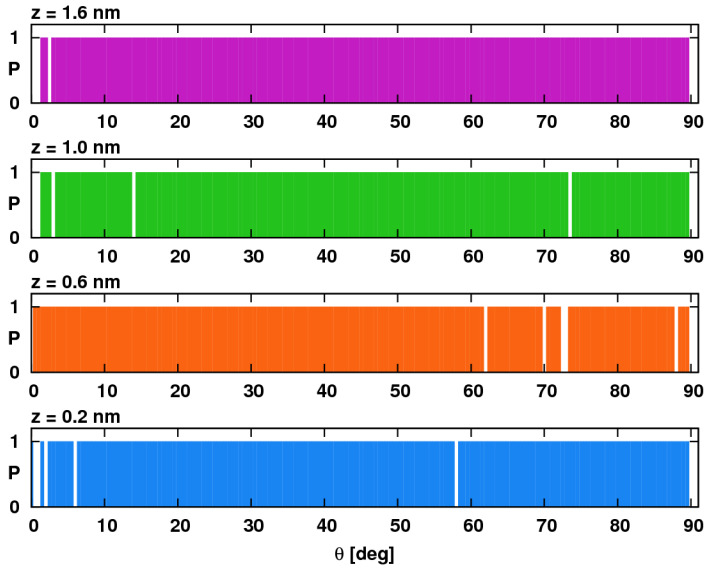

Fig. S 12: Binning plot showing the presence (P=1), or absence (P=0), of tilt angles for TCDD at z = 0.2, 0.6, 1.0, and 1.6 nm obtained in z-constrained calculations. The bin size was 0.5°.

Fig. S13 shows the tilt angle sampling at z = 2.0, 2.4, 3.0, and 3.8 nm for TCDD. The head group region (z = 2.0, and 2.4 nm) shows some empty bins at high θ values. This result could have been predicted considering the low probability that an angle greater than 70° could be here taken by TCDD. The intermolecular interactions in the head group region are strong enough to permit the membrane stability and a substantial deformation is not expected. Overall, the sampling of θ is nearly complete (always greater than 94%).

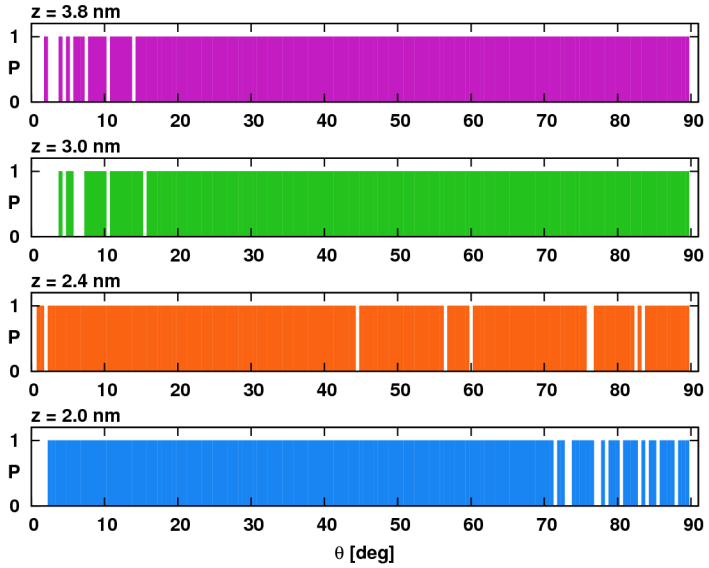

Fig. S 13: Binning plot showing the presence (P=1), or absence (P=0), of tilt angles TCDD at z = 2.0, 2.4, 3.0, and 3.8 nm obtained in z-constrained calculations. The bin size was 0.5°.

Table. S 1: Percentages of filled bins calculated for Figs S12 and S13.

z [nm]	Percentage [%]
0.2	97.2
0.6	97.2
1.0	98.3
1.6	99.4
2.0	93.7
2.4	95.5
3.0	96.5
3.8	95.5

Calculation of intermolecular cluster energies in N10 systems

In order to evaluate the stability of the clusters formed by different species, we calculated the average energy associated with the removal of one molecule from the largest aggregate formed during a MD simulation. For a cluster containing N molecules, we considered the difference between the intermolecular cluster energy, and the energy associated with the sub-clusters containing N-1 molecules. The N-1 sub-clusters were obtained by removing in turn one molecule from the original cluster. The energy difference, ΔE_N, was then calculated as:

$$\Delta E_N = \frac{\sum_k E_N - E^k_{N-1}}{N} = E_N - \frac{\sum_k E^k_{N-1}}{N}, \tag{S5}$$

where E_N represented the Coulomb, or the Lennard-Jones, intermolecular energy for the whole cluster, whereas E^k_{N-1} the same energy calculated removing the k-th molecule from the cluster. For each species, the calculation was performed on 500 selected MD frames where the largest cluster was present as a whole aggregate. It should be noted that this calculation was performed without considering the interactions between the solute molecules and DPPC, water, and ions. In order to determine the size and the number of clusters in a given frame and select it, we calculated the center-of-mass distances between each pair of molecules. Within each pair, the molecules separated by less than 1 nm were assigned to the same cluster. These sub-clusters (N = 2) were therefore joined in larger aggregates, in order to obtain the composition of all clusters within a given frame.

For TCDD, TCBP, THDD, and THDO, a significant number of frames was found with a maximum cluster size of 10. For ANTH, conversely, the maximum cluster size was 9. A shell script was instructed to select only the frames with clusters of the maximum size, and prepare single-point input files for GROMACS. The Coulomb and Lennard-Jones energies were then extracted by means of GROMACS tools. Finally, the values of ΔE_N were averaged over all the selected frames to provide a unique value for the Coulomb, and the Lennard-Jones interactions.

References

[1] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comp. Chem.*, 2005 **26**, 1701-1718; B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comp.*, 2008, **4**, 435-447.

[2] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684-3690.

[3] G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

[4] J. P. Ulmschneider and M. B. Ulmschneider, *J. Chem. Theory Comput.*, 2009, **5**, 1803-1813.

[5] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.

[6] W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225-11236; W. L. Jorgensen and N. A. McDonald, *Theochem*, 1998, **424**, 145-155; W. L. Jorgensen and N. A. McDonald, *J. Phys. Chem. B*, 1998, **102**, 8049-8059; R. C. Rizzo and W. L. Jorgensen, *J. Am. Chem. Soc.*, 1999, **121**, 4827-4836; M. L. Price, D. Ostrovsky and W. L. Jorgensen, *J. Comp. Chem.*, 2001, **22**, 1340-1352; E. K. Watkins and W. L. Jorgensen, *J. Phys. Chem. A*, 2001, **105**, 4118-4125; G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474-6487; C. Caleman, P. J. van Maaren, M. Hong, J. S. Hub, L. T. Costa and D. van der Spoel, *J. Chem. Theor. Comput.*, 2012, **8**, 61-74.

[7] S. J. Marrink and H. J. C. Berendsen, *J. Phys. Chem.*, 1994, **98**, 4155-4168; S. J. Marrink and H. J. C. Berendsen, J. *Phys. Chem.*, 1996, **100**, 16729-16738.

[8] M. Paloncýová, K. Berka and M. Otyepka, *J. Chem. Theory Comput.*, 2012, **8**, 1200-1211.

[9] M. Orsi, W. E. Sanderson and J. W. Essex, *J. Phys. Chem. B*, 2009, **113**, 12019-12029.

[10] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, UK, 3rd edn, 1989.

[11] R. H. Bartels, J. C. Beatty and B. A. Barsky, *An introduction to splines for use in computer graphics and geometric modelling*, Morgan Kaufmann, 1998, pp. 211-245.

[12] R. Qiao, A. P. Roberts, A. S. Mount, S. J. Klaine and P. C. Ke, *Nano Lett.*, 2007, **7**, 614-619

[13] C. Neale, W. F. D. Bennett, D. P. Tieleman and R. Pomès, *J. Chem. Theory Comput.*, 2011, **7**, 4175-4188; J. P. M. Jämbeck and A. P. Lyubartsev, *J. Phys. Chem. Lett.*, 2013, **4**, 1781-1787.

[14] M. Casalegno, G. Raos and G. Sello, *Phys. Chem. Chem. Phys.*, 2015, **17**, 2344-2348.