

NMRlipids Databank: Making data-driven analyses of membrane properties accessible for all

Anne Kiirikki¹, ...², and O. H. Samuli Ollila^{1,*}

¹University of Helsinki, Institute of Biotechnology, Helsinki, Finland

²Affiliation, department, city, postcode, country

*samuli.ollila@helsinki.fi

ABSTRACT

We present a databank of lipid bilayer simulations from the NMRlipids open collaboration project.

Introduction

Cellular membranes are composed of hundreds of different types of lipid molecules which regulate membrane properties, morphology and biological functions¹. Membrane lipid composition can be also related to diseases, such as cancer and neurodegenerative disorders, and potential therapeutics that affect membrane compositions have been proposed². However, membrane containing systems are often difficult to study experimentally because membranes are in disordered fluid state in biological conditions and sample preparation can be highly complicated due to rich phase behaviour and complex interactions between lipids and other biomolecules. Therefore, connections between complex lipid interactions and biological functions are still poorly understood. Molecular dynamics (MD) simulations have been particularly useful in understanding membrane systems, yet their accuracy is often compromised by the quality of model parameters and other artefacts^{3,4}. On the other hand, the accuracy of models is becoming increasingly important as researches are progressing from the simulations of individual molecules to the models of whole cells or organelles using interdisciplinary approaches^{4–6}. In such systems exhibiting very complex behaviour, inaccuracies are more difficult to detect and multiplication of small errors may have major effects the conclusions.

In contrast to experimental structural biology where standard protocols to share and quality evaluate the resolved structures are established in the PDB databank⁷, such best practices are yet to be defined for MD simulations. Importance of such approaches is widely recognized^{8–15} and solutions for sharing data are emerging for proteins in solution^{16,17} or membranes^{15,18,19}, nucleic acids²⁰, cyclodextrins²¹, and arbitrary biomolecules²². However, automatic quality evaluation^{16,20} and programmatic access are still rare. Therefore, tools for automatic quality evaluation of membrane simulations or training sets for machine learning models of membrane containing systems are not yet available.

Here we present the NMRlipids databank, a community driven open access databank with programmatic access containing atomistic resolution MD simulations of lipid bilayers. The programmatic access enables various data-driven approaches, providing new tools for researchers in wide range of fields in academia and industry from cell membrane biology to lipid nanoparticle formulations and data-driven computational chemistry and machine learning. Here we demonstrate how data-driven analysis of water anisotropic diffusion in membrane systems from the NMRlipids databank can extend the scope of MD simulations to MRI imaging and pharmacokinetics, and how the NMRlipids databank can be utilized to analyse rare phenomena that are beyond the scope of standard MD simulation studies. Furthermore, we perform automatic quality evaluation of membrane simulations which guides the selection of best models for specific applications and development of simulation parameters and methodology.

While the NMRlipids databank currently contains only bilayer systems, it can be readily expanded to contain other molecules than lipids, such as disordered proteins or sugars. The combination of overlay databank structure and open collaboration introduced here can be applied also when building databanks in other fields, particularly when storage of raw data requires significant resources, best practices in the field are not defined, and incentives to share data do not exist. Therefore, the approach opens up possibilities to create databanks enabling data-driven and machine learning applications in new fields.

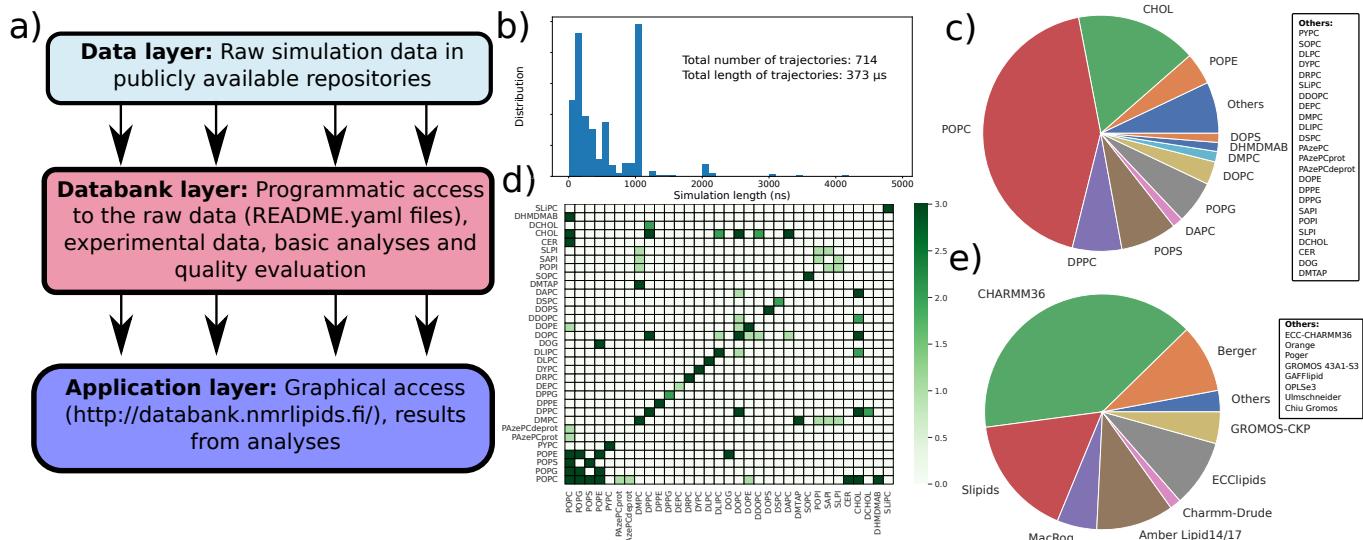


Figure 1. a) Structure of an overlay databank. More detailed structure of the layer 2 in the NMRLipids databank is illustrated in Fig. S6 in the SI. b) Distribution of the lengths of the trajectories, total number of trajectories and total length of the simulations in the NMRLipids databank. c) Distribution of lipids present in the trajectories in the NMRLipids databank. Lipids occurring in five or less simulations ('others') are listed in the right. d) Currently available binary mixtures in the NMRLipids databank. e) Distribution of force fields in the simulations in the NMRLipids databank.

Results

Programmatic access to MD simulation data of membranes composed of most common biologically abundant lipids

NMRLipids databank is a community driven databank containing atomistic MD simulations of biologically relevant lipid membranes emerging from the NMRLipids open collaboration^{23–27}. Nearly thousand simulation trajectories with the total length approaching half a millisecond can be accessed programmatically using Python API or graphical user interface (<http://www.databank.nmrlipids.fi/>). Following the FAIR principles²⁸ and NMRLipids project protocol, not only the content and computer programs related to the databank but also the whole construction process of the databank are open access²³.

Currently available single component lipid membranes and binary mixtures in the NMRLipids databank are illustrated in Fig. 1D. While these compose the majority of available trajectories, also mixtures up to five lipids are available. The distribution of available simulations containing a specific lipid is shown in Fig. 1C. This roughly resembles the biological abundancies of different lipid types, PC being the most common followed by cholesterol, PE, PS, PG, PI, and other lipids depending on organism and organelle¹. The force fields used in the simulations cover all the essential parameters used in lipid simulations, ranging from highly popular CHARMM36 parameters to united atom and polarizable force fields. Therefore, the averages calculated over the databank can be considered as mean predictions from available lipid models for an average cell membrane.

The efficient upcycling of large MD simulation trajectories in the NMRLipids databank is enabled by the overlay structure illustrated in Fig. 1A. Raw simulation data in the *Data layer* can be stored in any publicly available location with long term stability and permanent links to the data, such as digital object identifiers. The *Databank layer* is the core of the databank containing all the relevant information about the simulations, including links to the raw data, relevant metadata describing the systems, universal naming conventions for lipids and their atoms, computer programs to create and analyse the entries, basic properties calculated from all simulations (area per lipid, C-H bond order parameters, x-ray scattering form factors and membrane thickness), and quality evaluation of simulations against experimental data. In practise, this information is stored in the git repository which is currently available at <https://github.com/NMRLipids/Databank>. The *Application layer* contains outputs from the databank, such as graphical user interface (<http://www.databank.nmrlipids.fi/>) and results from analyses described in sections below. More detailed description on the NMRLipids databank structure is available in the supplementary information.

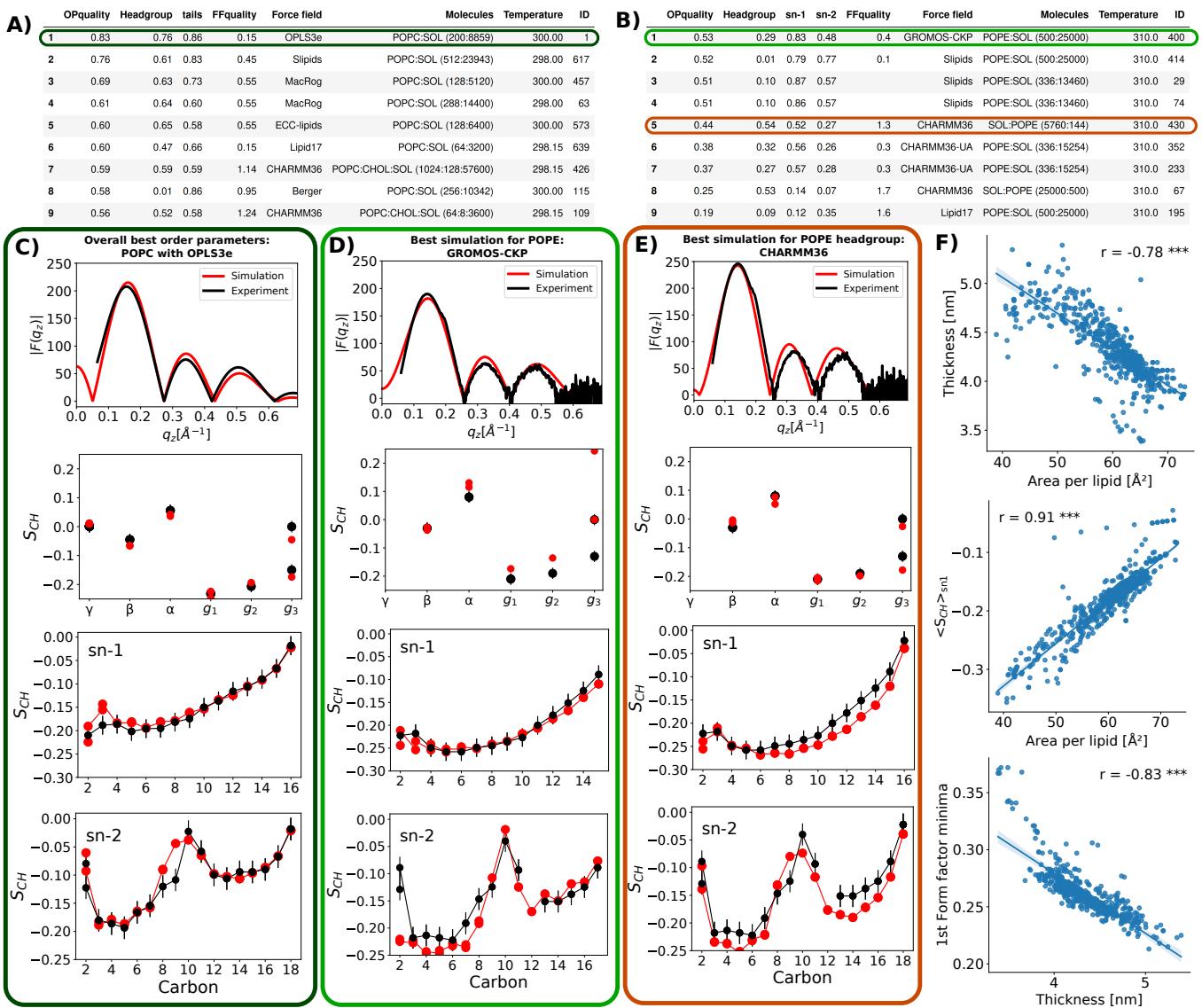


Figure 2. A) Best simulations ranked based on overall order parameter quality. B) Best simulations ranked based on the overall order parameter quality of POPE lipid. C)-E) Evaluation against experimental data exemplified for a simulation with the best overall order parameter quality (C), the best quality for POPE lipid (D), and the headgroup quality for POPE (E). F) Scatter plots and Pearson correlation coefficients for the membrane area per lipid, thickness, first minima of x-ray scattering form factor and average order parameter of the *sn*-1 acyl chain extracted from the NMRlipids databank. All correlation coefficients have p-value below 0.001. For more correlations see Fig. S1.

Selection of best simulation parameters for applications using the NMRlipids databank

Understanding the accuracy of used force field parameters is crucial when estimating the reliability and significance MD simulation results. However, the lack of universal quality measures for MD simulations and complicated landscape of force field quality for membranes hamper the quality of simulations which has resulted to many controversial results³. For example, previous quality evaluations of membrane simulations have concluded that CHARMM36 parameters give the best description for lipid headgroup conformational ensembles²⁷, GROMOS-CKP parameters best capture the membrane packing in POPS bilayers³, and that OPLS3e parameters overcome CHARMM36 in structural quality for POPC but predicts overestimated ion binding²⁹. To enable the navigation in this complex force field quality landscape, we define here quantitative quality measures that can be used to rapidly find the best available parameters for a specific application or to guide force field development. For detailed definitions of quality measures, see the supplementary information.

Ranking of all simulations with the highest scores based on evaluation against C-H bond order parameters from NMR

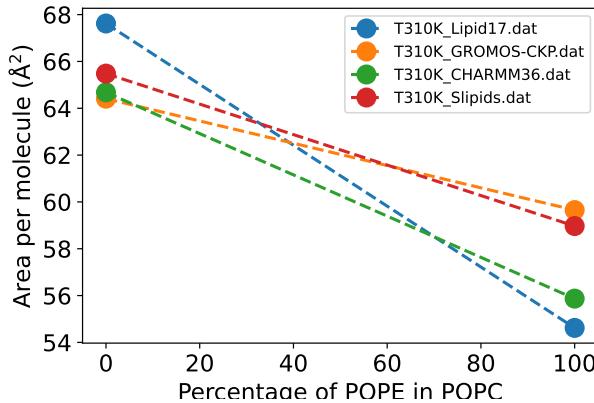


Figure 3. Differences in area per lipids of POPC and POPE bilayers between different force fields.

experiments is shown in Fig. 2 A and ranking of best POPE lipid bilayer simulations in Fig. 2 B. Direct comparisons with experiments for selected simulations are shown in Figs. 2 C-D. C-H bond order parameters are sensitive to conformational ensembles of individual lipid molecules²⁴, but they correlate well also with the membrane thickness and lateral packing as shown in Fig. 2 F. This is complemented by comparing the absolute values of x-ray scattering form factors between simulations and experiments that are related to the overall membrane dimensions via electron density profile. Particularly, the location of minima in these form factors are correlated with the membrane dimensions as shown in Figs. 2 F, S1, and S2.

The power of the NMRLipids databank in selecting the best simulation parameters for a specific application is demonstrated for mixtures of PE and PC lipids in Fig. 3 showing area per lipids from available POPC and POPE bilayer simulations in the databank from different force fields at 310 K. Predictions from different force fields deviate significantly in terms of absolute values and the slope of decrease upon addition of POPE. Because the area per lipid correlates with the average order parameter of *sn*-1 chain (Fig. 2 F), simulations with the best predictions for area per lipids can be selected based on order parameter quality evaluation. Quality evaluation in Fig. 2 and direct comparison in Fig. S3 reveal that Slipids ranks 1st for POPC with a clear difference to others, and 2nd for POPE with only marginally lower quality than GROMOS-CKP, therefore being the best selection for studies where packing effects of PE lipids are relevant. Simulations with CHARMM36 and GROMOS-CKP predict too packed bilayer for POPC (overestimated order in Fig. S3), while area per lipid for POPC is overestimated in Lipid17. For POPE, CHARMM36 and Lipid17 predict too packed membranes.

In conclusion, the automatic quality evaluation in the NMRLipids databank enables rapid selection of the best simulation parameters to specific applications without extensive and tedious manual force field evaluation. This possibility will promote more reliable simulation results for wide range of applications, support force field development and parametrization of coarse grained force fields against the best atomistic MD simulations. The automatically extracted quantitative quality measures will be particularly useful to guide automatic parametrization procedures.

Detection rare phenomena using NMRLipids databank: Cholesterol flip-flops

Lipid flip-flops from one bilayer leaflet to another play an important role in lipid trafficking and regulating membrane properties¹. Phospholipid flip-flops are slow, occurring with the timescales of hours or days, while cholesterol, diacylglycerol and ceramide flip-flops are faster, yet the reported timescales range between minutes to sub-millisecods^{1,30-32}. These timescales were previously accessible only by coarse-grained simulations or free energy calculations³¹, yet atomistic simulations reporting cholesterol flip-flop events have been published recently³²⁻³⁴. These studies report increase in cholesterol flip-flop rates with increasing unsaturation level and decreasing cholesterol concentration^{32,33}, but correlations between cholesterol flip-flop rates and membrane properties have not been systematically studied. Here we demonstrate that the large amount of simulation data available in the NMRLipids databank with programmatic access makes the analysis of these correlations accessible for all.

Flip-flops were observed for cholesterol, DCHOL (18,19-di-nor-cholesterol), DOG (1,2-dioleoyl-sn-glycerol), and SDG (1-stearoyl-2-docosahexaenoyl-sn-glycerol) in the NMRLipids databank. Cholesterol flip-flop rates range between 0.001-1.6 μs^{-1} with the mean of 0.16 μs^{-1} and median of 0.07 μs^{-1} . These values are in line with the previously reported values from atomistic MD simulations³²⁻³⁴. Flip-flop rates of DCHOL, 0.2 μs^{-1} , was close to the average value of cholesterol, while average rates for diacylglycerols DOG and SDG were higher than for cholesterol, 0.4 μs^{-1} and 0.5 μs^{-1} , respectively. Flip-flops were not observed for other lipids, giving the upper limits for PC lipid flip-flop rate of 9 s^{-1} and for ceramide (N-palmitoyl-D-erythro-sphingosine) of 0.002 μs^{-1} . Therefore, data in the NMRLipids databank suggests that the lipid flip-flop

rate decreases in the order of diacylglycerols > cholesterol > other lipids including ceramides. However, amount of data for diacylglycerols (8 simulations with Lipid17 force field) and ceramide (3 simulations with CHARMM force field) is less than for cholesterol (83 simulations), thus we cannot fully exclude the effect of force field or composition on this comparison.

Nevertheless, we use the wide range of available simulations to analyse how cholesterol flip-flop rate depends on membrane properties. Cholesterol flip-flop rates and their averages over fixed ranges of x-axis values are plotted as a function of membrane thickness, lateral density and order in Figs. 4 B-D. The results reveal a non-linear correlation between decreasing cholesterol flip-flop rate and membrane packing. Flip-flop rates increase with an order of magnitude when membrane packing density decreases and a major jump is observed at low membrane packing. The order of magnitude changes in cholesterol flip-flop rate with the membrane composition may have major implications in understanding lipid trafficking and membrane biochemistry^{32,34}. Because the results from the NMRLipids databank are averaged over large range of membrane compositions and force fields, they show that the strong dependence of cholesterol flip-flop rate on membrane properties is not limited to certain lipid compositions or force fields used in previous studies³²⁻³⁴.

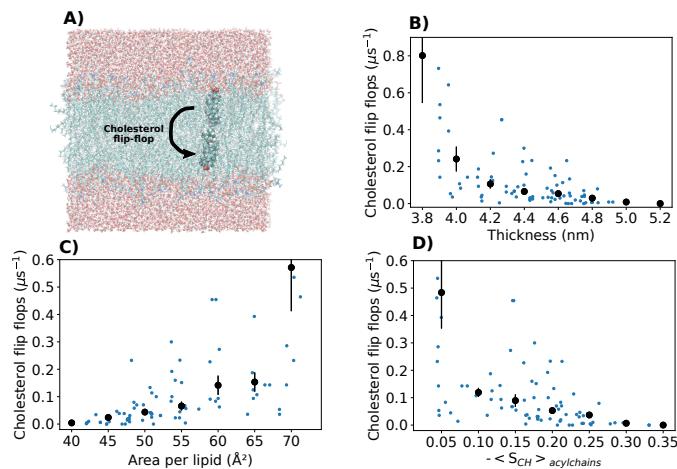


Figure 4. **A** Illustration of cholesterol flip-flop. **B-D** Cholesterol flip-flops analyzed from the databank as a function of membrane thickness, area per lipid, and acyl chain order. Values from simulations with non-zero flip-flop rates are shown with blue dots. Averages over fixed range of x-axis values are shown with black dots.

Cholesterol flip-flops are phenomena whose analyses from MD simulations has only recently became possible for groups with substantial resources and special expertise available³²⁻³⁴. Our analysis demonstrates how the NMRLipids databank makes analyses of such rare phenomena accessible to wide range of scientists in different fields of science and industry who do not have the access to required resources and expertise to produce the large amounts of MD simulation data.

Extending the scope of MD simulations to new fields using the NMRLipids databank: Water diffusion anisotropy in membrane systems

The anisotropy in water diffusion in parallel and perpendicular directions with respect to membranes plays a role in the drug translocation through biological material, particularly in skin³⁵⁻³⁸, and in MRI imaging³⁹. MD simulations are rarely used to analyze anisotropic diffusion of water as only few permeation events for water are typically observed in a single MD simulation trajectory^{40,41}, thereby making the collection of sufficient amount data challenging. Here we show that the NMRLipids databank can be used to analyze how anisotropic water diffusion in a multilamellar membrane systems depends on membrane properties.

To this end, we first calculated the water permeability through membranes from all simulations in the NMRLipids databank. The resulting non-zero values range between 0.3 and 322 μm/s with the mean and median of 14 μm/s and 8 μm/s, respectively. These values agree with the previously reported simulation results^{40,41} and have the same order of magnitude as experimentally determined diffusive permeability coefficients, but are on average larger than the values reported for PC lipids in liquid crystalline phase, 0.19-0.33 μm/s⁴³. Observed permeabilities and their averages over fixed ranges of values at x-axis are shown in Figs. 5 B-E as a function of temperature, membrane thickness, area per lipid, and acyl chain order. As expected, the permeability increases with the temperature, giving $17 \pm 4 k_B T$ for the average energy barrier for the water permeation from the Arrhenius plot in Fig. 5 B. On the other hand, the permeability of water decreases on average when membranes become more packed, i.e., with decreasing area per lipid and increasing thickness and acyl chain order (Figs. 5 C-E). Permeation of water through bilayers depends on membrane properties also according previous studies, but it is not agreed whether the area per lipid⁴⁴ or bilayer thickness⁴⁵ would be the main parameter determining the permeability. Our analysis over the NMRLipids databank, containing significantly more data than that was available in previous studies, suggest non-linear dependence on

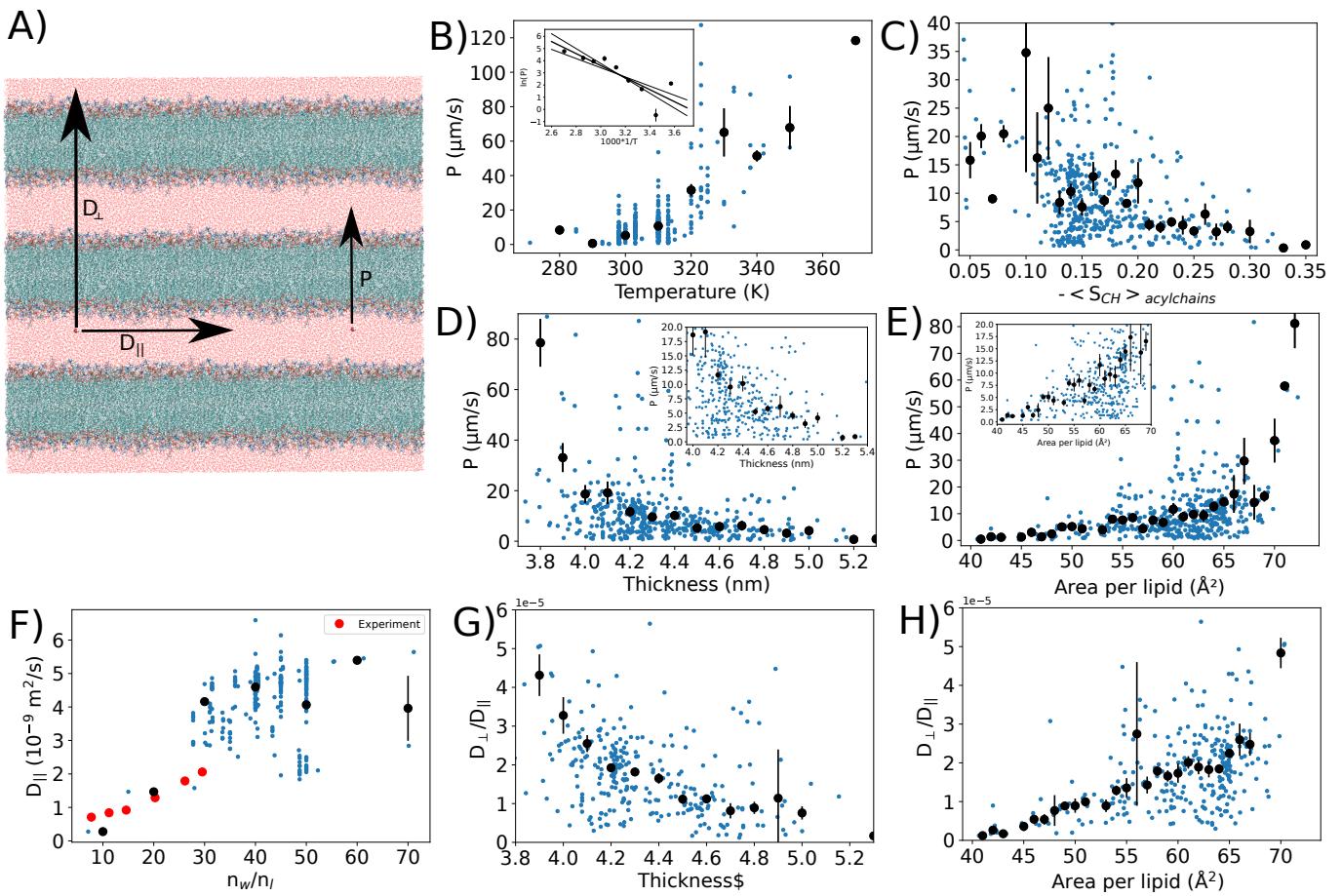


Figure 5. **A** Water diffusion, D_{\perp} , and permeability, P , through membranes, and lateral diffusion along the membrane, D_{\parallel} , illustrated in a multilamellar stack of lipid bilayers. **B-E** Water permeation through membranes analyzed from the databank as a function of temperature, thickness, area per lipid, and acyl chain order. Values from simulations with non-zero permeation values are shown with blue dots. Averages over fixed range of x-axis values are shown with black dots. Insert in B) shows the Arrhenius plot of permeation ($\ln(P)$ vs. $1/T$) that gives $17 \pm 4 k_B T$ for the average activation energy for water permeation through lipid bilayer. Inserts in C) and D) show the region where the dependence could be considered approximately linear. **F** Lateral diffusion of water as a function of hydration level. Experimental points for DMPC bilayers at 313 K at different hydration levels are shown⁴². **G-H** Diffusion anisotropy of water as a function of thickness and area per lipid.

both of these parameters, yet the linear correlation would be a good approximation for thicknesses above ~ 3.9 nm and area per lipids below $\sim 69 \text{ \AA}^2$ (insets in Figs. 5 C-D). Clear dependencies of permeability on charged lipids, cholesterol, POPE, or hydration level were not observed in Fig. S4 in the supplementary information.

To analyze how water diffusion anisotropy depends on membrane properties in a multi-lamellar lipid bilayer system, we calculated also the water diffusion parallel to the membrane surface from all simulations in the NMRlipids databank. The parallel diffusion coefficient of water, D_{\parallel} , decreases with reduced hydration and increases with the temperature, but dependencies on membrane area per lipid, thickness, or fraction of charged lipids were not observed in Figs. 5 and S5. Simulation results are close to the experimental values with low hydration levels in Fig. 5 F, but increase to the values of approximately two times higher than experimental bulk water diffusion value ($3.1 \cdot 10^{-9} \text{ m}^2/\text{s}$ at 313 K⁴⁶) with high hydration levels. This is not surprising as the most common water model used in membrane simulations, TIP3P, overestimates the bulk water diffusion⁴⁷. To estimate the diffusion anisotropy of water, D_{\perp}/D_{\parallel} , in multi-lamellar membrane system, the permeability coefficients of water through membranes were translated to perpendicular diffusion coefficients, D_{\perp} , using the Tanner equation^{48,49}. The resulting perpendicular diffusion coefficients are approximately five orders of magnitude smaller than lateral diffusion coefficients of water (Figs. 5 G and H), which is at the upper limit of the anisotropic estimated from the experimental data³⁷. Significant increase in the diffusion anisotropy with membrane packing is observed, as D_{\perp}/D_{\parallel} drifts away from one with decreasing area per lipid and increasing thickness in Figs. 5 G and H. This follows from decreasing water permeability with membrane packing

(Figs 5 C and D), while lateral diffusion remains approximately constant (Fig. S5 A and C).

In conclusion, our results suggest that the bilayer packing has a substantial effect on anisotropic water diffusion in multi-membrane lipid systems. Several folds larger anisotropy in membranes with higher lateral density is expected to play a role in pharmagokinetic models not only for water but also for other hydrophilic molecules³⁷. Furthermore, the understanding of this anisotropy may help in developing new MRI imaging methods³⁹.

Discussion

The focus of biomolecular simulations is moving from studies of individual molecules to larger complexes and even whole cells and organelles^{4–6}. On the other hand, machine learning based models predicting behaviour of biomolecules and automatic approaches to parametrize models are emerging^{3,50}. The NMRLipids databank will support the development in all these directions. The automatic quality evaluation and ranking of simulations in the NMRLipids databank guides the optimization of new force field parameters and the selection of best parameters for large biomolecular complexes where simulation quality becomes increasingly important due the multiplication of small errors in large simulations. On the other hand, the NMRLipids databank serves as a training set for diverse machine learning applications. For example, a machine learning model predicting electron density profiles from form factors would be highly useful in interpretation of scattering experiments. On the other hand, more elaborated models can be trained to predict arbitrary membrane properties using the data from the NMRLipids databank.

Providing programmatic access to large scale MD simulation data in the NMRLipids databank can lead to applications in unprecedented directions in fields where MD simulations are less commonly utilized. Here we exemplified this by analysing the anisotropic diffusion of water (Fig. 5), which is relevant in pharmacokinetic modeling and in MRI imaging where MD simulations are not yet commonly used^{37,39}. On the other hand, programmatic access in the NMRLipids databank enables automatic analyses over larger sets of simulation data in terms of quantity (e.g., simulation length and number of conformations) and content (e.g., lipid compositions and ion concentrations) that is currently possible in a single research group. Added value of such access to simulation data is demonstrated here for the analysis of simulation qualities (Fig. 2), water permeation through membranes (Fig. 5) and cholesterol flip-flop events (Fig. 4). Different types of applications enabled by the NMRLipids databank in wide range of fields are exemplified in Table 1. The increasing amount of data is expected to further increase the scope of potential applications of the NMRLipids databank in wide range of fields ranging from molecular biology and biotechnology to material science and biomolecular imaging.

Main practical limitations in building open access databanks of molecular dynamics simulations have been the required commitment in long term support for hardware and software maintenance, and the lack of incentives for researchers to share the data. The NMRLipids databank circumvents these challenges with the overlay databank design and open collaboration approach developed in the NMRLipids project²³. In this model, the file storage is distributed to publicly available stable repositories (*Data layer* in Fig. 1 A) and maintenance of the databank does not depend on individual scientists or groups because all its version controlled content is available with an open access licence (*Databank layer* in Fig. 1 A). Incentive to share the data is created in the NMRLipids open collaboration by offering authorship in published articles to the contributors following the NMRLipids project protocol²³. The current NMRLipids databank contains only lipid bilayer simulations, but the concept can be applied also to other biomolecules, such as disordered proteins and membrane-protein systems, or other fields where similar barriers to establish publicly accessible databanks exist.

Methods

Structure of the databank

Structure of the NMRLipids databank is illustrated more detailed in Fig. S6 in the supplementary information. The core content of the databank (*Databank layer* in Fig. 1) locates as a git repository at <https://github.com/NMRLipids/Databank/> and is permanently stored in Zenodo repository (www.zenodo.org)². Whenever a specific file is referred here, the file path within the NMRLipids databank repository is given. The scripts in the NMRLipids databank are mainly written in Python and many of them utilize the MDAnalysis module^{56,57}.

Essential information of each simulation in the NMRLipids databank is stored in a human and machine readable README.yaml file located at */Data/Simulations* in the NMRLipids databank repository. These files contain access to all information that are needed for further analyses of simulations. The content of these files is described in detail in table S1 in the supplementary information. Raw MD simulation data are stored in external publicly available and stable repositories (*Data layer* in Fig. 1), such as Zenodo (www.zenodo.org), from where it can be downloaded whenever needed using the links in README.yaml files.

Type of application	Practical examples	Target group
Analyses of rare phenomena	Lipid flip-flops, water permeation	Membrane scientists
Correlations between membrane properties	Membrane structural properties, water dynamics (Figs. 2 and 5)	Membrane scientists
Applications that are outside typical scope of MD simulations	Anisotropic water diffusion for pharmacokinetics and MRI imaging applications	Scientist in fields where MD simulations are not usually applied
Selection of the best simulation model for a specific application	Best model for POPC lipids (Fig. 2), headgroup conformations ²⁷ , packing of PS ³ and PE (Fig. 3) containing membranes.	Scientists using MD simulations
Guidance for force field development	Improvements in ion binding to lipids ^{51,52} and lipid headgroup conformational ensembles ^{53–55}	Scientists developing parameters for MD simulations
Training and target data for coarse grained models	Optimizing parameters of coarse grained models against NMRlipids databank, extracting continuum parameters for membranes.	Scientists developing and using coarse grained MD simulations
Training set for machine learning applicatons	programmatic access to the data and results enables training of machine learning type of models for various applications, such as predictions of membrane properties from composition	Scientists building and using machine learning applications for biomolecules.

Table 1. Examples on applications of the NMRlipids databank.

Molecule and atom naming convention

When analysing simulations, atoms and molecules needs to be often called by their names defined in the simulation trajectory. However, these names often vary between the force fields because the universal naming convention has not been defined for lipids. To enable automatic analyses over simulations with different atom and molecule names in the NMRlipids databank, we have defined unique naming conventions for molecules and atoms therein. Unique abbreviations used in the NMRlipids databank for each molecule are listed in table S2 in the supplementary information. Atom names in a simulation are connected to unique atom names defined for each molecule using mapping files located at `/Scripts/BuildDatabank/mapping_files` in the NMRlipids repository. Also residue names for each atom and whether they belong to headgroup, glycerol backbone, or acyl chain region are defined in the mapping files. For more details about mapping files, see <https://nmrlipids.blogspot.com/2022/04/new-yaml-format-of-mapping-files.html>). In practise, mapping file names and force field specific molecule names corresponding the unique abbreviation (table S2) are given in the COMPOSITION dictionary in README.yaml files.

Adding data into the NMRlipids databank

Anyone can add a new simulation into the NMRlipids databank by first creating an info.yaml file containing the manually entered information listed in table S1 and then adding this file to `/Scripts/BuildDatabank/info_files` folder in the NMRlipids databank repository via pull requests. Pull requests are manually accepted and automatically extracted data for README.yaml files, listed in table S1, will be calculated using the `/Scripts/BuildDatabank/AddData.py` script. Current content of the NMRlipids databank is composed of the simulations found from Zenodo repository with an appropriate licence. Most, but not all, of these trajectories originate from previous NMRlipids projects^{23,25–27}.

Experimental data

Experimental data used in the quality evaluation, currently composed of C–H bond order parameters and x-ray scattering form factors, are stored in */Data/experiments* in the NMRLipids databank repository. Similarly to simulations, each experimental data set has a README.yaml file containing all the relevant information about the experiment. The keys and their descriptions for the experimental data are given in table S3. NMR data currently in the NMRLipids databank are taken from Refs. 26,27,52,58,59 and x-ray scattering data from Refs. 60–64. In addition, some previously unpublished NMR and x-ray scattering data are used. These are measured with previously established methods as described in the supplementary information.

Analyzing the databank

Programs that analyze large data sets from the NMRLipids databank loop over README.yaml files of all simulations, download the raw simulation data to the local computer, and perform the analysis from the downloaded trajectories. When molecule or atom names are needed in the analysis, they will be read from the README.yaml file and mapping files defined therein.

Calculation of C–H bond order parameters

The C–H bond order parameters were calculated directly from the carbon and hydrogen positions using the definition

$$S_{\text{CH}} = \frac{1}{2} \langle 3 \cos^2 \theta - 1 \rangle, \quad (1)$$

where angular brackets denote the ensemble average, i.e., average over all sampled configurations of all lipids in a simulation, and θ is the angle between the C–H bond and the membrane normal. As in previous NMRLipids publications, the order parameters were first calculated separately for each lipid and the standard error of the mean over different lipids was used as the error bar²³. The script that calculates C–H bond order parameters from all simulations in the NMRLipids databank is available at */Scripts/AnalyzeDatabank/calcOrderParameters.py* in the NMRLipids databank repository. The resulting order parameters are stored for all simulations in files named *[lipid_name]OrderParameters.json* at folders in */Data/Simulations* in the NMRLipids databank repository.

Calculation of x-ray scattering form factors

X-ray scattering form factors were calculated with standard equation for symmetric lipid bilayers²⁴

$$F(q) = \int_{-D/2}^{D/2} \Delta\rho_e(z) \cos(zq_z) dz, \quad (2)$$

where $\Delta\rho_e(z)$ is the difference between total and solvent electron densities, and D is the simulation box size in the z-direction. For the calculation of density profiles, atom coordinates were first centred around the centre of mass of lipid molecules for every time frame, and a histogram of these centred positions weighted with the number of electrons in each atom was then calculated with the bin width of 1/3 Å. The script to calculate form factors for all simulations in the NMRLipids databank is available at *Scripts/AnalyzeDatabank/calc_FormFactors.py*. The resulting form factors are stored for all simulations in files named *FormFactor.json* at folders in */Data/Simulations* in the NMRLipids databank repository.

Calculation of a bilayer area per lipid and thickness

Area per lipids of bilayers were calculated by dividing the area of the simulation box with the total number of lipid molecules in the simulation. The script that calculates area per lipids from all simulations in the NMRLipids databank repository is available at *Scripts/AnalyzeDatabank/calcAPL.py* in the NMRLipids databank repository. The resulting area per lipids are stored for all simulations in files named *apl.json* at folders in */Data/Simulations*.

Thicknesses of lipid bilayers were calculated from the intersections of lipid and water electron densities. The script that calculates thickness of all simulations in the NMRLipids databank is available at *Scripts/AnalyzeDatabank/calc_thickness.py* in the NMRLipids databank repository. The resulting thicknesses are stored in files named *thickness.json* at folders in */Data/Simulations* in the NMRLipids databank repository.

Quality evaluation of C–H bond order parameters

As the first step to evaluate simulation qualities against experimental data, each simulation is connected to an experimental data if molar concentrations of all molecules are within ±3 percentage units, charged lipids have the same counterions, and temperatures are within ±2 degrees. For molar concentrations of water, the exact hydration level is considered only for systems with molar water to lipid ratio below 25, otherwise the systems are considered as fully hydrated. In practise, this is implemented by adding the path of the experimental data into the simulation README.yaml file using the */Scripts/BuildDatabank/searchDATABANK.py* script in the NMRLipids databank repository.

The quality of each C–H bond order parameter is estimated by calculating the probability of a simulated value to locate within the error bars of the experimental value. Because conformational ensembles of individual lipids are independent in a fluid lipid bilayer, $\frac{S_{\text{CH}} - \mu}{s/\sqrt{n}}$ has a Student's t-distribution with $n - 1$ degrees of freedom and the probability for a order parameter in simulation to locate within experimental error bars can be estimated from equation

$$P = f\left(\frac{S_{\text{CH}} - (S_{\text{exp}} + \Delta S_{\text{exp}})}{s/\sqrt{n}}\right) - f\left(\frac{S_{\text{CH}} - (S_{\text{exp}} - \Delta S_{\text{exp}})}{s/\sqrt{n}}\right), \quad (3)$$

where $f(t)$ is the Student's t-distribution, μ is the real mean of the order parameter, n is the number of independent sample points for each C–H bond which equals the number of lipids in a simulation, S_{CH} is the sample mean from Eq. 1, s is the variance of S_{CH} calculated over individual lipids, S_{exp} is the experimental value, and ΔS_{exp} its error. The error of $\Delta S_{\text{exp}} = 0.02$ is currently assumed for all experimental order parameters²⁴, yet more accurate may be available in the future⁶⁵. Because lipid bilayer simulations contain at least dozens of lipids, the Student's t-distribution could be safely approximated with a normal distribution. However, the normal distribution gives probability values that are below the numerical accuracy of computers when simulation values are far from experiments. To avoid such numerical instabilities, we use the first order Student's t-distribution having slightly higher probabilities for values far away from the mean. On the other hand, some force fields exhibit too slow dynamics which leads to large error bars in order parameter values⁶⁶. Such large error bars widen the Student's t-distribution in Eq. 3 thereby artificially increasing the probability to find the simulated value within experimental error bars. Therefore, the order parameters with simulation error bars above the experimental error 0.02 are not included in the quality evaluation.

To streamline the comparison between simulations, we define the qualities of different fragments (headgroup, acyl chains or total lipid) within each lipid type in a simulation as

$$P^{\text{frag}}[\text{lipid}] = \frac{\langle P[\text{lipid}] \rangle_{\text{frag}}}{p_{\text{frag}}[\text{lipid}]}, \quad (4)$$

where $\langle P[\text{lipid}] \rangle_{\text{frag}}$ is the average of individual C–H bond order parameters qualities within the fragment, $p_{\text{frag}}[\text{lipid}]$ is the percentage of order parameters for which the quality is available within the fragment, and frag can be sn-1, sn-2, headgroup or total (all order parameters within a molecule). The overall quality of different fragments in a simulation are then defined as a molar fraction weighted average over different lipid components

$$P^{\text{frag}} = \sum_{\text{lipid}} \chi_{\text{lipid}} P^{\text{frag}}[\text{lipid}], \quad (5)$$

where χ_{lipid} is the molar fraction of a lipid in the bilayer.

The quality evaluation of order parameters is implemented in */Scripts/BuildDatabank/QualityEvaluation.py* in the NMRLipids databank repository. The resulting qualities for each order parameters are stored in files named *[lipid_name]_OrderParameters_quality.json* for individual lipids in files named *[lipid_name]_FragmentQuality.json*, and for overall quality for fragments in files named *system_quality.json* at folders in */Data/Simulations* in the NMRLipids databank repository.

Quality evaluation of x-ray scattering form factors

Because experiments give form factors only in relative intensity scale, they should scaled before comparing with the simulation data. Here we use the scaling coefficient for experimental intensities defined in the SIMtoEXP program⁶⁷

$$k_e = \frac{\sum_{i=1}^{N_q} \frac{|F_s(q_i)| |F_e(q_i)|}{(\Delta F_e(q_i))^2}}{\sum_{i=1}^{N_q} \frac{|F_e(q_i)|^2}{(\Delta F_e(q_i))^2}}, \quad (6)$$

where $F_s(q)$ and $F_e(q)$ are form factors from a simulation and experiment, respectively, $\Delta F_e(q)$ is the error of the experimental form factor, and summation goes over the experimentally available N_q points.

Also a quality measure based on differences in simulated and experimental form factors accross available q-range is defined in the SIMtoEXP program⁶⁷. However, the lobe heights in simulated form factors depend on the simulation box size as shown in Fig. S2. Therefore, the quality measure defined in SIMtoEXP would also depend on the simulation box size. Nevertheless, locations of form factor minima are independent on simulation box size in Fig. S2. Here we use the location of the first form factor minima for the quality evaluation because automatic detection of the location of second minima is inaccurate for some experimental data due to fluctuations, such as for the POPE data in Figs. 2 D and E. The first minimum correlates well with the thickness of a membrane (Fig. 2 F), although the correlation of the second minima would be even stronger (Fig. S1). In practise, we first filter the fluctuations from the form factor data using Savitzky-Golay filter (window lenght 30 and polynomial

order 1) and locate the first minima above $0.1\text{ }\text{\AA}^2$ from both simulation and experimental data. The quality of a form factor is then defined as Euclidian distance between the minima in simulated and experimental form factors, $FF_q = |FF_{\min}^{\text{sim}} - FF_{\min}^{\text{exp}}|$.

The quality evaluation of form factors is implemented in `/Scripts/BuildDatabank/QualityEvaluation.py` in the NMRLipids databank repository. The resulting form factor qualities are stored in files named `FormFactorQuality.json` at folders in `/Data/Simulations` in the NMRLipids databank repository.

Analysing anisotropic diffusion of water in membranes environment from the NMRLipids databank

Water permeability through membranes was calculated from equation $P = r/2c_w$, where r is the rate of permeation events per time and area, and $c_w=33.3679(\text{nm})^{-3}$ which is the concentration of water in bulk⁴⁰. The number of permeation events in each trajectory was calculated using the code by Camilo et al.⁴¹, available at <https://github.com/crobertocamilo/MD-permeation>. The code that calculates permeabilities for all simulations in the NMRLipids databank is available at `/scripts/calcMD-PERMEATION.py` and the resulting permeabilities are stored at `/Data/MD-PERMEATION` in the repository containing all analyses specific for this publication at <https://github.com/NMRLipids/DataBankManuscript/>. This repository is organized similarly to the NMRLipids databank repository, enabling the upcycling of also the analyzed data without overloading the main NMRLipids databank repository.

The lateral diffusion of water along the membrane surface, $D_{||}$, was calculated with Einstein's equation using *-lateral* option in `gmx msd` program within the Gromacs software package⁶⁸. The code that calculates $D_{||}$ for water from all simulations in the NMRLipids databank is available at `/scripts/calcWATERdiffusion.py`, and the resulting diffusion coefficients are stored at `/Data/WATERdiffusion` in the repository at <https://github.com/NMRLipids/DataBankManuscript/>.

Water diffusion in the perpendicular direction of lipid bilayers in a multilamellar stack was estimated from the Tanner equation $D_{\perp} = \frac{D_{||}Pz_w}{D_{||}+Pz_w}$ ^{48,49}, where the water layer thickness, z_w , was estimated by subtracting bilayer thickness from the size of the simulation box in membrane normal direction.

Calculation of lipid flip-flops

Flip-flop rates were calculated using `AssignLeaflets` and `FlipFlop` tools from LiPyphilic package⁶⁹. Headgroup atoms of each molecule defined in the mapping file were used to determine in which leaflet the molecule locates. The midplane cut-off defining the region between leaflets was 1 nm and frame cut-off was 100. This means that if the headgroup of a molecule entered within the distance of 1 nm from the bilayer midplane and was found from the opposing leaflet after 100 steps, this event was considered as a successfull flip-flop event. A code that finds flip-flop rates from all simulations in the databank is available at `scripts/FlipFlop.py` and the resulting flip-flop events at `Data/Flipflops/` in the repository at <https://github.com/NMRLipids/DataBankManuscript/>.

References

- van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* **9**, 112–124, DOI: [10.1038/nrm2330](https://doi.org/10.1038/nrm2330) (2008).
- Torres, M. et al. Lipids in pathophysiology and development of the membrane lipid therapy: New bioactive lipids. *Membranes* **11** (2021).
- Antila, H. S. et al. Emerging era of biomolecular membrane simulations: Automated physically-justified force field development and quality-evaluated databanks. *The J. Phys. Chem. B* **126**, 4169–4183 (2022).
- Gupta, C., Sarkar, D., Tieleman, D. P. & Singharoy, A. The ugly, bad, and good stories of large-scale biomolecular simulations. *Curr. Opin. Struct. Biol.* **73**, 102338, DOI: <https://doi.org/10.1016/j.sbi.2022.102338> (2022).
- Johnson, G. T. et al. cellpack: a virtual mesoscope to model and visualize structural systems biology. *Nat. Methods* **12**, 85–91, DOI: [10.1038/nmeth.3204](https://doi.org/10.1038/nmeth.3204) (2015).
- Thornburg, Z. R. et al. Fundamental behaviors emerge from simulations of a living minimal cell. *Cell* **185**, 345–360.e28, DOI: <https://doi.org/10.1016/j.cell.2021.12.025> (2022).
- Montelione, G. T. et al. Recommendations of the wwpdb nmr validation task force. *Structure* **21**, 1563–1570, DOI: [10.1016/j.str.2013.07.021](https://doi.org/10.1016/j.str.2013.07.021) (2013).
- Feig, M., Abdullah, M., Johnsson, L. & Pettitt, B. Large scale distributed data repository: design of a molecular dynamics trajectory database. *Futur. Gener. Comput. Syst.* **16**, 101–110, DOI: [https://doi.org/10.1016/S0167-739X\(99\)00039-4](https://doi.org/10.1016/S0167-739X(99)00039-4) (1999).
- Tai, K. et al. Biosimgrid: towards a worldwide repository for biomolecular simulations. *Org. Biomol. Chem.* **2**, 3219–3221, DOI: [10.1039/B411352G](https://doi.org/10.1039/B411352G) (2004).

10. Silva, C. G. *et al.* P-found: The protein folding and unfolding simulation repository. In *2006 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology*, 1–8, DOI: [10.1109/CIBCB.2006.330978](https://doi.org/10.1109/CIBCB.2006.330978) (2006).
11. Abraham, M. *et al.* Sharing data from molecular simulations. *J. Chem. Inf. Model.* **59**, 4093–4099 (2019).
12. Hildebrand, P. W., Rose, A. S. & Tiemann, J. K. Bringing molecular dynamics simulation data into view. *Trends Biochem. Sci.* **44**, 902–913, DOI: [10.1016/j.tibs.2019.06.004](https://doi.org/10.1016/j.tibs.2019.06.004) (2019).
13. Hospital, A., Battistini, F., Soliva, R., Gelpí, J. L. & Orozco, M. Surviving the deluge of biosimulation data. *WIREs Comput. Mol. Sci.* **10**, e1449, DOI: <https://doi.org/10.1002/wcms.1449> (2020). <https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1449>.
14. Abriata, L. A., Lepore, R. & Dal Peraro, M. About the need to make computational models of biological macromolecules available and discoverable. *Bioinformatics* **36**, 2952–2954, DOI: [10.1093/bioinformatics/btaa086](https://doi.org/10.1093/bioinformatics/btaa086) (2020). <https://academic.oup.com/bioinformatics/article-pdf/36/9/2952/33180880/btaa086.pdf>.
15. Rodríguez-Espigares, I. *et al.* Gpcrmd uncovers the dynamics of the 3d-gpcrome. *Nat. Methods* **17**, 777–787, DOI: [10.1038/s41592-020-0884-y](https://doi.org/10.1038/s41592-020-0884-y) (2020).
16. Meyer, T. *et al.* Model (molecular dynamics extended library): A database of atomistic molecular dynamics trajectories. *Structure* **18**, 1399–1409, DOI: <https://doi.org/10.1016/j.str.2010.07.013> (2010).
17. van der Kamp, M. W. *et al.* Dynameomics: A comprehensive database of protein dynamics. *Structure* **18**, 423–435, DOI: <https://doi.org/10.1016/j.str.2010.01.012> (2010).
18. Newport, T. D., Sansom, M. S. & Stansfeld, P. J. The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Res.* **47**, D390–D397, DOI: [10.1093/nar/gky1047](https://doi.org/10.1093/nar/gky1047) (2018). <https://academic.oup.com/nar/article-pdf/47/D1/D390/27437085/gky1047.pdf>.
19. Suarez-Leston, F. *et al.* Supepmem: A database of innate immune system peptides and their cell membrane interactions. *Comput. Struct. Biotechnol. J.* **20**, 874–881, DOI: <https://doi.org/10.1016/j.csbj.2022.01.025> (2022).
20. Hospital, A. *et al.* BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Res.* **44**, D272–D278, DOI: [10.1093/nar/gkv1301](https://doi.org/10.1093/nar/gkv1301) (2016). <https://academic.oup.com/nar/article-pdf/44/D1/D272/16661850/gkv1301.pdf>.
21. Mixcoha, E., Rosende, R., Garcia-Fandino, R. & Piñeiro, A. Cyclo-lib: a database of computational molecular dynamics simulations of cyclodextrins. *Bioinformatics* **32**, 3371–3373, DOI: [10.1093/bioinformatics/btw289](https://doi.org/10.1093/bioinformatics/btw289) (2016). <https://academic.oup.com/bioinformatics/article-pdf/32/21/3371/7889578/btw289.pdf>.
22. Bekker, G.-J., Kawabata, T. & Kurisu, G. The biological structure model archive (bsm-arc): an archive for in silico models and simulations. *Biophys. Rev.* **12**, 371–375, DOI: [10.1007/s12551-020-00632-5](https://doi.org/10.1007/s12551-020-00632-5) (2020).
23. Botan, A. *et al.* Toward atomistic resolution structure of phosphatidylcholine headgroup and glycerol backbone at different ambient conditions. *J. Phys. Chem. B* **119**, 15075–15088 (2015).
24. Ollila, O. S. & Pabst, G. Atomistic resolution structure and dynamics of lipid bilayers in simulations and experiments. *Biochim. Biophys. Acta* **1858**, 2512 – 2528 (2016).
25. Catte, A. *et al.* Molecular electrometer and binding of cations to phospholipid bilayers. *Phys. Chem. Chem. Phys.* **18**, 32560–32569 (2016).
26. Antila, H. *et al.* Headgroup structure and cation binding in phosphatidylserine lipid bilayers. *J. Phys. Chem. B* **123**, 9066–9079 (2019).
27. Bacle, A. *et al.* Inverse conformational selection in lipid–protein binding. *J. Am. Chem. Soc.* **143**, 13701–13709 (2021).
28. Wilkinson, M. D. *et al.* The fair guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018, DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18) (2016).
29. Kurki, M., Poso, A., Bartos, P. & Miettinen, M. S. Structure of popc lipid bilayers in opls3e force field. *J. Chem. Inf. Model.* **0**, null, DOI: [10.1021/acs.jcim.2c00395](https://doi.org/10.1021/acs.jcim.2c00395) (0).
30. Steck, T. & Lange, Y. How slow is the transbilayer diffusion (flip-flop) of cholesterol? *Biophys. J.* **102**, 945–946, DOI: <https://doi.org/10.1016/j.bpj.2011.10.059> (2012).
31. Parisio, G., Ferrarini, A. & Sperotto, M. M. Model studies of lipid flip-flop in membranes. *Int. J. Adv. Eng. Sci. Appl. Math.* **8**, 134–146 (2016).

32. Gu, R.-X., Baoukina, S. & Tieleman, D. P. Cholesterol flip-flop in heterogeneous membranes. *J. Chem. Theory Comput.* **15**, 2064–2070 (2019).
33. Javanainen, M. & Martinez-Seara, H. Rapid diffusion of cholesterol along polyunsaturated membranes via deep dives. *Phys. Chem. Chem. Phys.* **21**, 11660–11669 (2019).
34. Baral, S., Levental, I. & Lyman, E. Composition dependence of cholesterol flip-flop rates in physiological mixtures. *Chem. Phys. Lipids* **232**, 104967, DOI: <https://doi.org/10.1016/j.chemphyslip.2020.104967> (2020).
35. Hansen, S., Lehr, C.-M. & Schaefer, U. F. Improved input parameters for diffusion models of skin absorption. *Adv. Drug Deliv. Rev.* **65**, 251–264, DOI: <https://doi.org/10.1016/j.addr.2012.04.011> (2013). Modeling the human skin barrier - Towards a better understanding of dermal absorption.
36. Wen, J., Koo, S. M. & Lape, N. How sensitive are transdermal transport predictions by microscopic stratum corneum models to geometric and transport parameter input? *J. Pharm. Sci.* **107**, 612–623, DOI: <https://doi.org/10.1016/j.xphs.2017.09.015> (2018).
37. Nitsche, L. C., Kasting, G. B. & Nitsche, J. M. Microscopic models of drug/chemical diffusion through the skin barrier: Effects of diffusional anisotropy of the intercellular lipid. *J. Pharm. Sci.* **108**, 1692–1712, DOI: <https://doi.org/10.1016/j.xphs.2018.11.014> (2019).
38. Roberts, M. S. *et al.* Topical drug delivery: History, percutaneous absorption, and product development. *Adv. Drug Deliv. Rev.* **177**, 113929, DOI: <https://doi.org/10.1016/j.addr.2021.113929> (2021).
39. Topgaard, D. Chapter 1 translational motion of water in biological tissues – a brief primer. In *Advanced Diffusion Encoding Methods in MRI*, 1–11 (The Royal Society of Chemistry, 2020).
40. Venable, R. M., Krämer, A. & Pastor, R. W. Molecular dynamics simulations of membrane permeability. *Chem. Rev.* **119**, 5954–5997 (2019).
41. Camilo, C. R. d. S., Ruggiero, J. R. & de Araujo, A. S. A method for detection of water permeation events in molecular dynamics simulations of lipid bilayers. *Braz. J. Phys.* **52**, 1–13 (2022).
42. Rudakova, M., Filippov, A. & Skirda, V. Water diffusivity in model biological membranes. *Appl. Magn. Reson.* **27**, 519 (2004).
43. Jansen, M. & Blume, A. A comparative study of diffusive and osmotic water permeation across bilayers composed of phospholipids with different head groups and fatty acyl chains. *Biophys. J.* **68**, 997–1008, DOI: [https://doi.org/10.1016/S0006-3495\(95\)80275-4](https://doi.org/10.1016/S0006-3495(95)80275-4) (1995).
44. Nagle, J. F., Mathai, J. C., Zeidel, M. L. & Tristram-Nagle, S. Theory of Passive Permeability through Lipid Bilayers . *J. Gen. Physiol.* **131**, 77–85 (2008).
45. Frallicciardi, J., Melcr, J., Siginou, P., Marrink, S. J. & Poolman, B. Membrane thickness, lipid phase and sterol type are determining factors in the permeability of membranes to small solutes. *Nat. Commun.* **13**, 1605 (2022).
46. Khakimov, A. M., Rudakova, M. A., Dorogintskii, M. M. & Filippov, A. V. Temperature dependence of water self-diffusion through lipid bilayers assessed by NMR. *Biophysics* **53**, 147–152, DOI: [10.1134/s000635090802005x](https://doi.org/10.1134/s000635090802005x) (2008).
47. Kadaoluwa Pathirannahalage, S. P. *et al.* Systematic comparison of the structural and dynamic properties of commonly used water models for molecular dynamics simulations. *J. Chem. Inf. Model.* **61**, 4521–4536 (2021).
48. Tanner, J. E. Transient diffusion in a system partitioned by permeable barriers. application to nmr measurements with a pulsed field gradient. *The J. Chem. Phys.* **69**, 1748–1754, DOI: [10.1063/1.436751](https://doi.org/10.1063/1.436751) (1978).
49. Wästerby, P., Orädd, G. & Lindblom, G. Anisotropic water diffusion in macroscopically oriented lipid bilayers studied by pulsed magnetic field gradient nmr. *J. Magn. Reson.* **157**, 156–159, DOI: <https://doi.org/10.1006/jmre.2002.2583> (2002).
50. Jumper, J. *et al.* Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).
51. Melcr, J. *et al.* Accurate binding of sodium and calcium to a popc bilayer by effective inclusion of electronic polarization. *J. Phys. Chem. B* **122**, 4546–4557 (2018).
52. Melcr, J., Ferreira, T. M., Jungwirth, P. & Ollila, O. H. S. Improved cation binding to lipid bilayers with negatively charged pops by effective inclusion of electronic polarization. *J. Chem. Theo. Comput.* **16**, 738–748 (2020).
53. Yu, Y. *et al.* Charmm36 lipid force field with explicit treatment of long-range dispersion: Parametrization and validation for phosphatidylethanolamine, phosphatidylglycerol, and ether lipids. *J. Chem. Theo. Comput.* **17**, 1581–1595 (2021).
54. Dickson, C. J., Walker, R. C. & Gould, I. R. Lipid21: Complex lipid membrane simulations with amber. *J. Chem. Theo. Comput.* **18**, 1726–1736 (2022).

55. Grote, F. & Lyubartsev, A. P. Optimization of lipids force field parameters describing headgroups of phospholipids. *J. Phys. Chem. B* **124**, 8784–8793 (2020).
56. Gowers, R. J. *et al.* Mdanalysis: a python package for the rapid analysis of molecular dynamics simulations (2019).
57. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. Mdanalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).
58. Scherer, P. & Seelig, J. Structure and dynamics of the phosphatidylcholine and the phosphatidylethanolamine head group in L-M fibroblasts as studied by deuterium nuclear magnetic resonance. *EMBO J.* **6** (1987).
59. Ferreira, T. M. *et al.* Cholesterol and POPC segmental order parameters in lipid membranes: solid state ^1H - ^{13}C NMR and MD simulation studies. *Phys. Chem. Chem. Phys.* **15**, 1976–1989 (2013).
60. Kučerka, N. *et al.* Lipid bilayer structure determined by the simultaneous analysis of neutron and x-ray scattering data. *Biophys. J.* **95**, 2356 – 2367 (2008).
61. Kučerka, N., Nieh, M. P. & Katsaras, J. Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochim. Biophys. Acta* **1808**, 2761–2771 (2011).
62. Pan, J. *et al.* Molecular structures of fluid phase phosphatidylglycerol bilayers as determined by small angle neutron and x-ray scattering. *Biochimica et Biophys. Acta (BBA) - Biomembr.* **1818**, 2135 – 2148 (2012).
63. Pan, J. *et al.* The molecular structure of a phosphatidylserine bilayer determined by scattering and molecular dynamics simulations. *Soft Matter* **10**, 3716–3725 (2014).
64. Kučerka, N. *et al.* Molecular structures of fluid phosphatidylethanolamine bilayers obtained from simulation-to-experiment comparisons and experimental scattering density profiles. *The J. Phys. Chem. B* **119**, 1947–1956 (2015).
65. Wurl, A., Saalwächter, K. & Mendes Ferreira, T. Time-domain r-pdlf nmr for molecular structure determination in complex lipid membranes. *Magn. Reson. Discuss.* **2022**, 1–18, DOI: [10.5194/mr-2022-19](https://doi.org/10.5194/mr-2022-19) (2022).
66. Antila, H. S., M. Ferreira, T., Ollila, O. H. S. & Miettinen, M. S. Using open data to rapidly benchmark biomolecular simulations: Phospholipid conformational dynamics. *J.Chem. Inf. Model.* **61**, 938–949 (2021).
67. Kučerka, N., Katsaras, J. & Nagle, J. Comparing membrane simulations to scattering experiments: Introducing the SIMtoEXP software. *J. Membr. Biol.* **235**, 43–50 (2010).
68. Bauer, P., Hess, B. & Lindahl, E. Gromacs 2022.3 manual, DOI: [10.5281/zenodo.7037337](https://doi.org/10.5281/zenodo.7037337) (2022).
69. Smith, P. & Lorenz, C. D. Lipyphilic: A python toolkit for the analysis of lipid membrane simulations. *J. Chem. Theory Comput.* **17**, 5907–5919, DOI: [10.1021/acs.jctc.1c00447](https://doi.org/10.1021/acs.jctc.1c00447) (2021).

Acknowledgements

Author contributions statement

Must include all authors, identified by initials, for example: A.A. conceived the experiment(s), A.A. and B.A. conducted the experiment(s), C.A. and D.A. analysed the results. All authors reviewed the manuscript.

Additional information