# NMRlipids Databank: Overlay Databank of Lipid Membrane Simulations Arising from Open Collaboration

**Anne Kiirikki**[1], **...**[2], **and O. H. Samuli Ollila**[1,*]

[1]University of Helsinki, Institute of Biotechonology, Helsinki, Finland
[2]Affiliation, department, city, postcode, country
[*]samuli.ollila@helsinki.fi

## ABSTRACT

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

## 1 Introduction

While MD simulations of systems with few biomolecules continue having major contributions in our understanding on biology and drug design, researches are now searching new avenues by building models not only for individual molecules but whole cells or organelles using interdisciplinary approaches[1–3].

The importance of sharing MD simulation data following the FAIR principles[4] has been widely recognized[5–12], and databanks are emerging[12–19]. The relevance of quality evaluation of simulation trajectories in databanks regarding technical details of simulations and accuracy of the underlying physical description of the system (force field) has become evident[6,10,13] and such quality evaluation has in some cases also been implemented[13,15]. However, straightforward quality comparisons between individual simulations or force fields within these databanks remain challenging. While importance of such databanks for MD simulations is widely recognized[5–12] and different kinds of approaches are emerging[12–19], generally accepted protocols and best practices are still under active development.

Here we present a solution for lipid bilayers based on overlay databank structure illustrated in Fig. 1. The concept of overlay databank is developed here to solve the practical challenges in generating databanks of MD simulation data enabling flexible analyses over large sets of simulation data, but it can be potentially used for wide range of situations. Particularly when storage of raw data requires significant resources and final outcomes or best practices are not yet clear, overlay databank approach lowers the barrier to start without compromizing the long term stability or scalability.

The practical relevance of the NMRlipids databank is exemplified by automatic quality evaluation and ranking of large amount of MD simulation data, data-driven analysis detecting correlations between properties of model cell membranes and analyses of rare phenomena that are beyond the scope of standard MD simulation studies. The NMRlipids databank provides 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.

## 2 Results

### 2.1 Overlay structure of the NMRlipids databank

The three layer structure of the NMRlipids databank, illustrated in Fig. 1 a, enables an efficient way to share and upcycle large MD simulation trajectories. Layer 1 contains raw simulation data that can locate in any publicly available repository or server offering long term stability and permanent links to the data, such as digital object identifiers. Layer 2 contains all the relevant information on the simulations, including links to the raw data in layer 1 and relevant metadata describing the systems. Layer 3 is the application layer containing outputs from the databank, such as graphical user interface (`http://www.databank.nmrlipids.fi/`) and results from analyses desribed in sections below.

Layer 2 is the core of the databank. In addition to the metadata and links to raw simulation data, it contains universal naming conventions for the molecules and atoms in the databank, computer programs to generate and analyse the databank, 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 is the git repository which is currently available at `https://github.com/NMRLipids/Databank`. Detailed description

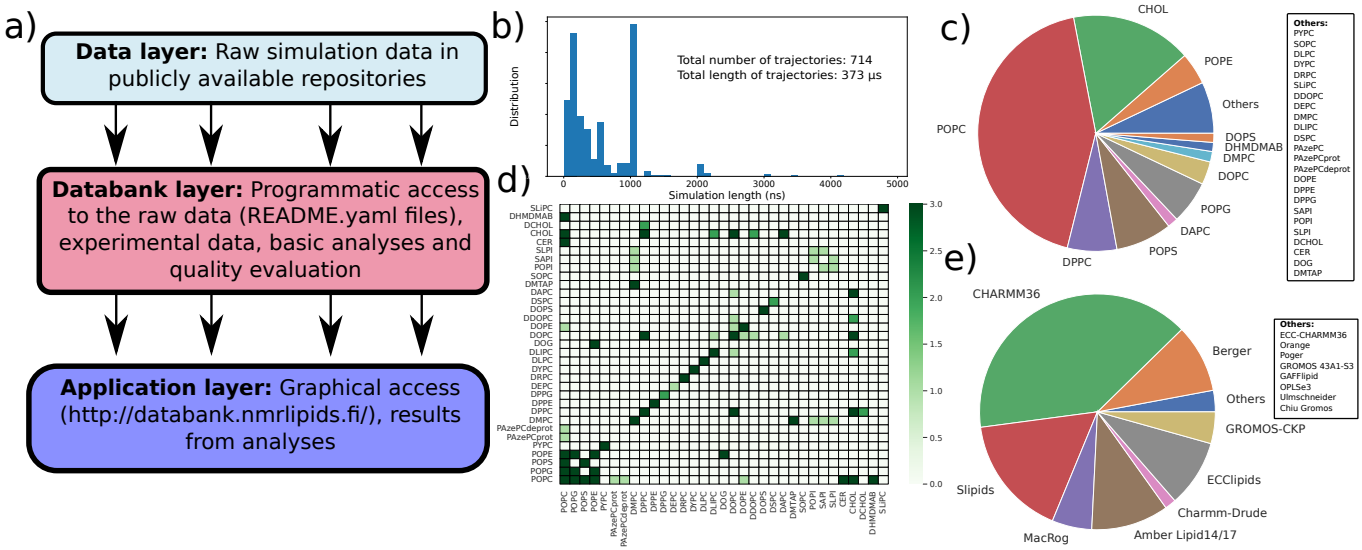


**Figure 1.** a) Structure of an overlay databank. More detailed structure of the layer 2 in the NMRlipids databank is illustrated in Fig. S6 in the SI. b) Distribution of the lengths of the trajectories, total number of trajectories and total lenght of the simulations in the NMRlipids databank. c) Distribution of lipids present in the trajectories in the NMRlipids databank. Lipids occuring 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. The figures and numbers are created on 9th of May 2022.

on its content is in the supplementary information. Applications in layer 3 uses the information in layer 2 to access raw data in layer 1 to perform automatic analyses over large sets of simulations in the databank.

### 2.2 Content of the databank

Currently the databank is composed of approximately 500 trajectories with the total length of approximately 231 microseconds of which most are contributed for the previous publications from the NMRlipids open collaboration[20–23]. The distribution of lipids, force fields, length of the trjajectories and available binary mixtures are shown in Fig. 1.

### 2.3 Quality evaluation of force fields

Quality of membrane simulations with different force fields have been evaluated against experimental data during parameterization and in separate comparison studies[20,21,24–27], but universal quality measure for membrane simulations is not defined and controversial results are often reported from simulations[28]. The lack of universal quality measure for membrane simulations complicates the selection of proper simulation parameters and the estimation of reliability in simulations, thereby being a major obstacle in many applications of membrane MD simulations. To enable the rapid quality evaluation of membrane MD simulations, a quantitative quality measure based on C-H bond order parameters from NMR experiments is defined in the NMRlipids databank. This is complemented by comparing the locations of x-ray scattering form factor minima against experiments. Both of these are robust experimental measurables that can be directly connected to the simulation data[24]. The used quality measures are defined in detail in the supplementary information.

Simulations with the highest scores based on order parameter evaluation for all simulations are shown in Fig. 2 A and for POPE lipids in Fig. 2 B. The direct comparison with experiments for the best simulation, POPC with OPLS3e, is shown in Fig. 2 C, where discrepancies with experiments are observed only for the third carbon in the *sn*-1 chain, and carbons 8 and 9 close to the double bond in *sn*-2 chain. The best overall quality for POPE is found in GROMOS-CKP simulation, for which the direct comparison in Fig. 2 D show minor differences in acyl chains, but major discrepancies in glycerol backbone and in the first carbon of *sn*-2 chain. These parts are better described for POPE by CHARMM36 parameters (direct comparison shown in Fig. 2 E), but its overall quality suffers from the overestimated order in acyl chains.

### 2.4 Finding the best models for PC and PE lipid mixtures

The power of the NMRlipids databank in finding the best models for a certain lipid bilayer system is exemplified here for mixtures of PE and PC lipids. Area per lipids from available POPC and POPE bilayer simulations in the databank from different

A)

| | OPquality | Headgroup | tails | FFquality | Force field | Molecules | Temperature | ID |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.83 | 0.76 | 0.86 | 0.15 | OPLS3e | POPC:SOL (200:8859) | 300.00 | 1 |
| 2 | 0.76 | 0.61 | 0.83 | 0.45 | Slipids | POPC:SOL (512:23943) | 298.00 | 617 |
| 3 | 0.69 | 0.63 | 0.73 | 0.55 | MacRog | POPC:SOL (128:5120) | 300.00 | 457 |
| 4 | 0.61 | 0.64 | 0.60 | 0.55 | MacRog | POPC:SOL (288:14400) | 298.00 | 63 |
| 5 | 0.60 | 0.65 | 0.58 | 0.55 | ECC-lipids | POPC:SOL (128:6400) | 300.00 | 573 |
| 6 | 0.60 | 0.47 | 0.66 | 0.15 | Lipid17 | POPC:SOL (64:3200) | 298.15 | 639 |
| 7 | 0.59 | 0.59 | 0.59 | 1.14 | CHARMM36 | POPC:CHOL:SOL (1024:128:57600) | 298.15 | 426 |
| 8 | 0.58 | 0.01 | 0.86 | 0.95 | Berger | POPC:SOL (256:10342) | 300.00 | 115 |
| 9 | 0.56 | 0.52 | 0.58 | 1.24 | CHARMM36 | POPC:CHOL:SOL (64:8:3600) | 298.15 | 109 |

B)

| | OPquality | Headgroup | sn-1 | sn-2 | FFquality | Force field | Molecules | Temperature | ID |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.53 | 0.29 | 0.83 | 0.48 | 0.4 | GROMOS-CKP | POPE:SOL (500:25000) | 310.0 | 400 |
| 2 | 0.52 | 0.01 | 0.79 | 0.77 | 0.1 | Slipids | POPE:SOL (500:25000) | 310.0 | 414 |
| 3 | 0.51 | 0.10 | 0.87 | 0.57 | | Slipids | POPE:SOL (336:13460) | 310.0 | 29 |
| 4 | 0.51 | 0.10 | 0.86 | 0.57 | | Slipids | POPE:SOL (336:13460) | 310.0 | 74 |
| 5 | 0.44 | 0.54 | 0.52 | 0.27 | 1.3 | CHARMM36 | SOL:POPE (5760:144) | 310.0 | 430 |
| 6 | 0.38 | 0.32 | 0.56 | 0.26 | 0.3 | CHARMM36-UA | POPE:SOL (336:15254) | 310.0 | 352 |
| 7 | 0.37 | 0.27 | 0.57 | 0.28 | 0.3 | CHARMM36-UA | POPE:SOL (336:15254) | 310.0 | 233 |
| 8 | 0.25 | 0.53 | 0.14 | 0.07 | 1.7 | CHARMM36 | SOL:POPE (25000:500) | 310.0 | 67 |
| 9 | 0.19 | 0.09 | 0.12 | 0.35 | 1.6 | Lipid17 | POPE:SOL (500:25000) | 310.0 | 195 |

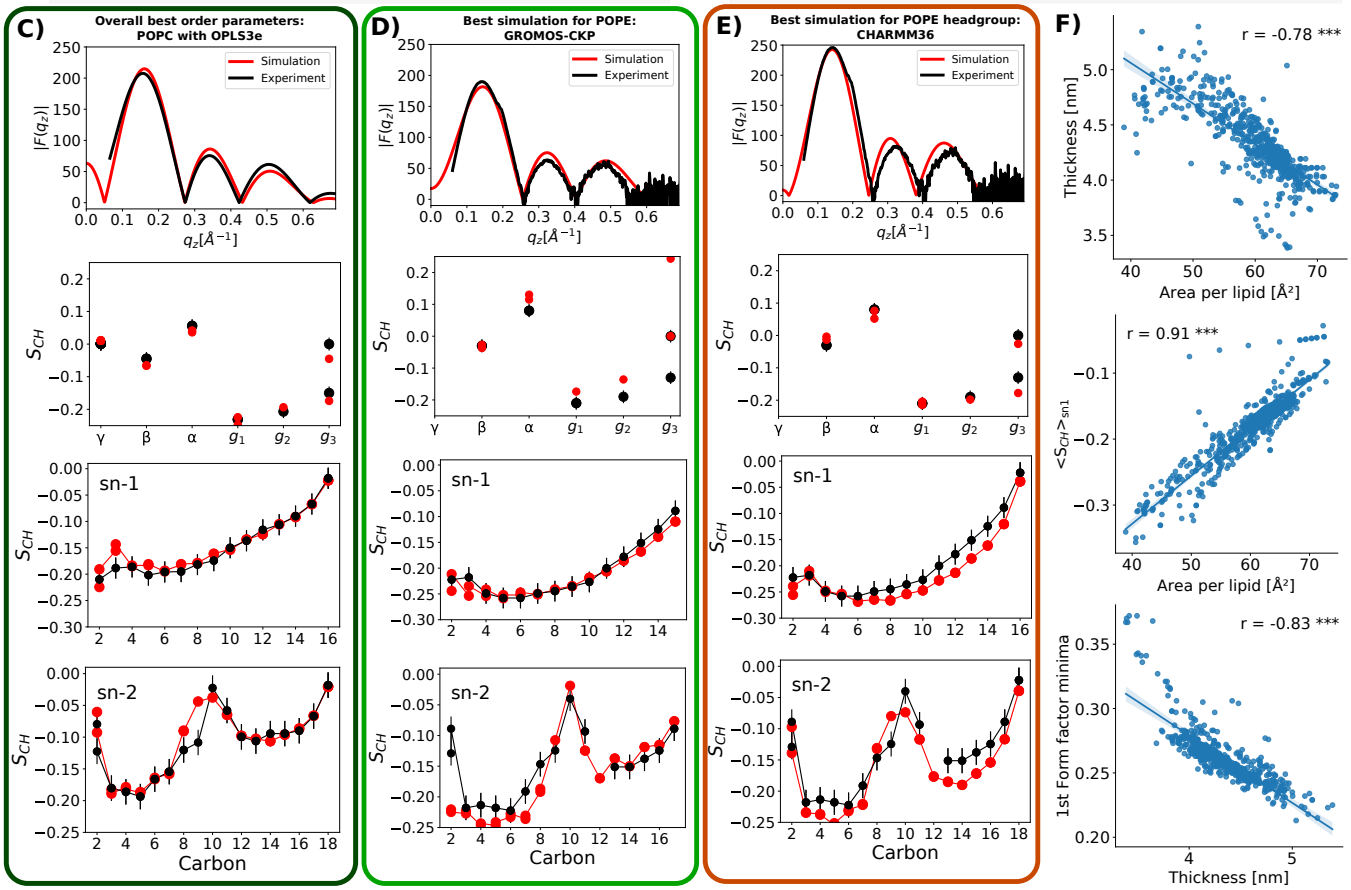


**Figure 2.** A) Best simulations ranked based on overall order parameter quality. B) Best simulations ranked based on the overall order parameter quality of POPE lipid. C) Scatter plots and Pearson correlation coefficients for the area per lipid, thickness, and experimental observables. All correlation coefficients have p-value below 0.001. For more correlations see Fig. S1. D)-F) Evaluation against experimental data exemplified for a simulation with the best overall order parameter quality (D), the best quality for POPE lipid (E), and the headgroup quality for POPE (F).

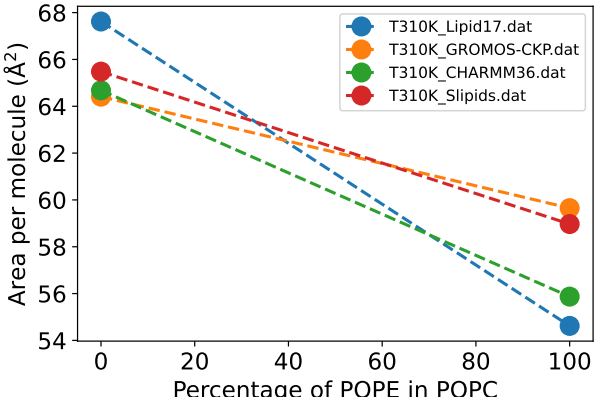


**Figure 3.** Differences in area per lipids of POPC and POPE bilayers between different force fields.

force fields at 310 K are shown in Fig. 3. Because the area per lipid strongly correlates with the average order of *sn*-1 chain (Fig. 2 F), we use order parameter qualities to evaluate which simulations predict the best area per lipids for POPC and POPE. Among the force fields with area per lipids shown in Fig. 3, Slipids ranks 1st for POPC with a clear difference to others, and 2nd for POPE with only marginally lower quality than GROMOS-CKP (Figs. 2 A and B). The direct comparison with experiments in Fig. S3 in the supplementary information reveals that GROMOS-CKP and CHARMM36 predicts too ordered (too low area per lipid) bilayer for POPC, while the order parameters in Lipid17 are lower than in experiments indicating overestimated area per lipid. For POPE, order parameters in Slipids and GROMOS-CKP are close to experiments suggesting that their are per lipids are most reasonable, while CHARMM36 and Lipid17 are too packed thereby overestimating the order parameters.

In conclusion, the quality evaluation based on the NMRlipids databank suggests that the Slipids parameters would give the most reliable description for membrane packing of POPC:POPE mixtures. Slipids was a relatively good model also for POPC:POPS mixtures in similar comparisons utilizing the preliminary version of the NMRlipids databank[28], although charged membranes are more complicated as the counterion binding affinity plays a significant role in membrane packing[22]. On the other hand, CHARMM36 overcomes Slipids when the focus is on conformational ensembles of headgroups in different types of lipids because the accuracy of glycerol backbone is low in Slipids[23]. Altogether, these examples demonstrate how the NMRlipids databank can be used to navigate in the complex landscape of lipid force field quality.

### 2.5 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 skin[29–32], and in MRI imaging[33]. Several models on how water dynamics depends on membrane properties have been proposed based on different experimental and theoretical methods, yet consensus has not been reached[34–39]. MD simulations have been also used in these endeavours, but systematic studies are limited by the accessible amount of data as only few water permeation events are typically observed in a single MD simulation trajectory[38,40]. Therefore, this serves as an excellent example where the large amount of available MD simulations in the NMRlipids databank can be utilized to answer important biological questions.

To this end, we first calculated the water permeability through membranes from all the simulations in the NMRlipids databank. Resulting permeabilities are shown as a function of temperature, membrane thickness, area per lipid, and acyl chain order in Figs. 4 B-E. The values for water permeabilities from simulations vary between 0.3 and 322 $\mu$m/s with the mean and median of 14 $\mu$m/s and 8 $\mu$m/s, respectively. These values have the same order of magnitude as the experimentally determined diffusive permeability coefficients, but are on average below the values reported for PC lipids in liquid crystalline phase, 190-330 $\mu$m/s[41]. To analyze how water permeability through membranes depend on bilayer properties on average, we calculated the average permeabilities over all systems with the value of temperature, thickness, area per lipid, or acyl chain order within a fixed range. These averaged permeabilities also shown in Figs. 4 with black dots. 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. 4 B. Water permeability decreases with increasing membrane packing, i.e., with decreasing area per lipid and increasing thickness and acyl chain order. Linear dependence of permeability on area per lipid and thickness is previously reported from simulations and experiments for some systems, but not for all[34,38,39]. Our results suggest on average linear dependence for thicknesses above $\sim$3.9 nm and area per lipids below $\sim$69 Å$^2$ (insets in Figs. 4 C-D), yet stronger dependence is observed for more loosely packed membranes. Clear dependencies of permeability on charged lipids, cholesterol, POPE, or

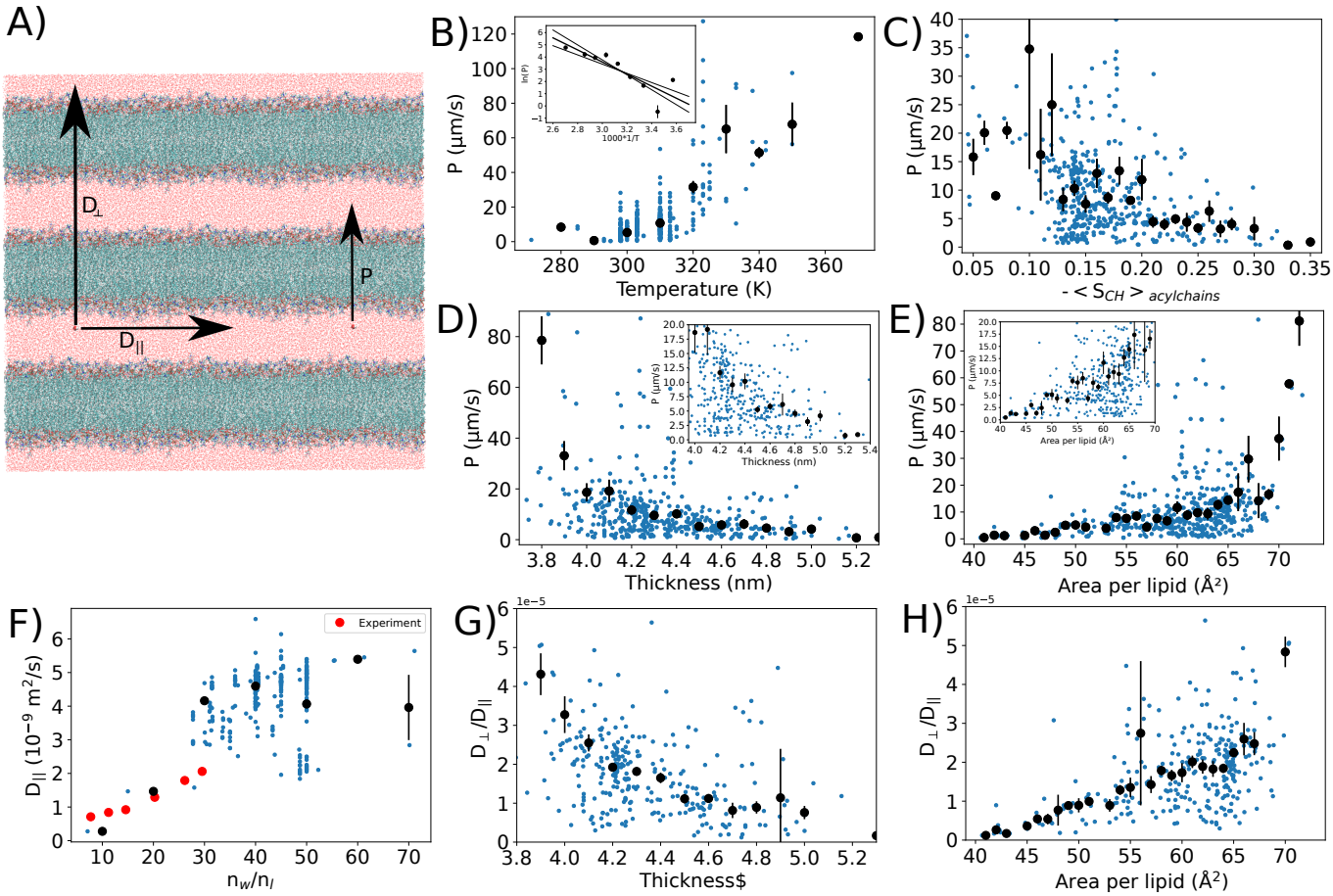


**Figure 4.** A **B-E** Water permeation through membranes analyzed from the databank as a function of temperature, thickness, area per lipid, and acyl chain order. Values from simulations with non-zero permeation values are shown with blue dots. Averages over fixed range of x-axis values are shown with black dots. Insert in B) shows the Arrhenius plot of permeation (ln (P) vs. 1/T) that gives $17 \pm 4\,k_bT$ 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 shown[42]. **G-H** Diffusion anistoropy of water as a function of thickness and area per lipid.

hydration level were not observed in Fig. S4 in the supplementary information.

To analyze the anisotropic diffusion of water in membrane environment, we also calculated the water diffusion parallel to the membrane surface from all simulations in the NMRlipids databank. The results at different hydration levels are shown in Fig. 4 F together with the experimental data[42]. Simulation results are close to the experimental values with low hydration levels, but increase to the values of approximately two times higher than experimental bulk water diffusion value ($3.1 \cdot 10^{-9}$ m$^2$/s at 313 K[43]) with high hydration levels. This is not surprising as the most common water model used in membrane simulations, TIP3P, overestimates the bulk water diffusion[44]. The water diffusion parallel to membrane surface increased with the temperature, but dependencies on membrane area per lipid, thickness, or fraction of charged lipids were not observed in Fig. S5 in the supplementary information. In order to estimate the diffusion anistoropy of water in multilamellar membrane system, we translated the water permeation coefficients through membranes to diffusion coefficients through multilamellar stack using the Tanner equation[45,46]. The resulting diffusion coefficients through the membranes are approximately five orders of magnitude slower than along the membrane (Figs. 4 G and H), which is at the upper limits of the anisotropic estimated from the experimental data[31]. Relatively high anisotropy is understandable as simulations give slightly slower permeation rates and higher lateral diffusion rates than experiments. Significant increase in the diffusion anisotropy with membrane packing is observed, as $D_\perp/D_{||}$ drifts away from one with decreasing area per lipid and increasing thickness in Figs. 4 G and H. This follows from decreasing water permeability with membrane packing (Figs 4 C and D), while lateral diffusion remains approximately constant (Fig. S5 A and C).

### 2.6 Flip-flops

Lipid flip-flops from one bilayer leaflet to another play an important role in lipid trafficking and regulating membrane properties[47]. Phospholipid flip-flops are slow, occuring 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[47–50]. Flip-flops have been previously studied mainly with coarse-grained simulations and free energy calculations due to time scale limitations in atomistic simulations[49], but direct analyses on cholesterol flip-flops in atomistic simulations have been also emerging in recent years[50–52]. However, systematic studies on how flip-flop rates depend on membrane properties are still challenging even with the state of the art computational resources. The NMRlipids databank can greatly advance such studies by giving an easy access to large amounts of data simultaneously.

## 3 Discussion

Applications of MD simulations in understanding biomolecular behaviour are steadily increasing, but protocols for quality evaluation and sharing simulation data fall behind the ones in more elaborated fields, such as structural biology. Development of such protocols has been limited by practical challenges, such as lack of universally defined quality measures for MD simulations, incompatible naming conventions in different models and simulation softwares, and high technical requirements for hardware to share large trajectories from MD simulations, as well as incentives for scientist to evaluate and share their MD simulation data.

The NMRlipids databank circumvents these challenges by utilizing the overlay databank design and open collaboration approach developed in the NMRlipids project[20]. In the overlay databank, the storage of large simulation files is outsourced to existing publicly available repositories while databank itself is essentially a git repository with open access licence containing relevant information about the trajectories, including the location of the raw data. This architecture streamlines the construction and maintenance of the databank by distributing the long term commitment for hardware and software maintenance. Incentive to share the data is created in the NMRlipids open collaboration by offering authorship in published articles to the contributors. Universal quality measures and naming conventions for molecules and atoms are defined in the NMRlipids databank to enable the automatic quality evaluation and other analyses. The current NMRlipids databank contains only lipid bilayer simulations, but the concept combining open collaboration and overlay databank structure can be applied also in other fields where similar barriers to establish publicly accessible databanks exist.

The NMRlipids databank will be of great benefit to scientists involved in MD simulations. They will be able to rapidly evaluate the quality of MD simulations in order to filter out potentially misleading results or facilitate force field parameter development[28]. The selection of best models for a specific application is demonstrated for PC/PE lipid mixtures in Fig. 2, and previously for PC/PS lipid mixtures[28], and lipid headgroups[23]. On the other hand, programmatic access in the NMRlipids databank enables automatic analyses over large sets of simulation data containing more data in terms of quantity (e.g., simulation length and number of conformations) with broader range in terms of content (e.g, lipid compositions and ion concentrations) that can be accessible for a single research group. Added value of the access to broad range of systems is exemplified in Figs. 2 and 4 revealing correlations between membrane properties, while large quantity of data enables the detection of rare phenomena with good statistics, as exemplified for water permeability in Fig. 4 and cholesterol flip-flop events in Fig. **??**.

Providing programmatic access to large scale MD simulation data can lead to applications in unprecedented directions. For example, water anisotropic diffusion analyzed in Fig. 4 is relevant not only in pharmacokinetics but also in the development of MRI imaging methods where MD simulations are not yet widely used[33]. We expect the access provided by the NMRlipids databank to facilitate novel applications also in other fields where MD simulations are less commonly applied, such as design of lipid nanoparticles for drug delivery or other applications, and material science. The increasing amount of data in terms of quantity and content will also increase the scope of potential applications of the NMRlipids databank in time.

The NMRlipids databank can also serve as a training set for diverse machine learning applications. Most straightforward applications include models that would predict connections between membrane properties that are already stored in the databank. 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 could be trained to predict arbitrary membrane properties using the data from the NMRlipids databank. Such models could be used to connect membrane composition to relevant membrane properties in wide range of fields from molecular biology to drug development and biomolecular imaging. The overlay databank design can open such avenues also for other biomolecules than lipids, such as disordered proteins and membrane-protein systems.

Different types of applications enabled by the NMRlipids databank in wide range of fields are exemplified in Table 1.

| 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 4) | Membrane scientists |
| Applications that are outside typical scope of MD simulations | Anisotropic water diffusion for pharmagokinetics 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[23], packing of PS[28] and PE (Fig. 3) containing membranes. | Scientists using MD simulations |
| Guidance for force field development | Improvements in ion binding to lipids[53,54] and lipid headgroup conformational ensembles[?] | 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.

# 4 Methods

## 4.1 Structure of the databank

Structure of the NMRlipids databank is illustrated more detailed in Fig. S6 in the supplementary information. The databank locates as a git repository at https://github.com/NMRLipids/Databank/ and is permanently stored in Zenodo repository (www.zenodo.org)[?]. 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[55,56].

The core content of the NMRlipids databank is stored in README.yaml files located at */Data/Simulations* in the NMRlipids databank repository. Essential information for each simulation entry in the databank is stored in these files. These yaml files are both human and machine readable, and contain access to all information that are needed for further analyses of the simulations. The keys available in these files and their explanations are listed in table S1 in the supplementary information. To keep the databank light, the raw MD simulation data is stored in an external location which can be any stable publicly available repository, although all the current data locates in Zenodo www.zenodo.org. Whenever needed, the raw data can be downloaded through the links stored in the README.yaml files.

## 4.2 Molecule and atom naming convention

Molecule and atom names used in a simulation are often needed when analyzing the simulation trajectories. However, the naming conventions of lipid molecules and atoms therein vary between authors and force fields because universal naming convention has not been defined. Because such convention is necessary for automatic analyses over simulations, we have defined the unique naming conventions for lipid molecules and atoms in the NMRlipids databank. For each molecule we have defined an unique abbreviation which are listed in table S2 in the supplementary information. On the other hand, unique names for each atom in a molecule are defined and connected to corresponding names in a simulation using mapping files. Mapping files are in yaml format where unique atom names are connected to force field specific names. Also residue names and fractions lipids (headgroup, glycerol bakcbone, and acyl chains) can be defined in the mapping file. For example, the mapping files used in the NMRlipids databank are located at */Scripts/BuildDatabank/mapping_files* in the NMRlipids repository. For more details about mapping files, see https://nmrlipids.blogspot.com/2022/04/new-yaml-format-of-mapping-files.html). In practise, information about molecule and atom names are given in the COMPOSITION dictionary in README.yaml files, where force field specific molecule name corresponding the unique abbreviation (table S2) and the name of the mapping file are given.

## 4.3 Adding data into the NMRlipids databank

In order to add a simulation entry into the NMRlipids databank, an initial info.yaml file containing user given information listed in table S1 needs to be created. For examples of these files, see */Scripts/BuildDatabank/info_files* in the NMRlipids databank repository. Using this file as an input, the */Scripts/BuildDatabank/AddData.py* script then calculates the rest of the information to be stored in the README.yaml files, also listed in table S1. The databank is open for submission of these info.yaml files via pull requests to the git repository. Large fraction of the current content is composed of the data uploaded to Zenodo in previous NMRlipids projects[20–23], yet also other simulation data from the Zenodo repository are added.

## 4.4 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. **?** and x-ray scattering data from Refs. **?**. 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.

## 4.5 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

$$S_{\mathrm{CH}} = \frac{1}{2}\left\langle 3\cos^2\theta - 1\right\rangle, \tag{1}$$

where $\theta$ is the angle between the C–H bond and the membrane normal and angular brackets denote the ensemble average, i.e., average over all sampled configurations of all lipids in a simulation. 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 bar[20].

In practise, the analysis code loops over all README.yaml files (i.e., simulations in the NMRlipids databank), downloads the raw simulation to a local computer, uses the information about the atom and molecule naming conventions in README.yaml to determine the location of carbon and hydrogen atoms, and then calculates the order parameters from Eq. 1. The script to calculate all C-H bond order parameters from 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.

### 4.6 Calculation of x-ray scattering form factors

X-ray scattering form factors were calculated with standard equation for symmetric lipid bilayers[24]

$$F(q) = \int_{-D/2}^{D/2} \Delta\rho_e(z) \cos(z q_z) \mathrm{d}z, \tag{2}$$

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 centred around the centre of mass of the POPC lipid molecules for every time frame, and a histogram of these centred positions was calculated with the bin width of $1/3$ Å. Similarly to order parameters, the script loops over all simulations, downloads the data on local computer, and uses the information in README.yaml file to calculate form factors. 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.

### 4.7 Calculation of membrane thickness, and area per lipids

Lipid bilayer thicknesses in all simulations are calculated from the intersections of lipid and water electron densities using the script *Scripts/AnalyzeDatabank/calc_thickness.py* at the NMRlipids databank repository. The resulting thicknesses are stored for all simulations in files named *thickness.json* at folders in */Data/Simulations* in the NMRlipids databank repository. Lipid bilayer area per lipids in all simulations are calculated by dividing the simulation box area with the total number of molecules considered as lipids using the script *Scripts/AnalyzeDatabank/calcAPL.py* at the NMRlipids databank repository. The resulting area per lipids are stored for all simulations in files named *apl.json* at folders in */Data/Simulations* in the NMRlipids databank repository.

### 4.8 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 $\pm 3$ percentage units, charged lipids have the same counterions, and temperatures are within $\pm 2$ 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, $\frac{S_{\mathrm{CH}}-\mu}{s/\sqrt{n}}$ 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

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

where f(t) is the Student's t-distribution, $\mu$ 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, $S_{\mathrm{CH}}$ is the sample mean from Eq. 1, $s$ is the variance of $S_{\mathrm{CH}}$ calculated over individual lipids, $S_{\mathrm{exp}}$ is the experimental value, and $\Delta S_{\mathrm{exp}}$ its error. The error of $\Delta S_{\mathrm{exp}} = 0.02$ is currently assumed for all experimental order parameters[24]. 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 values[57]. 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

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

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

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

where $\chi_{\text{lipid}}$ 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.jso* 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.

### 4.9 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 program[58]

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

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 program[58]. However, the lobe heigths 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 $0.1\,AA^2$ from both simulation and experimental data. The quality of a form factor is then defined as Eucledian distance between the minima in simulated and experimental form factors, $FF_q = |FF^{\text{sim}}_{\text{min}} - FF^{\text{exp}}_{\text{min}}|$.

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.

### 4.10 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 $c_w$ is the concentration of water in bulk with the value of $33.3679\,(\text{nm})^{-3}$[38]. The number of permeation events in each trajectory was calculated using the code by Camilo et al.[40], available at https://github.com/crobertocamilo/MD-permeation. This code was used within the loop that goes over all simulations in the NMRlipids databank and extracts the required information using the information in README.yaml files. The code is available at */scripts/calcMD-PERMEATION.py* and 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 also of the analyzed data without overloading the main NMRlipids databank repository.

The parallel diffusion along membrane surface was calculated using Einstein's equation with *gmx msd* program with the *-lateral* option from Gromacs softwater package[59]. The code is available at */scripts/calcWATERdiffusion.py* and 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 $D_{\perp} = \frac{D_{||} P z_w}{D_{||} + P z_w}$ [45,46], where the water layer thickness, $z_w$, was estimated by subtracting bilayer thickness from the size of the simulation box in membrane normal direction.

### 4.11 Calculation of lipid flip-flops

## References


1. 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 (2015).

2. 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).

3. 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).

4. 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 (2016).

5. 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 (1999).

6. Tai, K. *et al.* Biosimgrid: towards a worldwide repository for biomolecular simulations. *Org. Biomol. Chem.* **2**, 3219–3221, DOI: 10.1039/B411352G (2004).

7. 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 (2006).

8. Abraham, M. *et al.* Sharing data from molecular simulations. *J. Chem. Inf. Model.* **59**, 4093–4099 (2019).

9. 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 (2019).

10. 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.

11. 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 (2020). https://academic.oup.com/bioinformatics/article-pdf/36/9/2952/33180880/btaa086.pdf.

12. 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 (2020).

13. 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).

14. 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).

15. 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 (2016). https://academic.oup.com/nar/article-pdf/44/D1/D272/16661850/gkv1301.pdf.

16. 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 (2016). https://academic.oup.com/bioinformatics/article-pdf/32/21/3371/7889578/btw289.pdf.

17. 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 (2018). https://academic.oup.com/nar/article-pdf/47/D1/D390/27437085/gky1047.pdf.

18. 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 (2020).

**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.** 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).

**21.** Catte, A. *et al.* Molecular electrometer and binding of cations to phospholipid bilayers. *Phys. Chem. Chem. Phys.* **18**, 32560–32569 (2016).

**22.** Antila, H. *et al.* Headgroup structure and cation binding in phosphatidylserine lipid bilayers. *J. Phys. Chem. B* **123**, 9066–9079 (2019).

**23.** Bacle, A. *et al.* Inverse conformational selection in lipid–protein binding. *J. Am. Chem. Soc.* **143**, 13701–13709 (2021).

**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.** Pluhackova, K. *et al.* A critical comparison of biomembrane force fields: Structure and dynamics of model dmpc, popc, and pope bilayers. *The J. Phys. Chem. B* **120**, 3888–3903 (2016).

**26.** Sandoval-Perez, A., Pluhackova, K. & Böckmann, R. A. Critical comparison of biomembrane force fields: Protein–lipid interactions at the membrane interface. *J. Chem. Theory Comput.* **13**, 2310–2321 (2017).

**27.** Leonard, A. N., Wang, E., Monje-Galvan, V. & Klauda, J. B. Developing and testing of lipid force fields with applications to modeling cellular membranes. *Chem. Rev.* **119**, 6227–6269 (2019).

**28.** 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).

**29.** 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.

**30.** 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).

**31.** 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).

**32.** 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).

**33.** 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).

**34.** Mathai, J. C., Tristram-Nagle, S., Nagle, J. F. & Zeidel, M. L. Structural Determinants of Water Permeability through the Lipid Membrane . *J. Gen. Physiol.* **131**, 69–76 (2007).

**35.** Nitsche, J. M. & Kasting, G. B. Permeability of fluid-phase phospholipid bilayers: Assessment and useful correlations for permeability screening and other applications. *J. Pharm. Sci.* **102**, 2005–2032, DOI: https://doi.org/10.1002/jps.23471 (2013).

**36.** Nitsche, J. M. & Kasting, G. B. A universal correlation predicts permeability coefficients of fluid- and gel-phase phospholipid and phospholipid-cholesterol bilayers for arbitrary solutes. *J. Pharm. Sci.* **105**, 1762–1771, DOI: https://doi.org/10.1016/j.xphs.2016.02.012 (2016).

**37.** Shinoda, W. Permeability across lipid membranes. *Biochimica et Biophys. Acta (BBA) - Biomembr.* **1858**, 2254–2265, DOI: https://doi.org/10.1016/j.bbamem.2016.03.032 (2016). Biosimulations of lipid membranes coupled to experiments.

**38.** Venable, R. M., Krämer, A. & Pastor, R. W. Molecular dynamics simulations of membrane permeability. *Chem. Rev.* **119**, 5954–5997 (2019).

**39.** 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).

**40.** 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).

**41.** 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 (1995).

**42.** Rudakova, M., Filippov, A. & Skirda, V. Water diffusivity in model biological membranes. *Appl. Magn. Reson.* **27**, 519 (2004).

**43.** Khakimov, A. M., Rudakova, M. A., Doroginitskii, 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 (2008).

**44.** 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).

**45.** 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 (1978).

**46.** 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).

**47.** 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 (2008).

**48.** 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).

**49.** 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).

**50.** Gu, R.-X., Baoukina, S. & Tieleman, D. P. Cholesterol flip-flop in heterogeneous membranes. *J. Chem. Theory Comput.* **15**, 2064–2070 (2019).

**51.** Javanainen, M., Lamberg, A., Cwiklik, L., Vattulainen, I. & Ollila, O. H. S. Atomistic model for nearly quantitative simulations of langmuir monolayers. *Langmuir* **34**, 2565–2572 (2018).

**52.** 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).

**53.** 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).

**54.** 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).

**55.** Gowers, R. J. *et al.* Mdanalysis: a python package for the rapid analysis of molecular dynamics simulations (2019).

**56.** 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).

**57.** 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).

**58.** Kučerka, N., Katsaras, J. & Nagle, J. Comparing membrane simulations to scattering experiments: Introducing the SIMtoEXP software. *J. Membr. Biol.* **235**, 43–50 (2010).

**59.** Bauer, P., Hess, B. & Lindahl, E. Gromacs 2022.3 manual, DOI: 10.5281/zenodo.7037337 (2022).


## 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