SUPPLEMENTARY INFORMATION: 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

S1 Correlations between area per lipid and thickness with order parameters and form factors

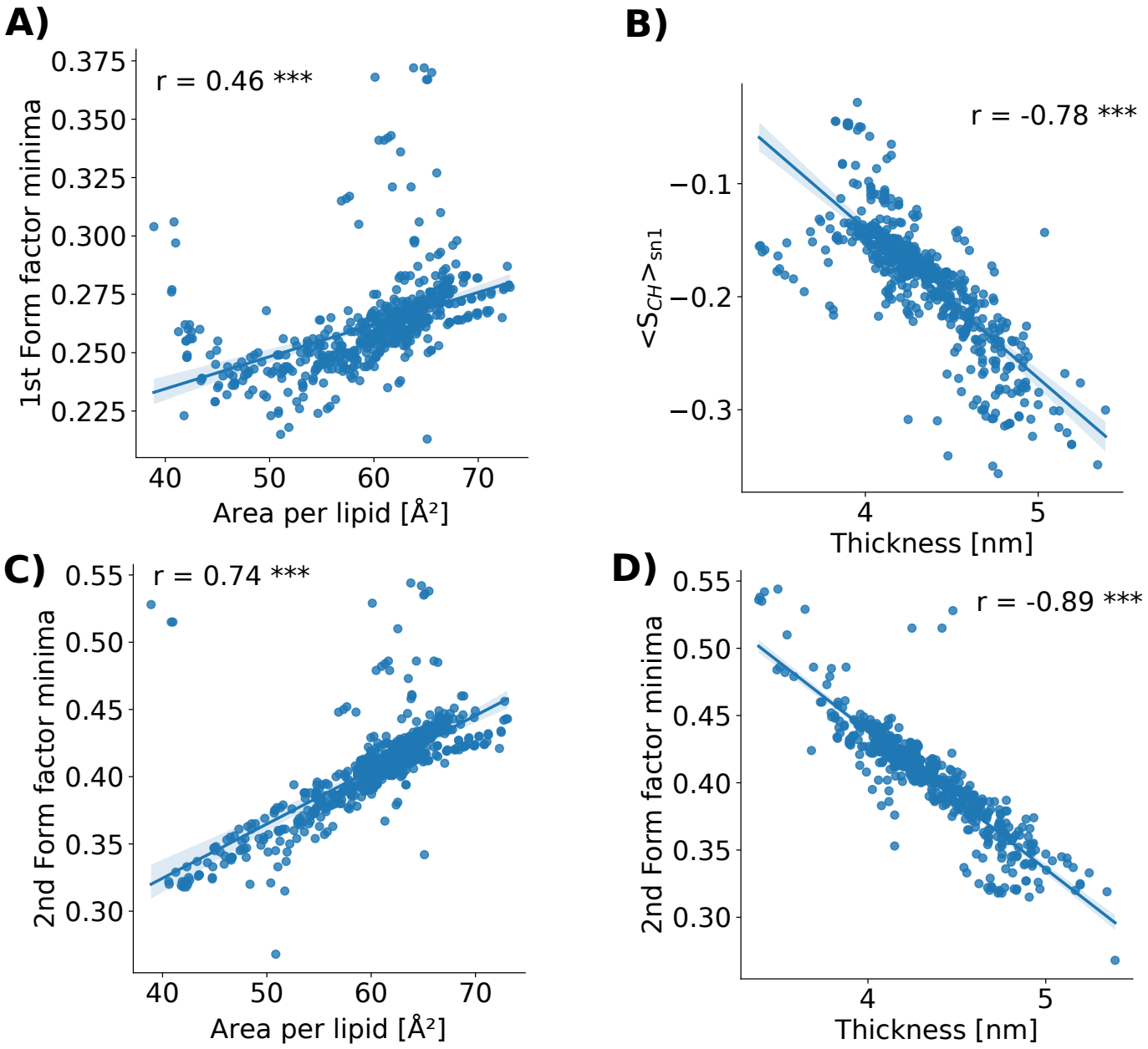

Figure S1. Dependence of form factor, electron density profiles along membrane normal, and C-H bond orderparameters (from top to bottom) on simulation box size with different cholesterol concentration.

S2 Dependence of form factor and order parameters in the size of simulation

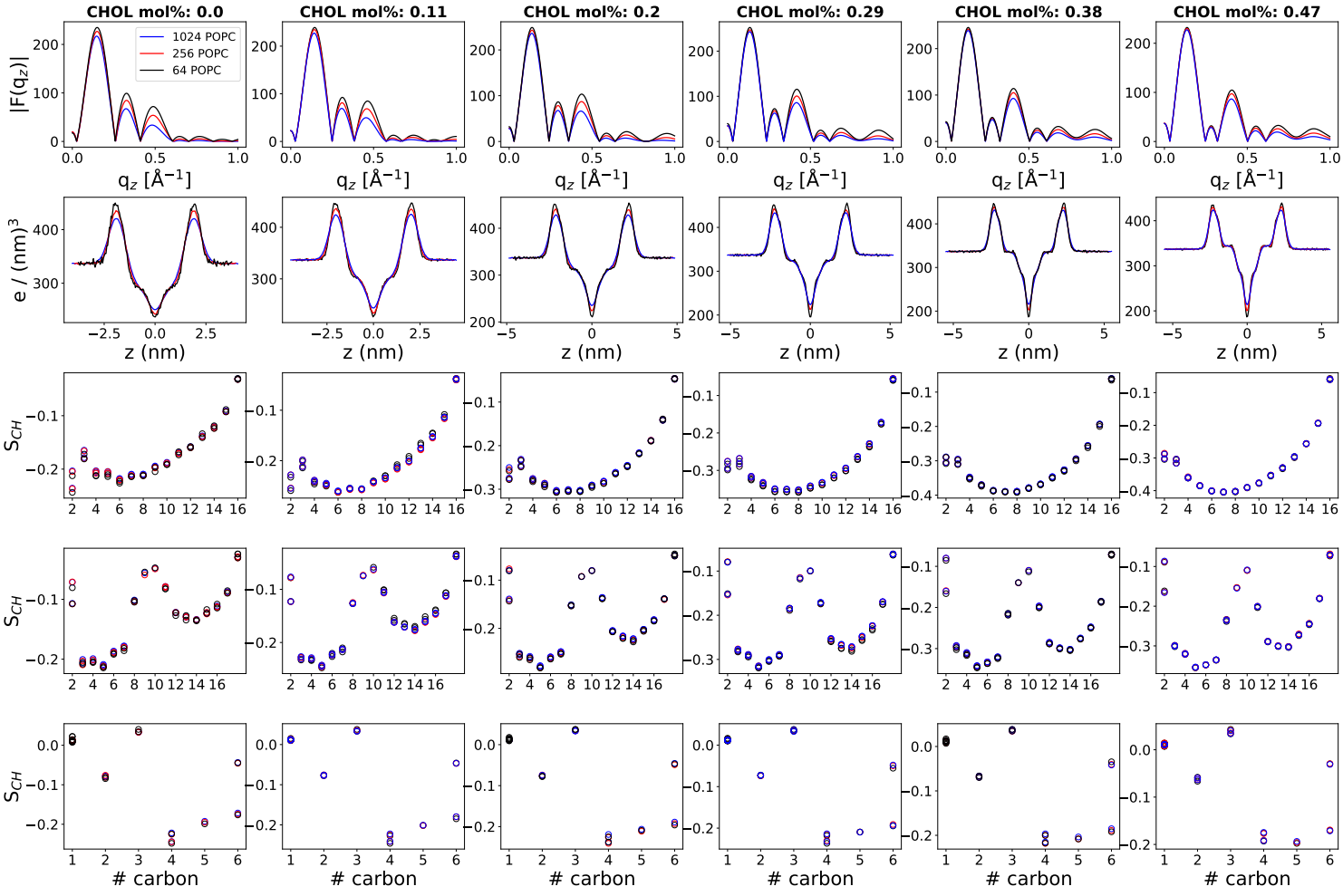

Figure S2. Dependence of form factor, electron density profiles along membrane normal, and C-H bond orderparameters (from top to bottom) on simulation box size with different cholesterol concentration. Simulations with 64, 256 or 1024 POPC lipids taken from Ref. 1.

S3 Finding the best models for PC and PE mixtures

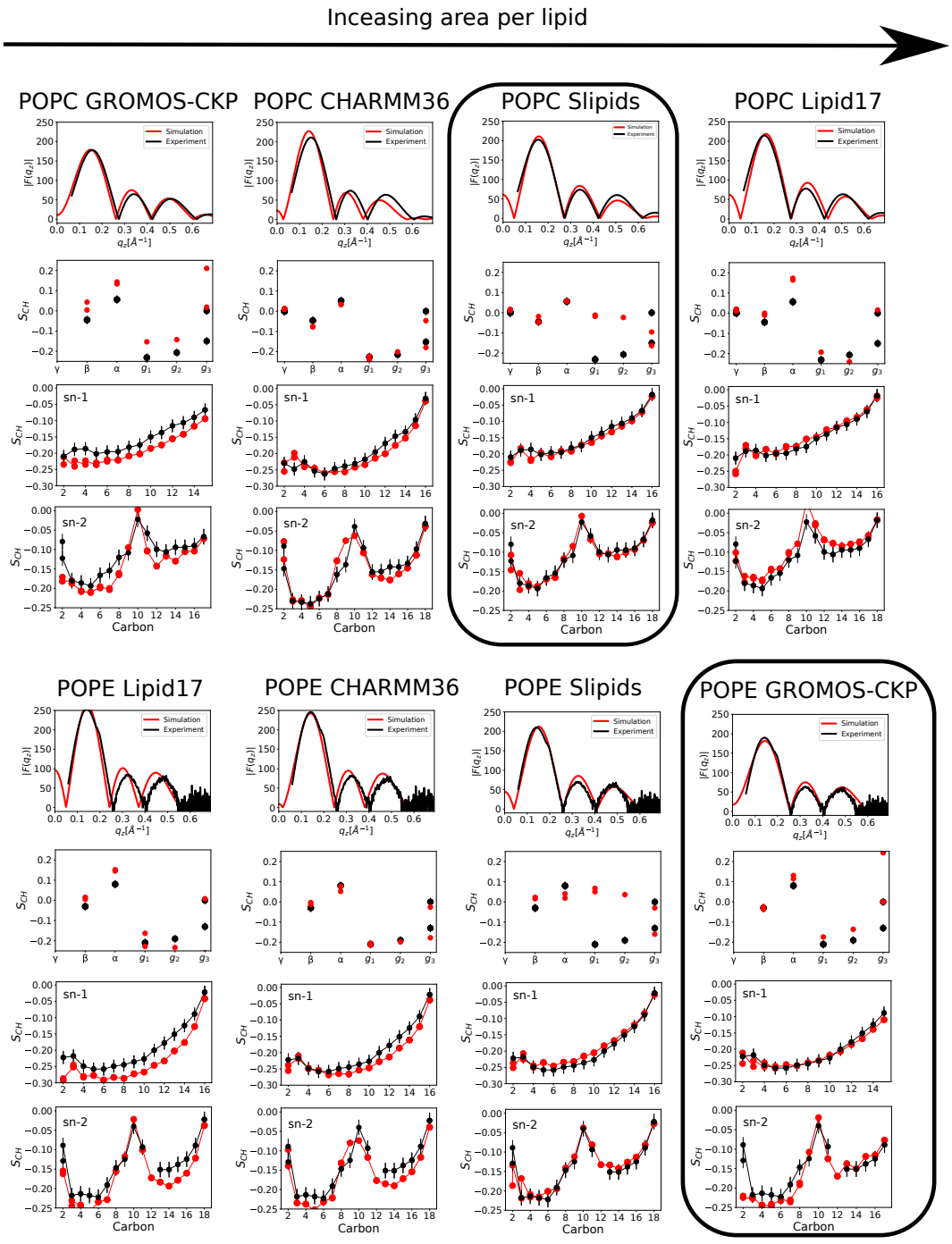

Figure S3. Simulations with the data for both POPC (top) and POPE (bottom) directly compared with the experimental data. The area per lipid in simulations increase from left to right. Simulations with the best overall quality for POPC and POPE order parameters are circled by a square.

S4 Water permeation through membranes

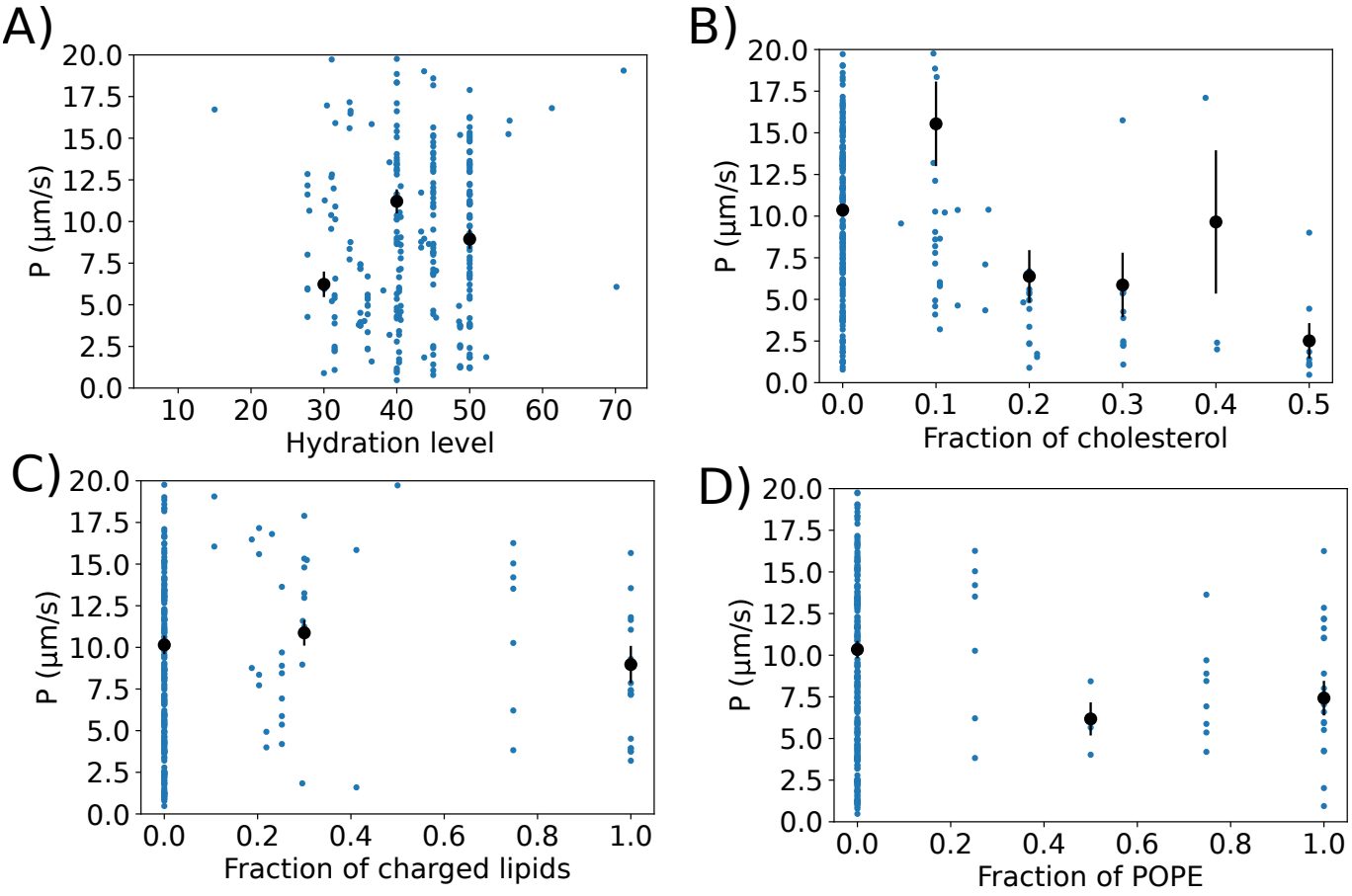

Figure S4. a) Structure of an overlay databank. More detailed structure of the layer 2 in the NMRlipids databank is illustrated in Fig. S6 in the SI. b) Distribution of the lengths of the trajectories, total number of trajectories and total lenght of the simulations in the NMRlipids databank. c) Distribution 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.

S5 Water diffusion along membranes

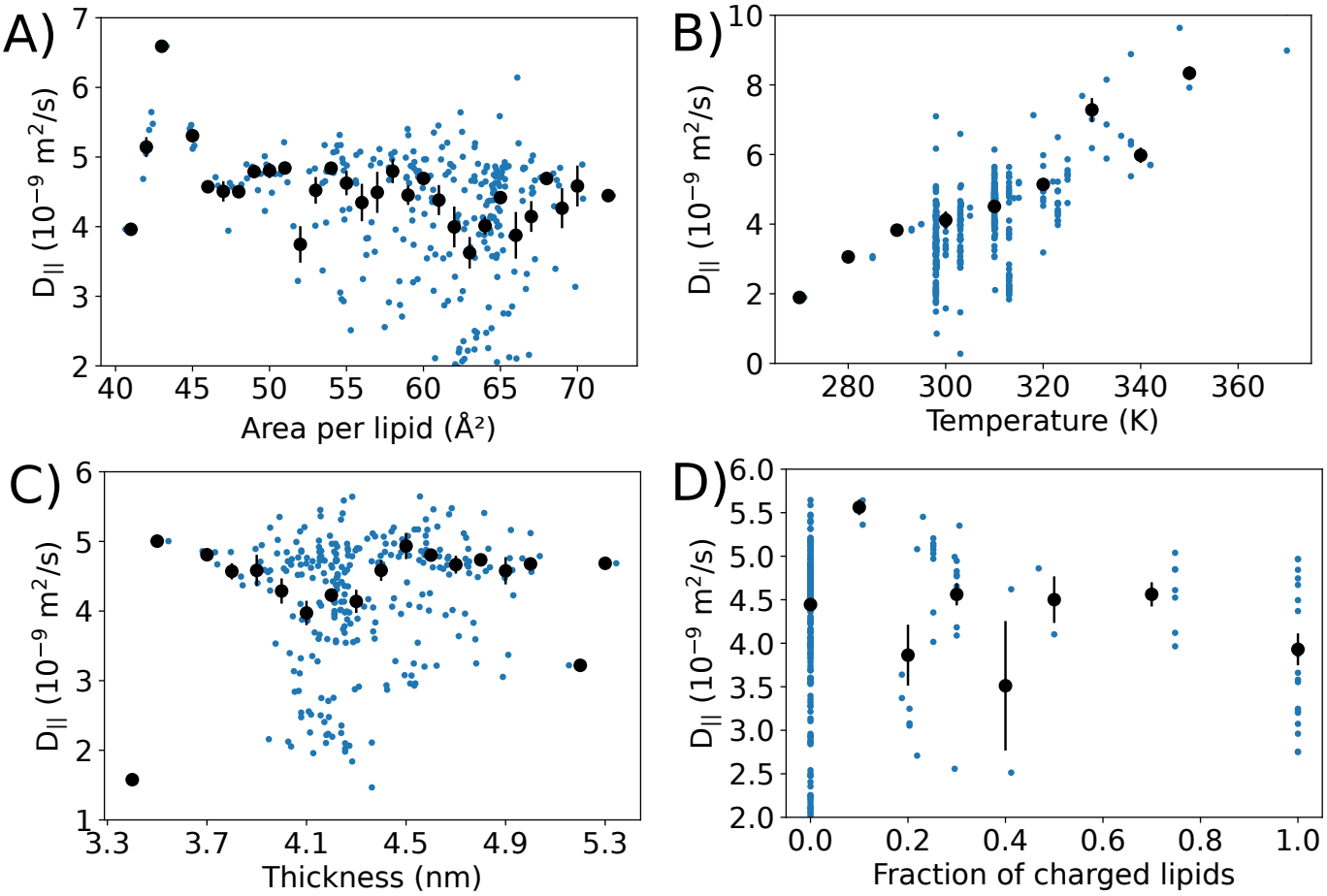

Figure S5. a) Structure of an overlay databank. More detailed structure of the layer 2 in the NMRlipids databank is illustrated in Fig. S6 in the SI. b) Distribution of the lengths of the trajectories, total number of trajectories and total lenght of the simulations in the NMRlipids databank. c) Distribution 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.

S6 Databank content

References

1. Javanainen, M. Simulations of POPC/cholesterol mixtures at 298 K, three system sizes, CHARMM36, DOI: 10.5281/zenodo.7035350 (2021).

[1]For example, if you upload 100-200 ns part of total 200 ns simulation, this should value should be 100.

[2]For example, if you upload 0-200 ns part of total 200 ns simulation where the first 100 ns should be considered as an equilibration, this value should be 100.

key	description	type
DOI	DOI from where the raw data is found	user given (compulsory)
SOFTWARE	Software used to run the simulation (e.g. Gromacs, Amber, NAMD, etc.)	
TRJ	Name of the trajectory file found from DOI	
TPR	Name of the topology file found from DOI (trp file in the case of Gromacs)	
PREEQTIME	Pre-equilibrate time simulated before the uploaded trajectory in nanoseconds. [1]	
TIMELEFTOUT	Equilibration period in the uploaded trajectory that should be discarded in analyses. [2]	
COMPOSITION	Molecules names used in the simulation and corresponding mapping files (see section 4.2)	
DIR_WRK	Temporary working directory in your local computer.	
UNITEDATOM_DICT	Information for constucting hydrogens for united atom simulations, empty for all atom simulations	
TYPEOFSYSTEM	Lipid bilayer or something else	
PUBLICATION	Give reference to a publication(s) related to the data.	User given (optional)
AUTHORS_CONTACT	Name and email of the main author(s) of the data.	
SYSTEM	System description on free text format	
SOFTWARE_VERSION	Version of the used software	
FF	Name of the used force field	
FF_SOURCE	Source of the force field parameters, e.g, CHARMM-GUI, webpage, citation to a publication, etc.	
FF_DATE	Date when force field parameters were accessed on the gives source (day/month/year).	
FFmolename	Molecule specific force field information, e.g., water model with FFSOL and sodium parameters with FFSOD.	
CPT	Name of the Gromacs checkpoint file.	
LOG	Name of the Gromacs log file.	
TOP	Name of the Gromacs top file.	
GRO	Name of the Gromacs gro file.	
TRAJECTORY_SIZE	Size of the trajectory file in bytes	automatically extracted data.
TRJLENGTH	Lenght of the trajectory (ps).	
TEMPERATURE	Temperature of the simulation.	
NUMBER_OF_ATOMS	Number of atoms in the simulation.	
DATEOFRUNNIG	Date when added into the databank	
EXPERIMENT	Potentially connected experimental data	
COMPOSITION	Numbers of lipid molecules (NPOPC, NPOPG, etc.) per membrane leaflet are calculated by determining on which side of the center of mass of the membrane the center of mass of the head group of each lipid molecule is located. Numbers of other molecules such as solvent and ions (NSOL, NPOT, NSOD, etc.) are read from the topology file.	

Table S1. Keys stored in the README.yaml files of simulations.

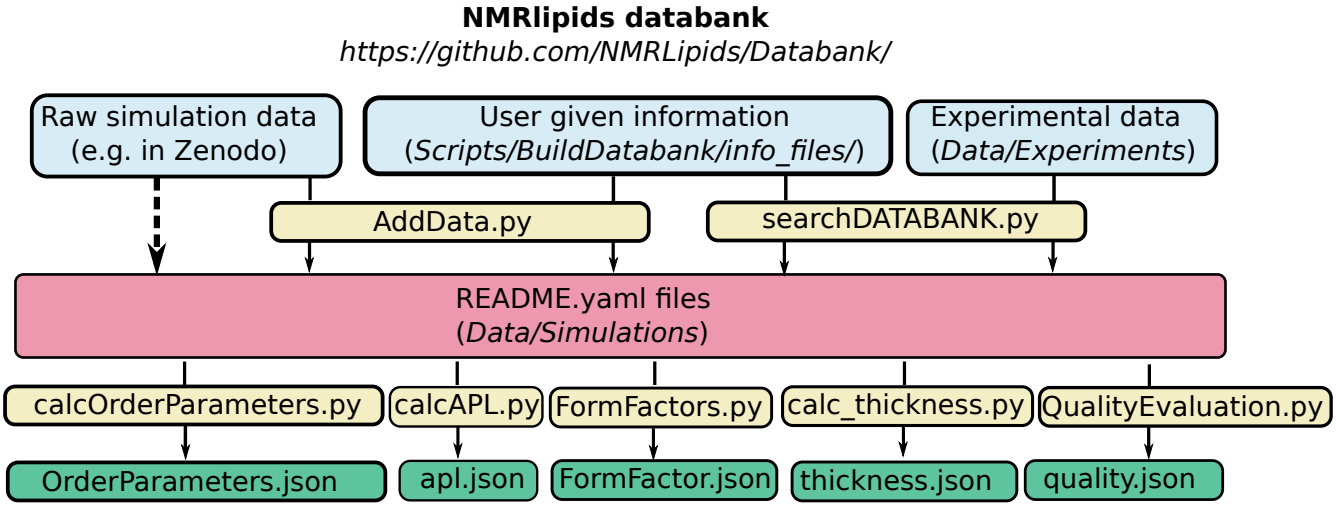

Figure S6. Structure of the NMRlipids databank. Manually added input data (blue boxes) includes basic information on the simulation, permanent links to the raw data, and experimental data if available. The databank entries (red box) and analysis results (green boxes), locating at https://github.com/NMRLipids/Databank/tree/main/Data/Simulations are automatically generated by the computer programs included in the NMRlipids databank (yellow boxes). Because raw data are not permanently stored but can be accessed based on the information in the databank, this connection is marked with the dashed line.

Abbreviation	Molecule name
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol
POPS	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
CHOL	cholesterol
DHMDMAB	dihexadecyldimethylammonium
POT	potassium ion
SOD	sodium ion
CLA	chloride ion
CAL	calcium ion
SOL	water

Table S2. Abbreviations used in the databank

key	description
DOI	DOI of the publication related to the experimental data.
TEMPERATURE	Temperature of the experiment.
MOLAR_FRACTIONS	Dictionary of molar fractions of bilayer components
ION_CONCENTRATIONS	Dictionary of ion concentrations of the system
TOTAL_LIPID_CONCENTRATION	Total concentration of lipid components. If exact concentration is not known, but experiments are performed in excess water, 'full hydration' can be given.
COUNTER_IONS	Type of counter ions if present.

Table S3. Keys stored in the README.yaml files of experiments.