# 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**

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

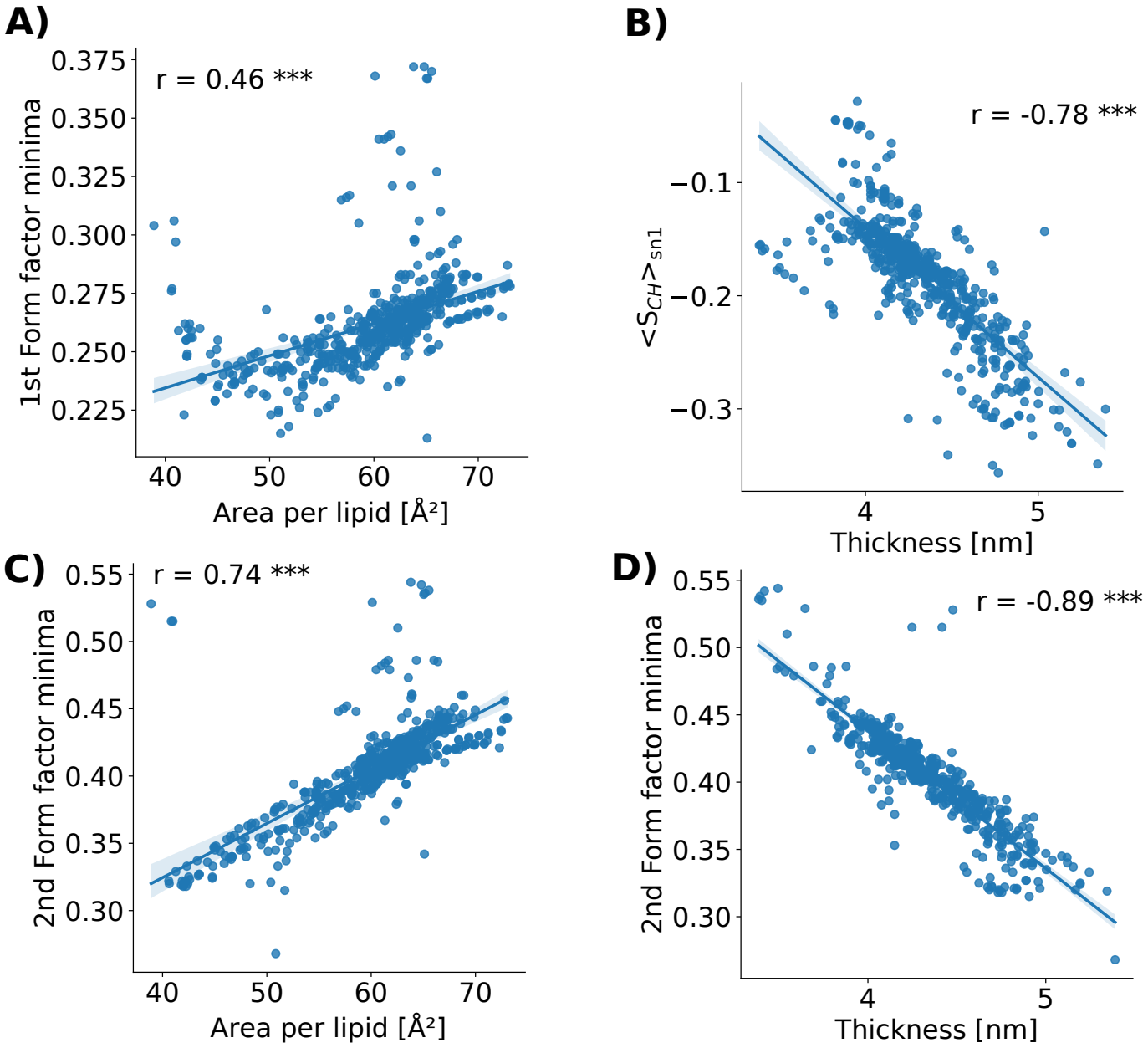


**Figure S1.** Dependence of form factor, electron density profiles along membrane normal, and C-H bond orderparameters (from top to bottom) on simulation box size with different cholesterol concentration.

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

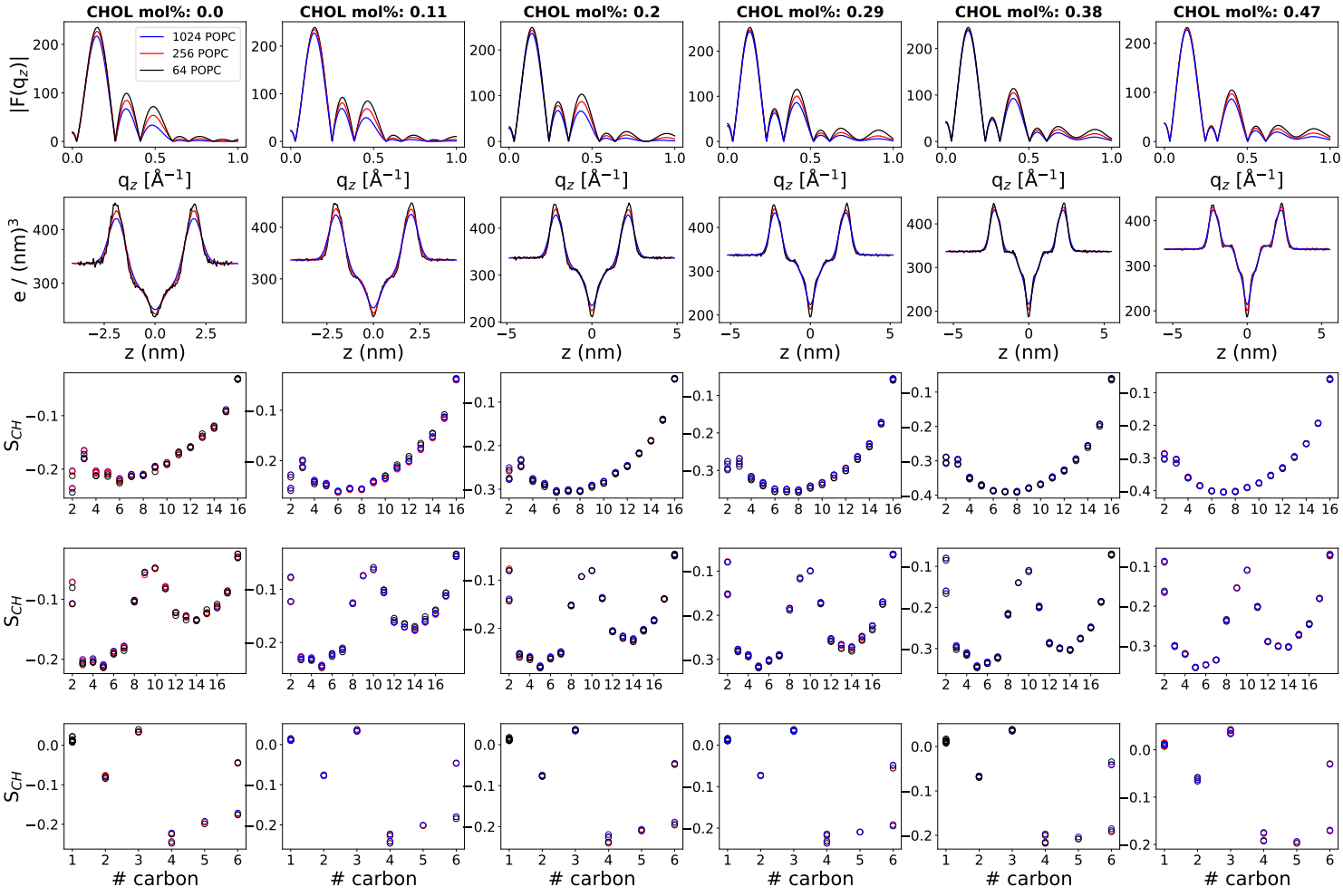


**Figure S2.** Dependence of form factor, electron density profiles along membrane normal, and C-H bond orderparameters (from top to bottom) on simulation box size with different cholesterol concentration. Simulations with 64, 256 or 1024 POPC lipids taken from Ref. **?**.

## Finding the best models for PC and PE mixtures

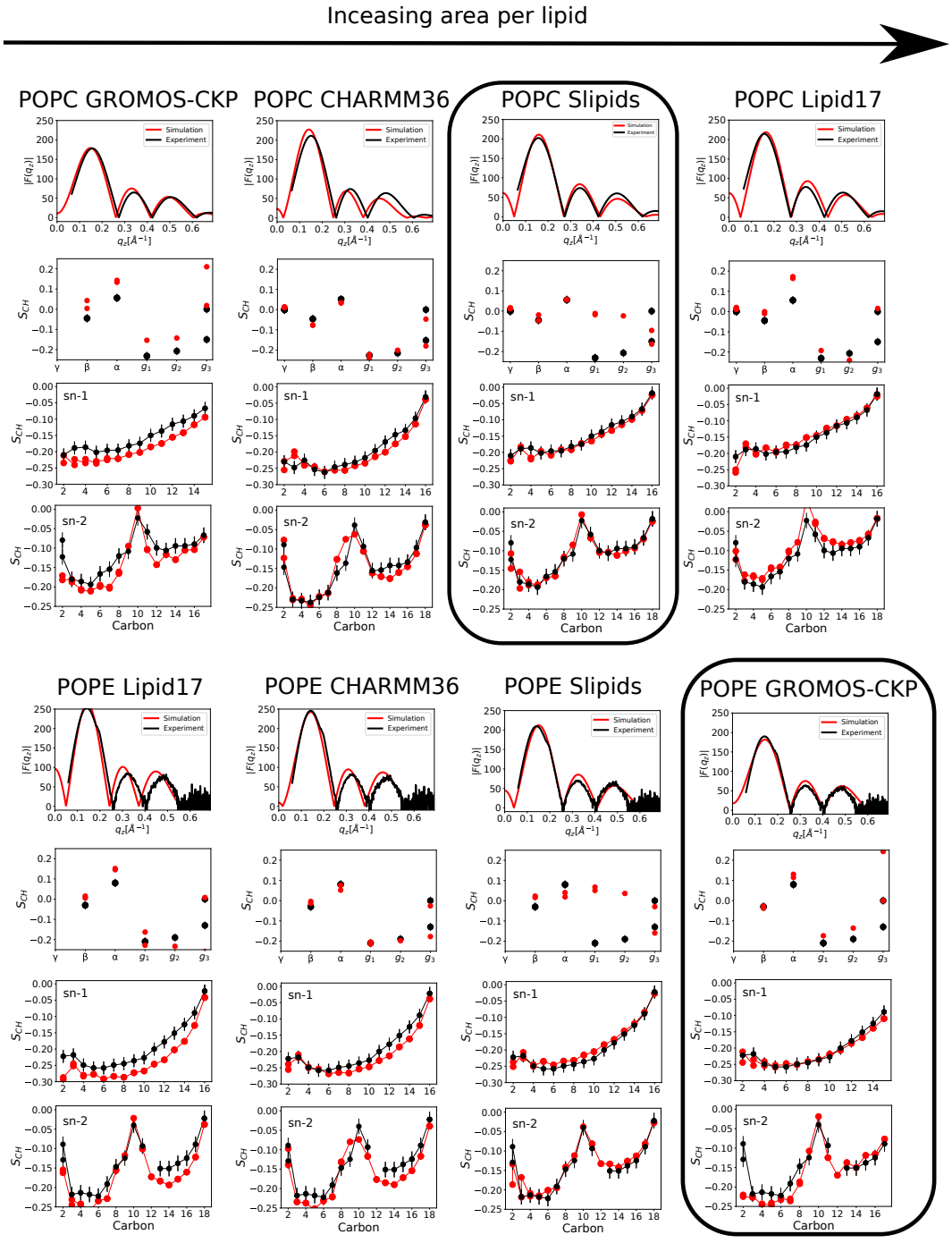


**Figure S3.** Simulations with the data for both POPC (top) and POPE (bottom) directly compared with the experimental data. The area per lipid in simulations increase from left to right. Simulations with the best overall quality for POPC and POPE order parameters are circled by a square.

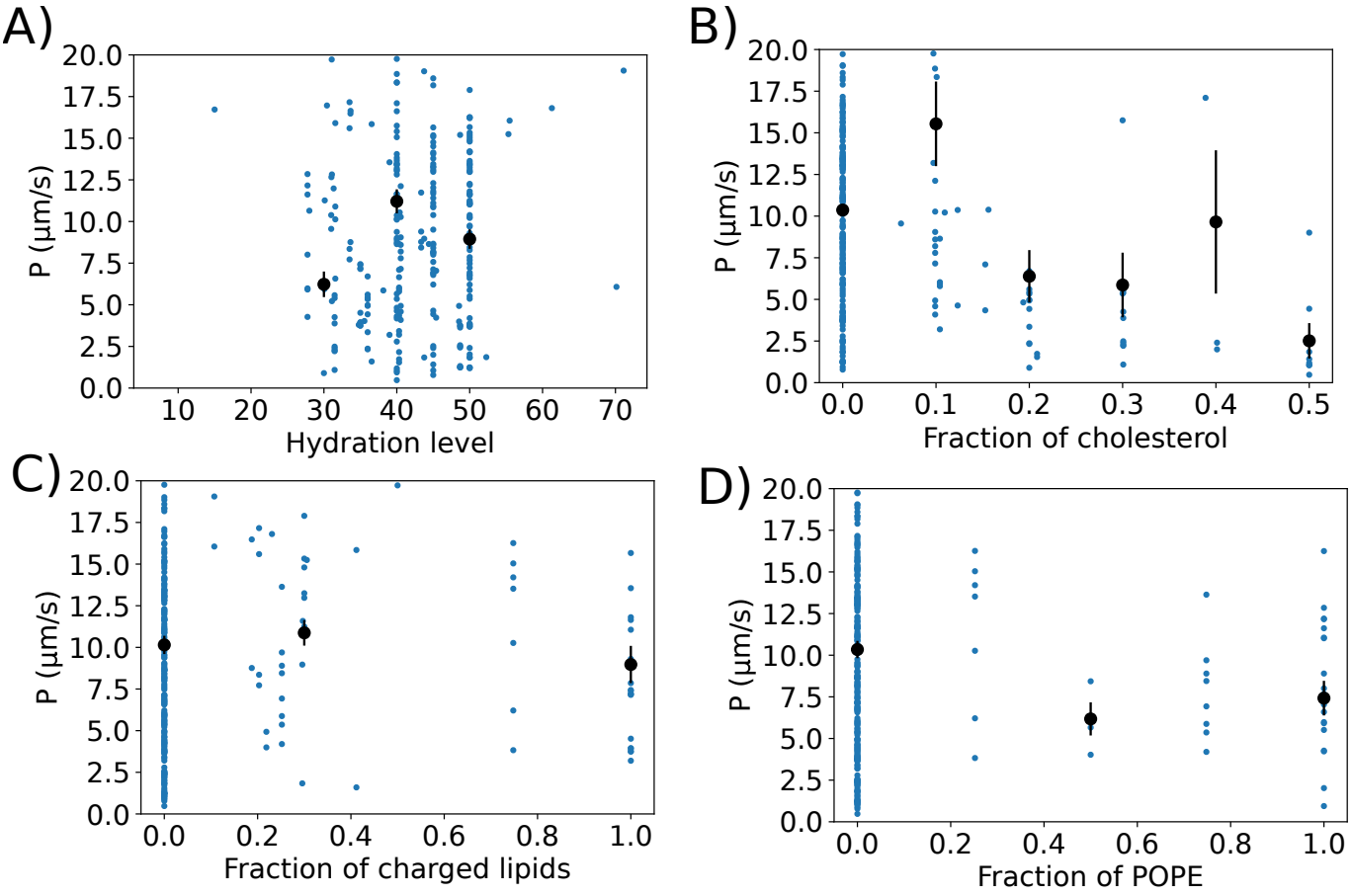


**Figure S4.** a) Structure of an overlay databank. More detailed structure of the layer 2 in the NMRlipids databank is illustrated in Fig. S6 in the SI. b) Distribution of the lengths of the trajectories, total number of trajectories and total lenght of the simulations in the NMRlipids databank. c)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

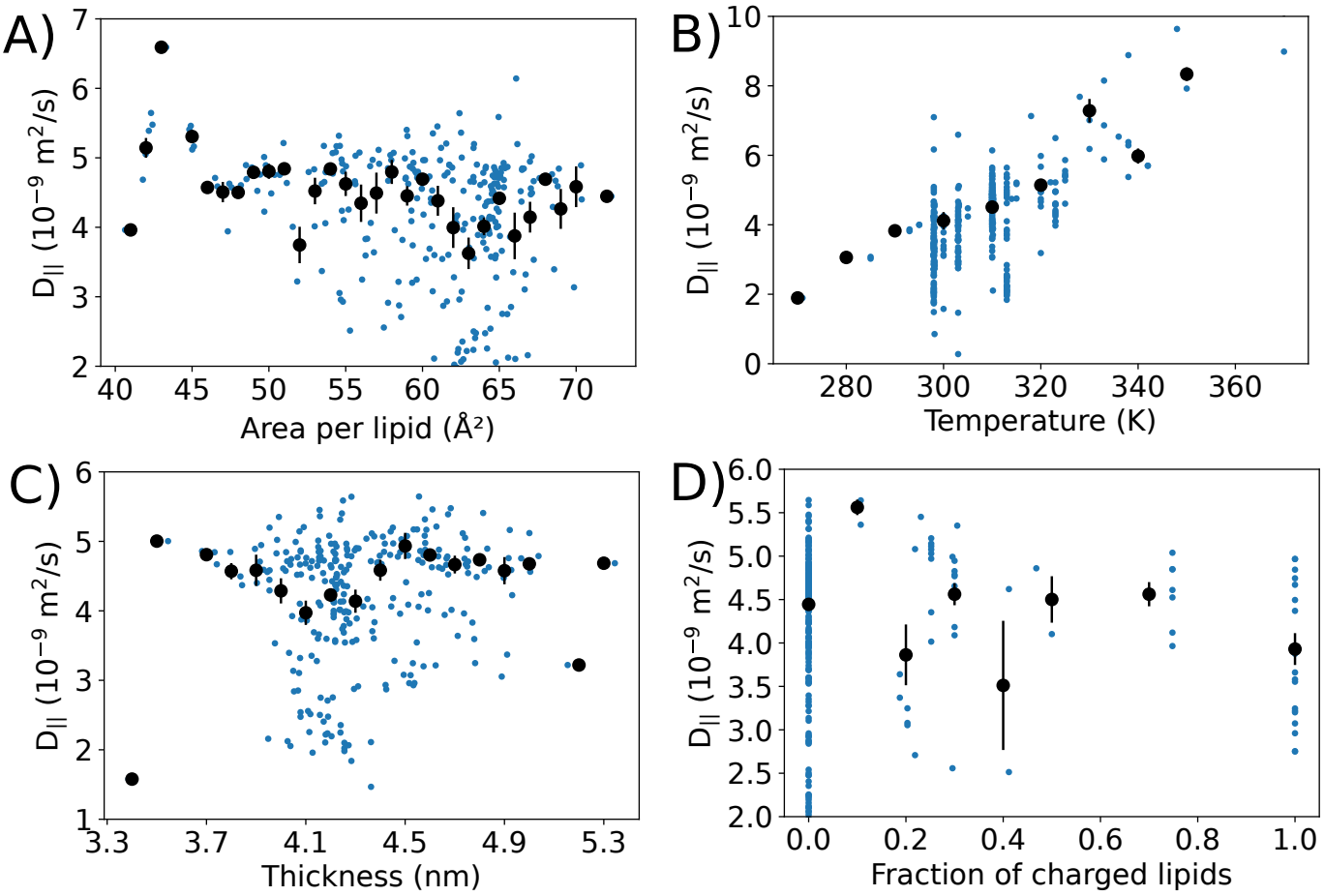


**Figure S5.** a) Structure of an overlay databank. More detailed structure of the layer 2 in the NMRlipids databank is illustrated in Fig. S6 in the SI. b) Distribution of the lengths of the trajectories, total number of trajectories and total lenght of the simulations in the NMRlipids databank. c)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

## Databank content

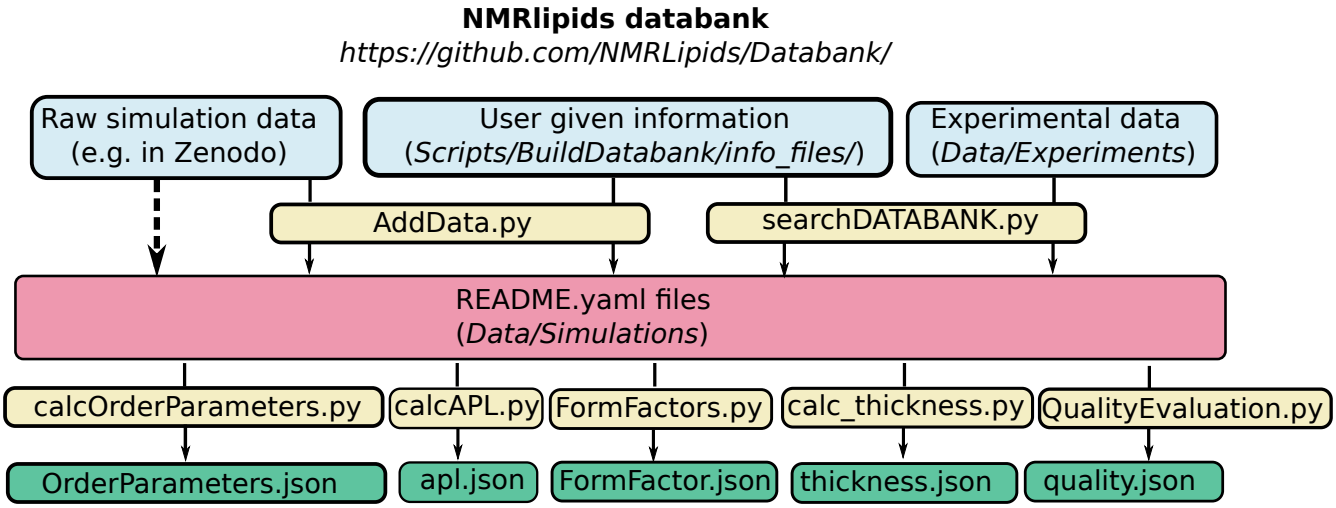


**Figure S6.** Structure of the NMRlipids databank. Manually added input data (blue boxes) includes basic information on the simulation, permanent links to the raw data, and experimental data if available. The databank entries (red box) and analysis results (green boxes), locating at https://github.com/NMRLipids/Databank/tree/main/Data/Simulations 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.

| 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 (trr or xtc for Gromacs, dcd for OpenMM) | |
| TPR (Gromacs) | Name of the tpr topology file found from DOI for Gromacs simulations | |
| PDB (OpenMM) | Name of the pdb file found from DOI for OpenMM simulations | |
| PREEQTIME | Pre-equilibrate time simulated before the uploaded trajectory in nanoseconds. | |
| TIMELEFTOUT | Equilibration period in the uploaded trajectory that should be discarded in analyses. | |
| COMPOSITION | Dictionary connecting universal molecule and atom names to the ones used in simulation | |
| DIR_WRK | Temporary working directory in your local computer. | |
| UNITEDATOM_DICT | Information for constucting hydrogens for united atom simulations using buildH program[?]. 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. | |
| GRO | Name of the Gromacs gro file. | |
| TOP | Name of top file for Gromacs or psf file for OpenMM. | |
| CRD | Name of crd file for OpenMM. | |
| WARNINGS | Dictionary containing information about unusual features in the trajectory, such as ambiguous atom names, membrane normal not oriented in z-direction, old Gromacs version used, etc. | |
| 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 in both leaflets and numbers of other molecules are added to the dictionary. | |
| ID | Unique ID number to ease the analyses. | |

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

## Experimental data

### NMR experiments

Acyl chain order parameters of POPE and POPG were analyzed from the experiments that were used to determine headgroup order parameters in Ref. **?**.

### X-ray scattering experiments

X-ray scattering form factors contributed to the NMRlipids III project are included in the databank (http://nmrlipids.blogspot.com/2022/09/nmrlipids-iii-including-lipid-lateral.html).

## References


1. Javanainen, M. Simulations of POPC/cholesterol mixtures at 298 K, three system sizes, CHARMM36, DOI: 10.5281/zenodo.7035350 (2021).
2. Santuz, H., Bacle, A., Poulain, P. & Fuchs, P. F. buildh: Build hydrogen atoms from united-atom molecular dynamics of lipids and calculate the order parameters. *J. Open Source Softw.* **6**, 3521, DOI: 10.21105/joss.03521 (2021).
3. Bacle, A. *et al.* Inverse conformational selection in lipid–protein binding. *J. Am. Chem. Soc.* **143**, 13701–13709 (2021).

| 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 |
| PYPC | 1-(16:0)-2-(16:1$^{\Delta}$9)-sn-glycero-3-phosphocholine |
| PAzePCprot | 1-palmitoyl-2-azelaoyl-sn-glycero-3-phosphocholine protonated |
| PAzePCdeprot | 1-palmitoyl-2-azelaoyl-sn-glycero-3-phosphocholine deprotonated |
| DMPC | 1,2-dimyristoyl-sn-glycero-3-phosphocholine |
| DPPC | 1,2-dipalmitoyl-sn-glycero-3-phosphocholine |
| DPPE | 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine |
| DPPG | 1,2-dipalmitoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt) |
| DEPC | 1,2-dierucoyl-sn-glycero-3-phosphocholine |
| DRPC | 1,2-(14:1$^{\Delta}$9)-sn-glycero-3-phosphocholine |
| DYPC | 1,2-(16:1$^{\Delta}$9)-sn-glycero-3-phosphocholine |
| DLPC | 1,2-dilauroyl-sn-glycero-3-phosphocholine |
| DLIPC | 1,2-dilinoleoyl-sn-glycero-3-phosphocholine |
| DOG | 1,2-dioleoyl-sn-glycerol |
| DOPC | 1,2-dioleoyl-sn-glycero-3-phosphocholine |
| DOPE | 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine |
| DDOPC | 1,2-didocosahexaenoyl-sn-glycero-3-phosphocholine |
| DOPS | 1,2-dioleoyl-sn-glycero-3-phospho-L-serine |
| DSPC | 1,2-distearoyl-sn-glycero-3-phosphocholine |
| DAPC | 1,2-diarachidonoyl-sn-glycero-3-phosphocholine |
| SLiPC | 1-(18:0)-2-(18:2 $^{\Delta 9,12}$)-sn-glycero-3-phosphocholine |
| DMTAP | 1,2-dimyristoyl-3-trimethylammonium-propane |
| SOPC | 1-stearoyl-2-oleoyl-sn-glycero-3-phosphocholine |
| POPI | |
| SAPI | |
| SLPI | |
| SDG | 1-stearoyl-2-docosahexaenoyl-sn-glycerol |
| SDPE | 1-stearoyl-2-docosahexaenoyl-sn-glycero-3-phosphoethanolamine |
| CER | N-palmitoyl-D-erythro-sphingosine |
| CHOL | cholesterol |
| DCHOL | 18,19-di-nor-cholesterol |
| DHMDMAB | dihexadecyldimethylammonium |
| POT | potassium ion |
| SOD | sodium ion |
| CLA | chloride ion |
| CAL | calcium ion |
| CES | caesium ion |
| SOL | water |

**Table S2.** Abbreviations for molecules 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.