

# **SUPPLEMENTARY INFORMATION: NMRLipids**

## **Databank makes data-driven analysis of biomembrane properties accessible for all**

**Anne Kiirikki<sup>1</sup>, Hanne Antila<sup>2</sup>, Lara Bort<sup>2</sup>, Pavel Buslaev<sup>3</sup>, Fernando Favela<sup>4</sup>, Tiago Mendes Ferreira<sup>5</sup>, Patrick Fuchs<sup>6,7</sup>, Rebeca Garcia-Fandino<sup>8,9</sup>, Ivan Gushchin<sup>2</sup>, Batuhan Kav<sup>2</sup>, Milla Kurki<sup>2</sup>, Alexander Kuzmin<sup>2</sup>, Jesper Madsen<sup>10,11</sup>, Markus S. Miettinen<sup>2</sup>, Thomas Piggot<sup>12</sup>, Suman Samantray<sup>2</sup>, Fabián Suarez-Leston<sup>2</sup>, and O. H. Samuli Ollila<sup>1,\*</sup>**

<sup>1</sup>University of Helsinki, Institute of Biotechnology, Helsinki, Finland

<sup>2</sup>Affiliation, department, city, postcode, country

<sup>3</sup>Nanoscience Center and Department of Chemistry, University of Jyväskylä, P.O. Box 35, Jyväskylä, 40014 , Finland

<sup>4</sup>Departamento de Ciencias Básicas, Tecnológico Nacional de México - ITS Zacatecas Occidente, Sombrerete, Zacatecas, 99102, México

<sup>5</sup>NMR group - Institute for Physics, Martin Luther University Halle-Wittenberg, Halle (Saale), 06120, Germany

<sup>6</sup>Sorbonne Université, Ecole Normale Supérieure, PSL University, CNRS, Laboratoire des Biomolécules (LBM), Paris, 75005, France

<sup>7</sup>Université de Paris, UFR Sciences du Vivant, Paris, 75013, France

<sup>8</sup>Center for Research in Biological Chemistry and Molecular Materials (CiQUS), Universidade de Santiago de Compostela, Santiago de Compostela, E-15782, Spain

<sup>9</sup>CiQUP, Centro de Investigação em Química, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Porto, 4169-007, Portugal

<sup>10</sup>Department of Chemistry, The University of Chicago, Chicago, Illinois, 60637, United States of America

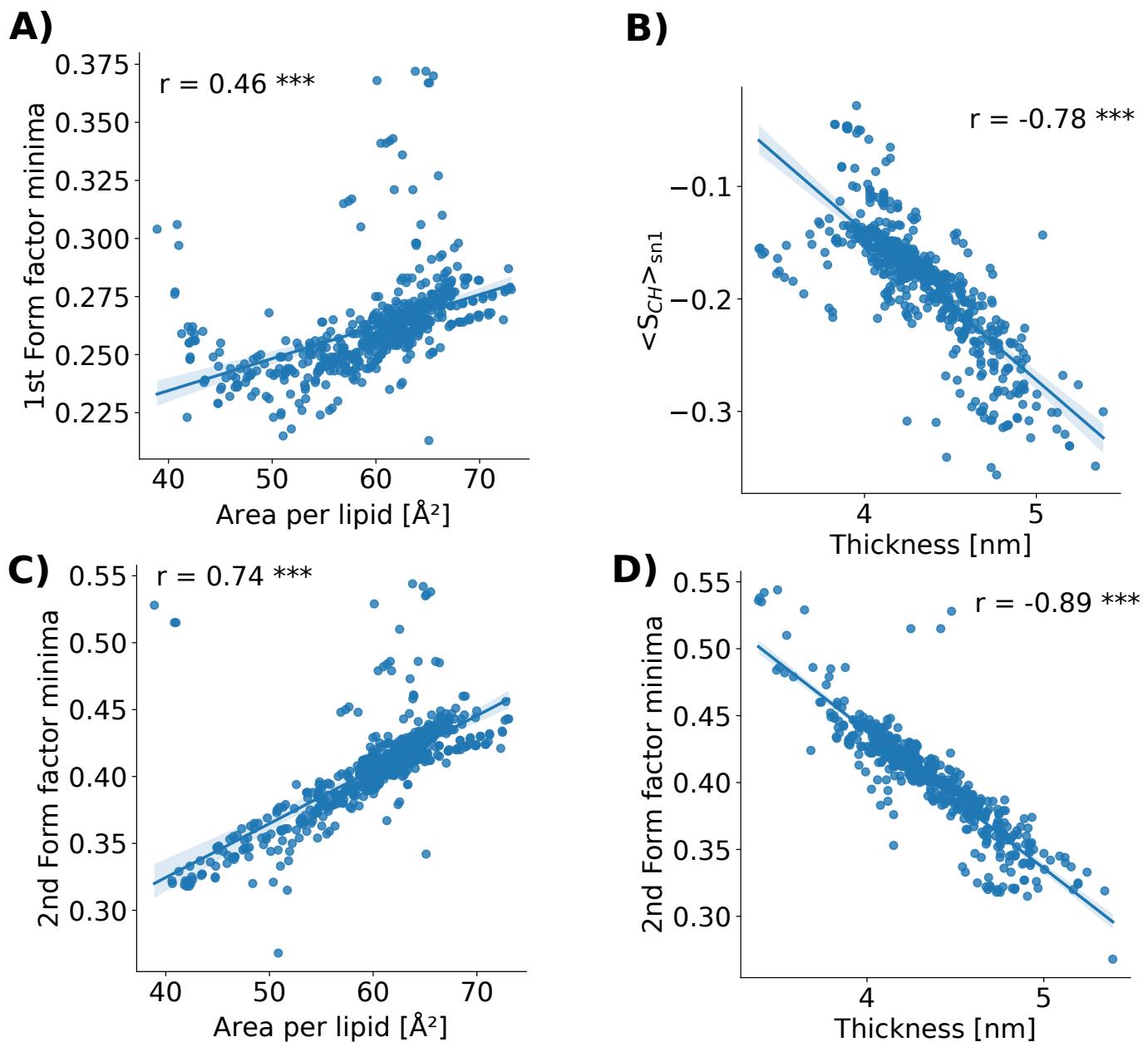
<sup>11</sup>Global and Planetary Health, College of Public Health, University of South Florida, Tampa, Florida, 33612, United States of America

<sup>12</sup>Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom

\* samuli.ollila@helsinki.fi

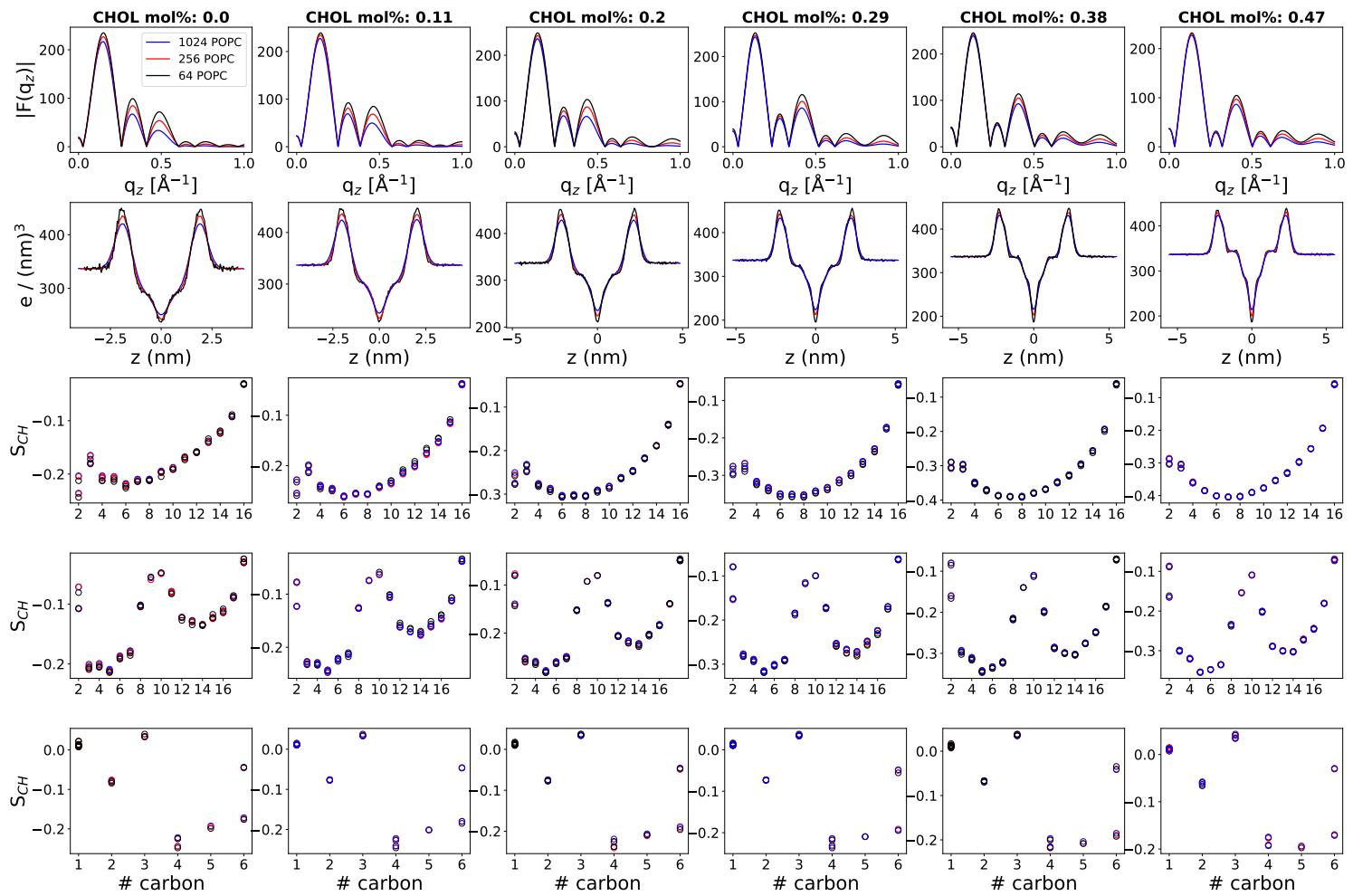
## **ABSTRACT**

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



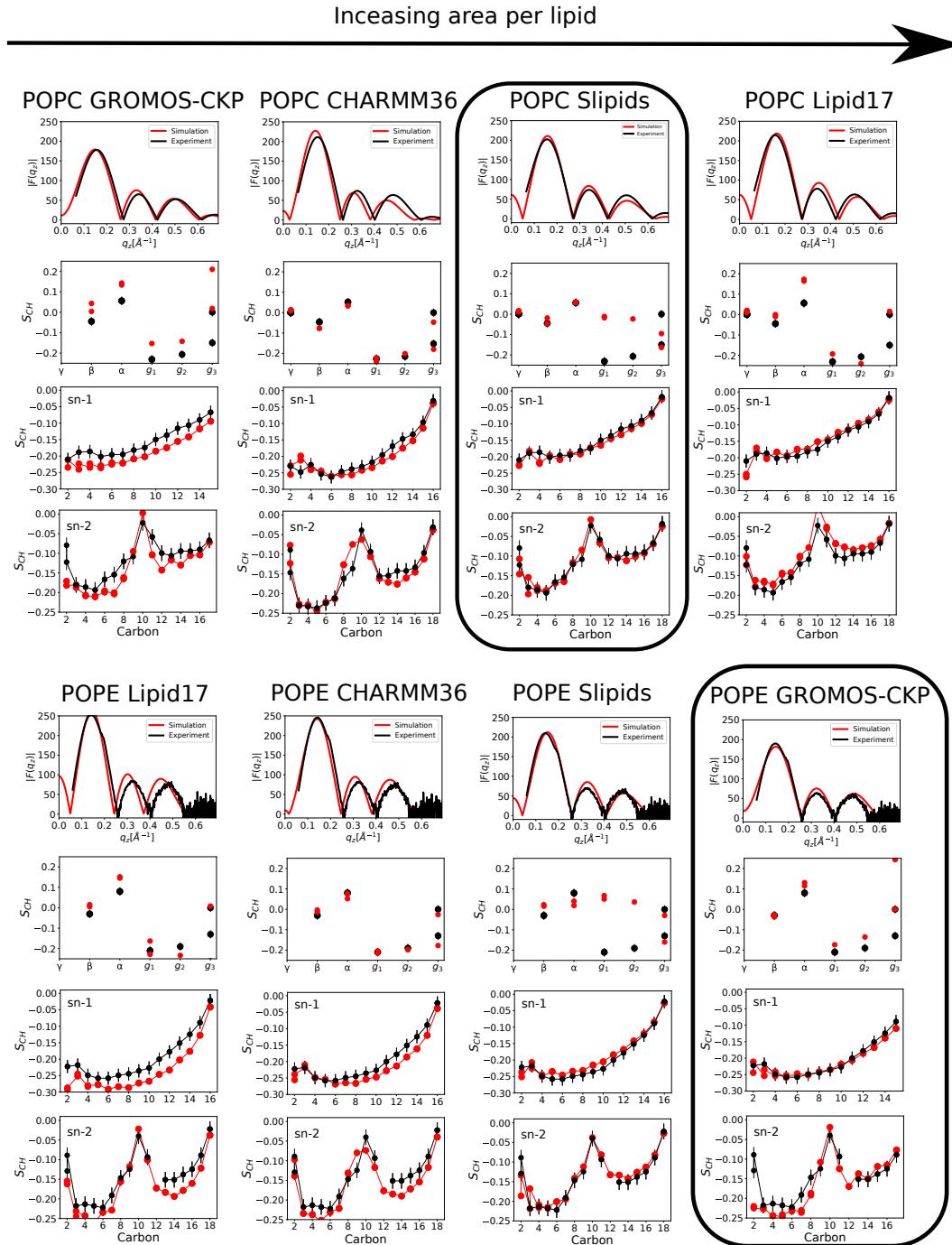
**Figure S1.** Scatter plots and Pearson correlation coefficients,  $r$ , for the membrane area per lipid with X-ray scattering form factor minima (A and B), and for thickness with the average order parameter of the sn-1 acyl chain (B) and with the second minimum from X-ray scattering form factors (D) extracted from the NMRLipids databank. All correlation coefficients have p-value below 0.001 as indicated by \*\*\*.

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



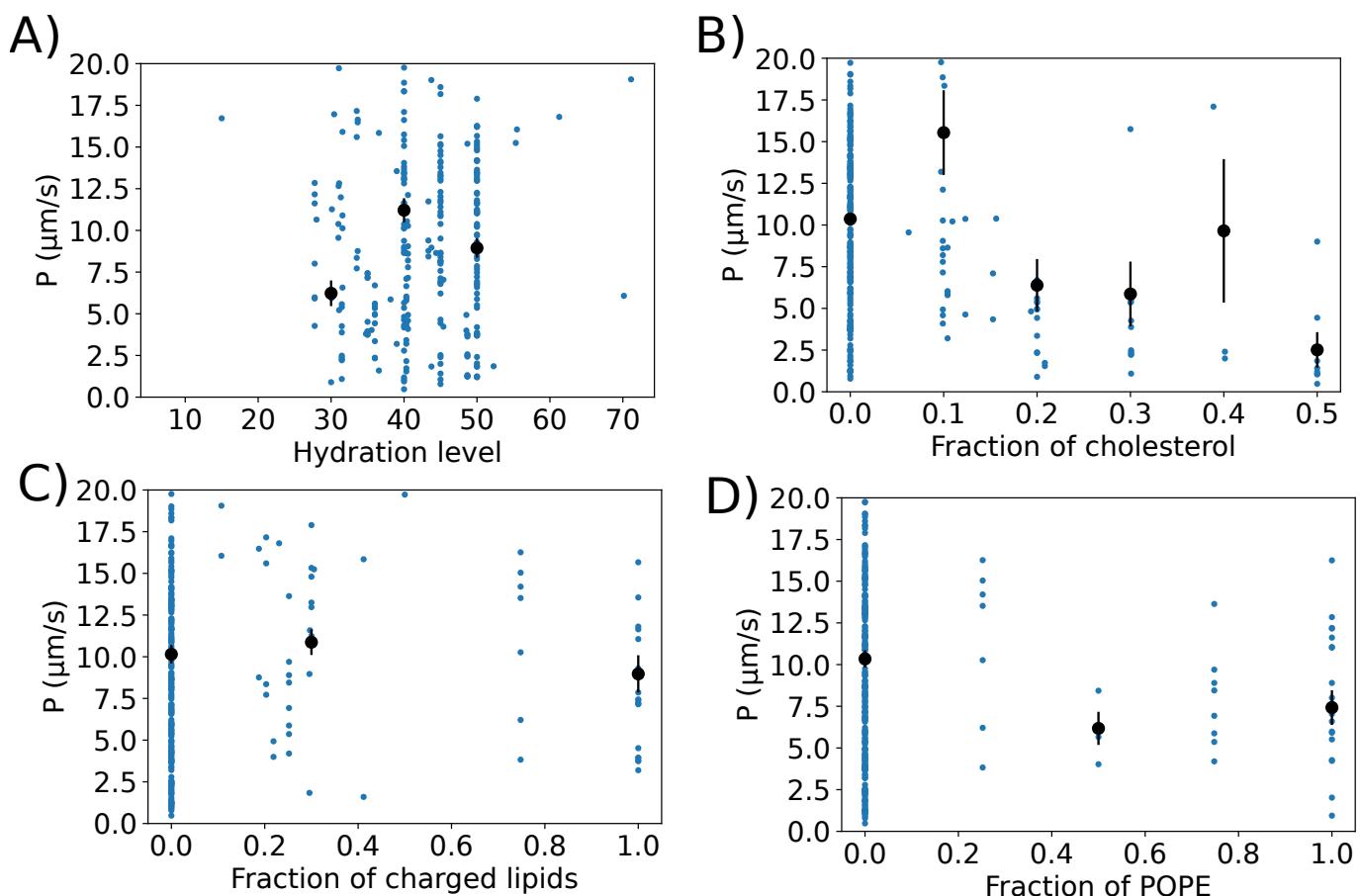
**Figure S2.** Dependence of form factor, electron density profiles along membrane normal, and C-H bond orderparameters (from top to bottom) on simulation box size with different cholesterol concentration. Simulations with 64, 256 or 1024 POPC lipids taken from Ref. 1.

### S3 Finding the best models for PC and PE mixtures



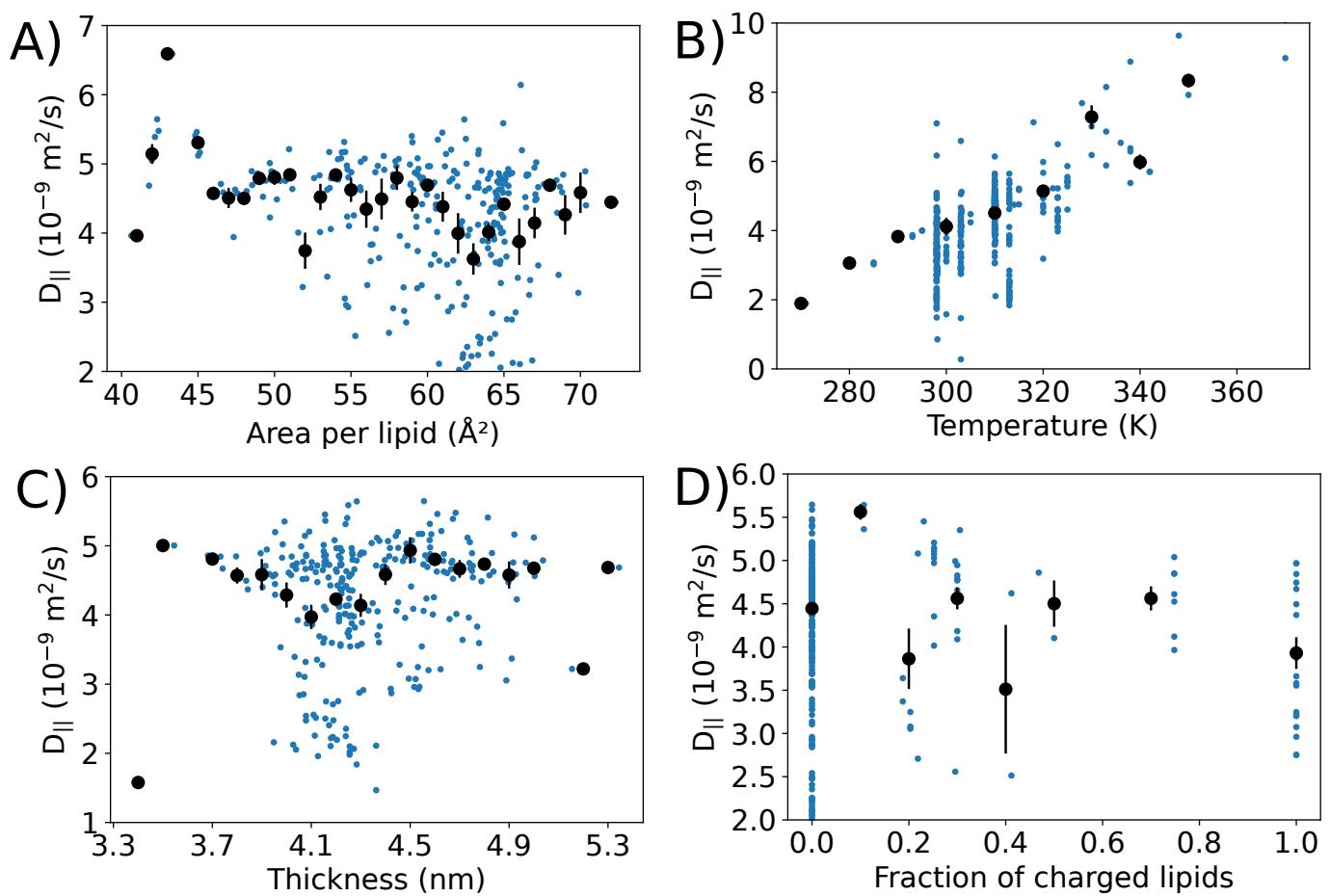
**Figure S3.** Simulations with the data for both POPC (top) and POPE (bottom) directly compared with the experimental data. The area per lipid in simulations increase from left to right. Simulations with the best overall quality for POPC and POPE order parameters are circled by a square.

## S4 Water permeation through membranes



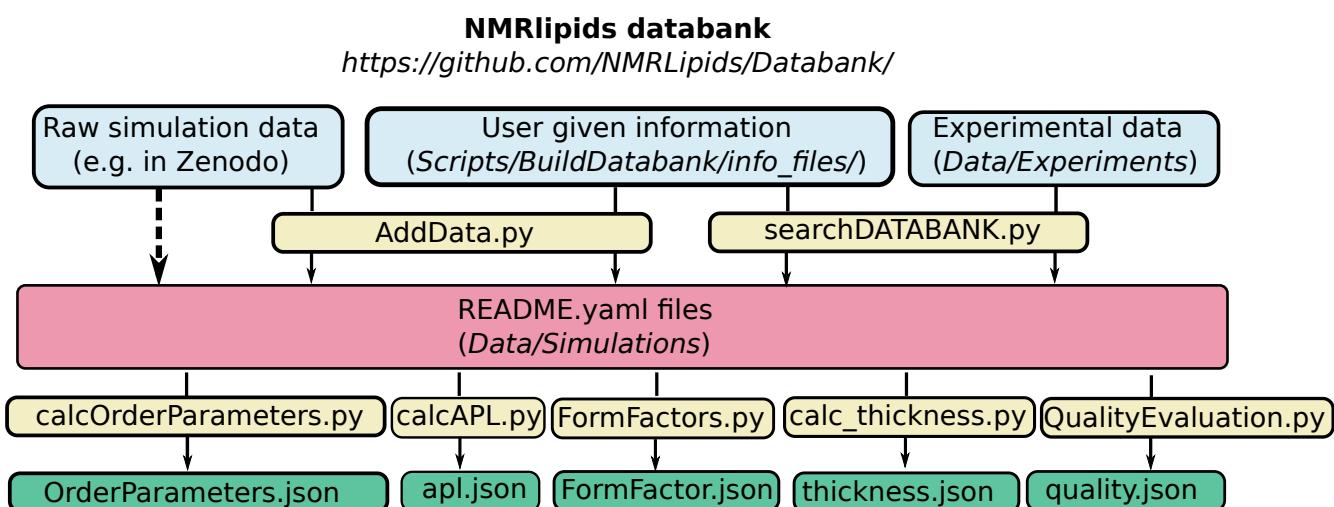
**Figure S4.** Water permeation through membranes analyzed from the databank as a function of (A) hydration level, (B) fraction of cholesterol, (C) fraction of charged lipids, and (D) fraction of POPE in membrane. Values from simulations with non-zero permeation values are shown with blue dots. Histogrammed values are shown with black dots.

## S5 Water diffusion along membranes



**Figure S5.** Lateral diffusion of water as a function of (A) area per lipid, (B) temperature, (C) membrane thickness, and (D) fraction of charged lipids in a membrane.

## S6 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 <sup>2</sup> . 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.

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.

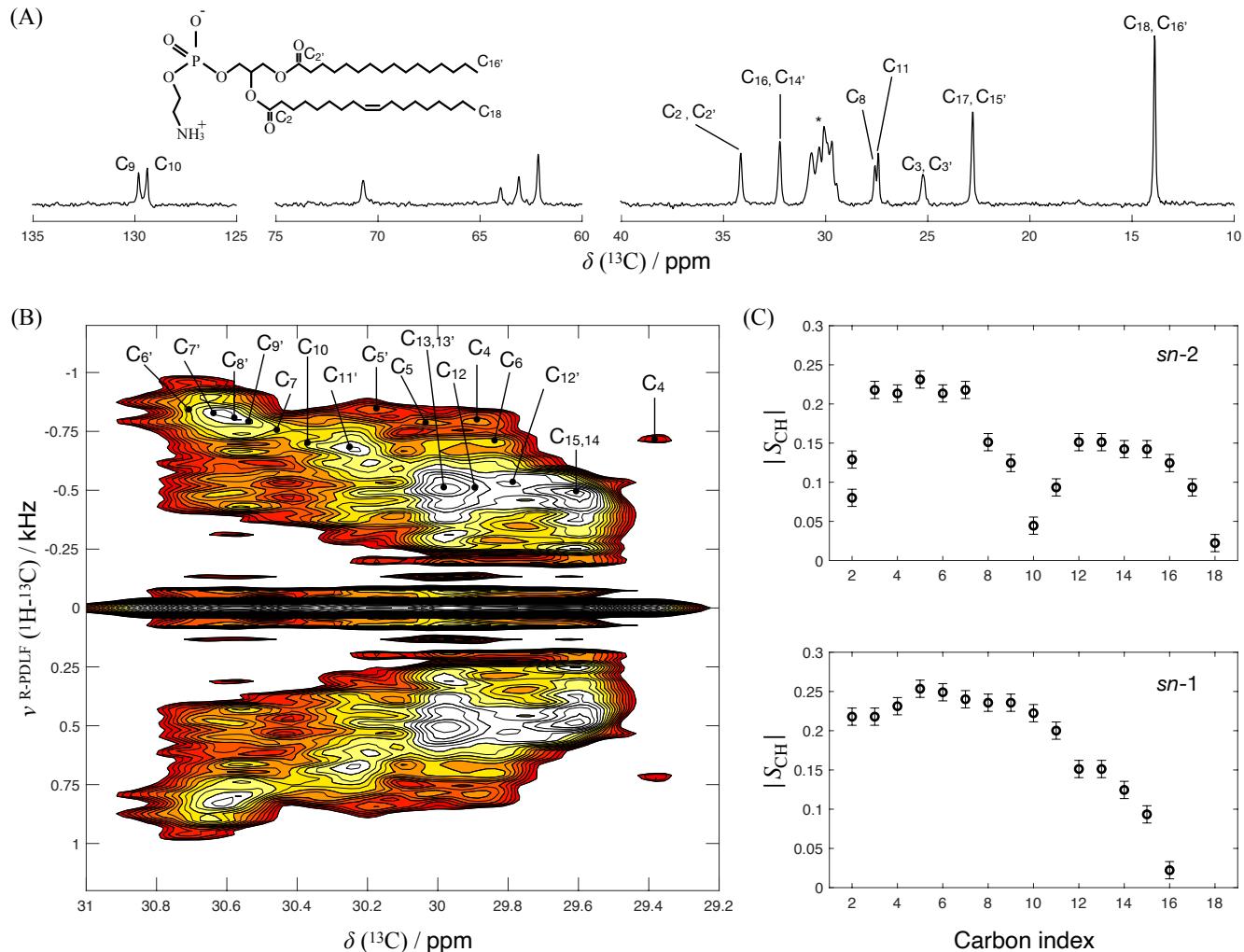
**Table S2.** Keys stored in the README.yaml files of experiments.

Abbreviation	Molecule name
<b>Lipids and surfactants</b>	
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 <sup>A</sup> 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 <sup>A</sup> 9)-sn-glycero-3-phosphocholine
DYPC	1,2-(16:1 <sup>A</sup> 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 <sup>A</sup> 9, <sup>B</sup> 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
<b>Other molecules</b>	
POT	potassium ion
SOD	sodium ion
CLA	chloride ion
CAL	calcium ion
CES	caesium ion
SOL	water

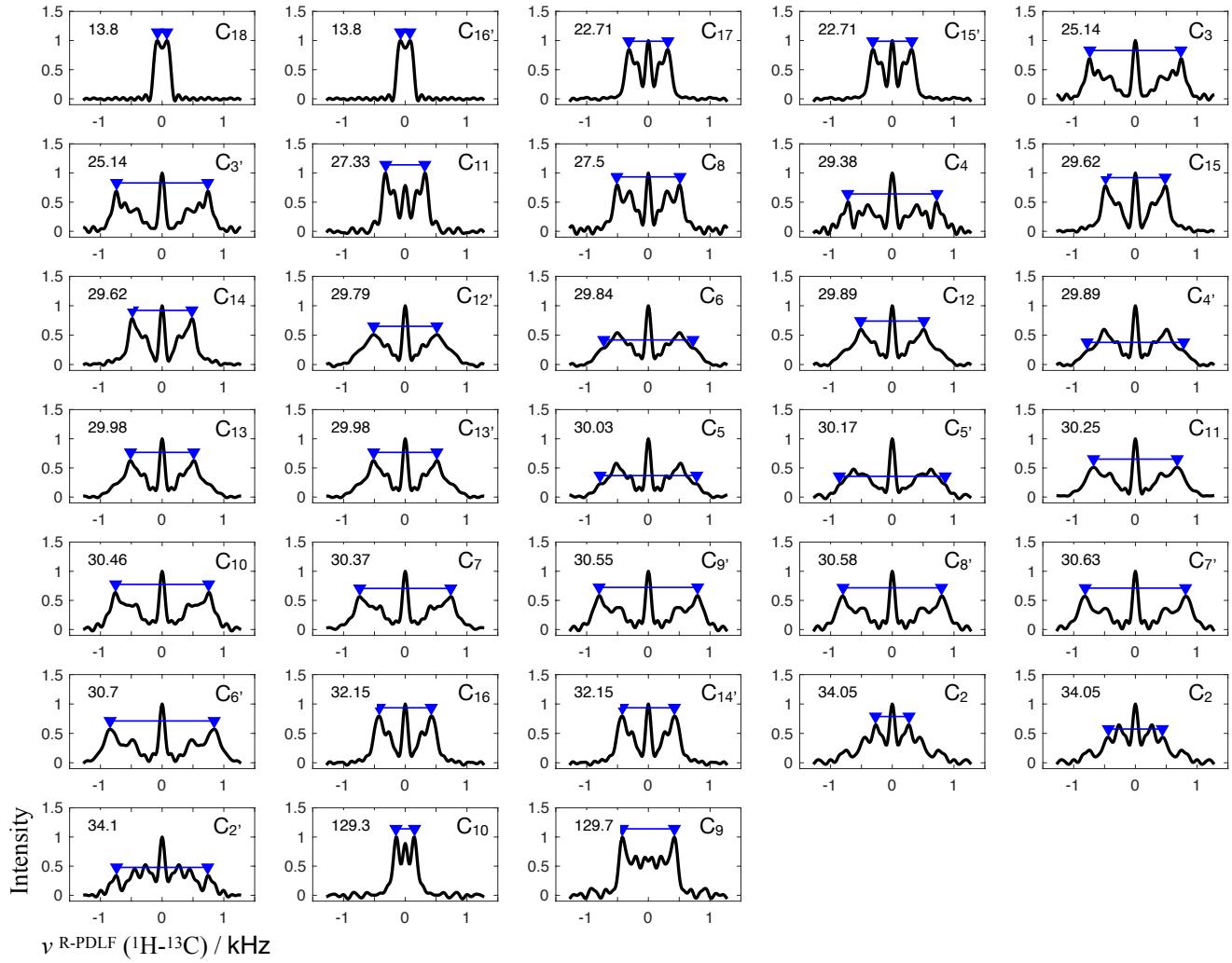
**Table S3.** Abbreviations for molecules used in the databank

## S7 NMR experiments

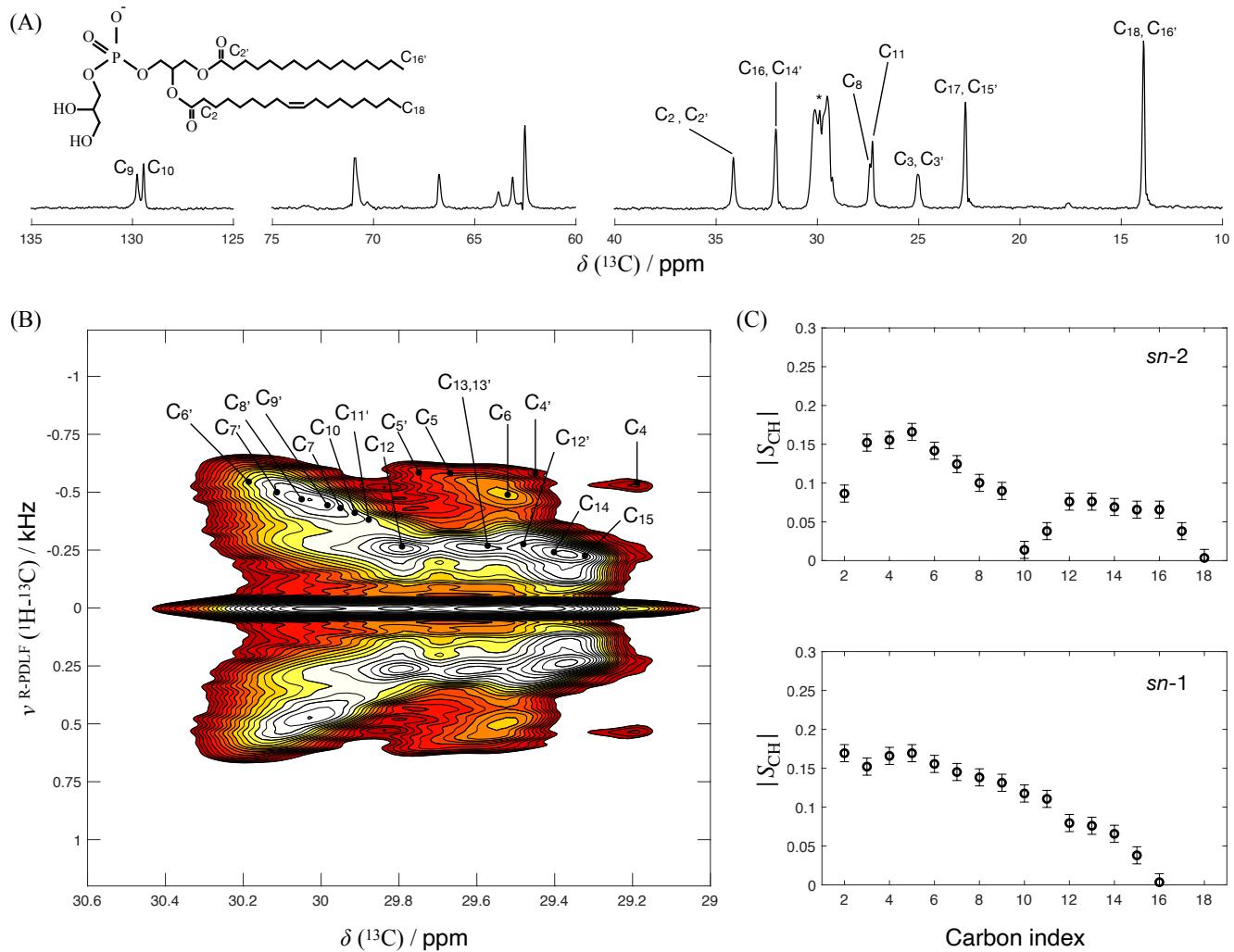
Acy chain order parameters of POPE and POPG were analyzed from the experiments that were previously used to determine headgroup order parameters<sup>3</sup>. The analysis of the crowded spectral region at 29-31 ppm was based on the previous assignment reported for POPC membranes<sup>4</sup>.



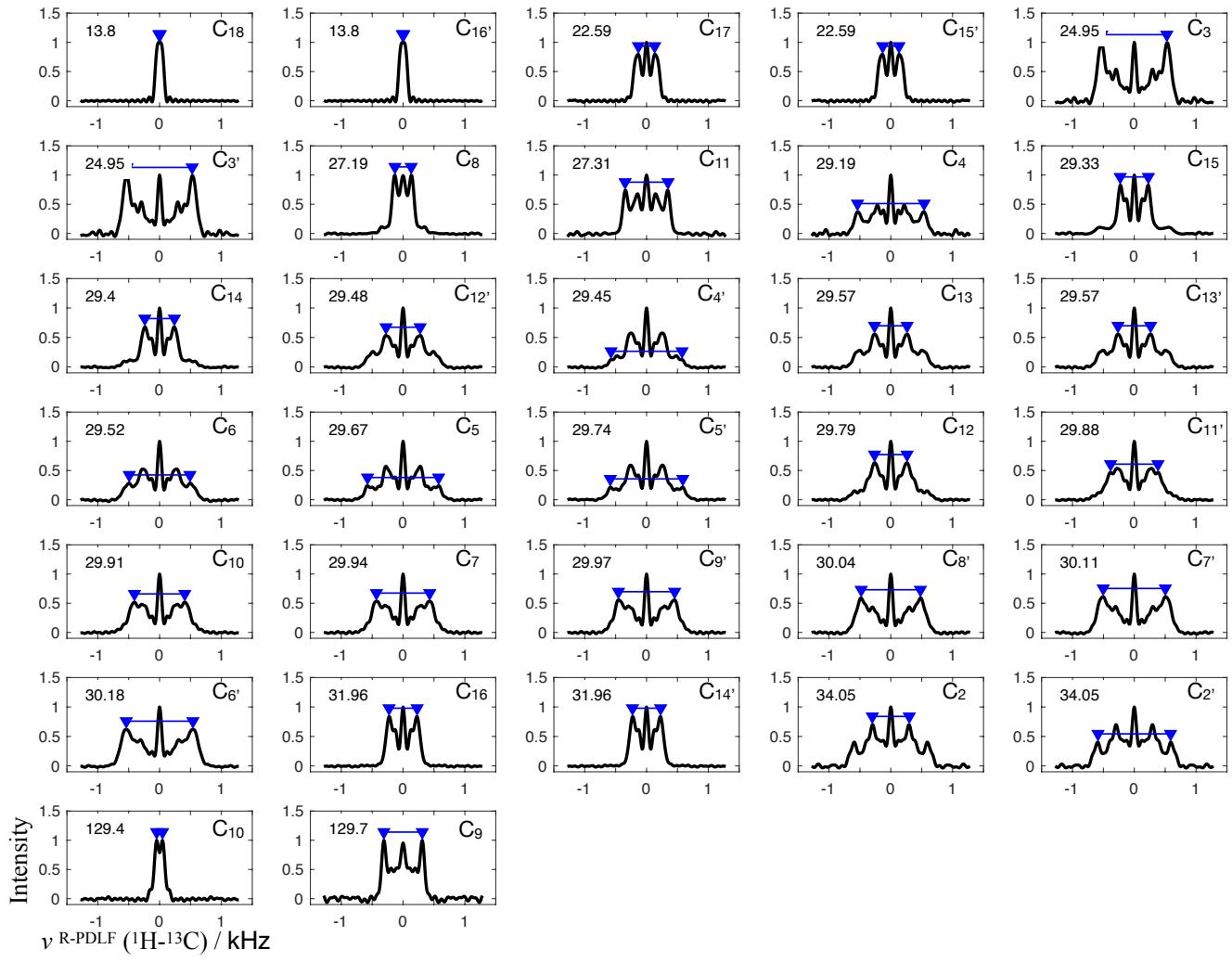
**Figure S7.** Determination of the POPE acyl chain order parameters from a R-PDLF spectrum measured at a magic angle spinning frequency of 5.15 kHz. (A)  $^{13}\text{C}$  rINEPT spectrum with peak assignment. The labels used are shown in the chemical structure of POPE. The chemical shift of the methyl groups was defined as 13.8 ppm. (B) Contour plot of the R-PDLF spectrum for the crowded spectral region. The assignment was based on a previous assignment reported for POPC membranes<sup>4</sup>. (C) C-H order parameter profile for the acyl chains of POPE. The splittings used for calculating the order parameters are shown in Fig. S8. The unassigned peaks belong to the headgroup and glycerol backbone carbons. A detailed assignment and order parameter analysis of these carbons was shown previously<sup>3</sup>.



**Figure S8.** Dipolar spectra obtained from the 2D R-PDLF spectrum from POPE in Fig. S7. The number at the top left of each plot denotes the corresponding chemical shift. The carbon label for each splitting is displayed on the top right. The labels are the same as in Fig. S7.



**Figure S9.** Determination of the POPG acyl chain order parameters from a R-PDLF spectrum measured at a magic angle spinning frequency of 5.15 kHz. (A)  $^{13}\text{C}$  rIN EPT spectrum with peak assignment. The labels used are shown in the chemical structure of POPG. The chemical shift of the methyl groups was defined as 13.8 ppm. (B) Contour plot of the R-PDLF spectrum for the crowded spectral region. The assignment was based on a previous assignment reported for POPC membranes<sup>4</sup>. (C) C-H order parameter profile for the acyl chains of POPG. The splittings used for calculating the order parameters are shown in Fig. S10. The unassigned peaks belong to the headgroup and glycerol backbone carbons. A detailed assignment and order parameter analysis of these carbons was shown previously<sup>3</sup>.



**Figure S10.** Dipolar spectra obtained from the 2D R-PDLF spectrum described in Fig. S9. The number at the top left of each plot denotes the corresponding chemical shift. The carbon label for each splitting is displayed on the top right. The labels are the same as in Fig. S9.

## References

1. Javanainen, M. Simulations of POPC/cholesterol mixtures at 298 K, three system sizes, CHARMM36, DOI: [10.5281/zenodo.7035350](https://doi.org/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](https://doi.org/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).
4. Ferreira, T. M. *et al.* Cholesterol and POPC segmental order parameters in lipid membranes: solid state  $^1\text{H}$ - $^{13}\text{C}$  NMR and MD simulation studies. *Phys. Chem. Chem. Phys.* **15**, 1976–1989 (2013).