Using open data to benchmark internal dynamics of phosphatidylcholine in molecular dynamics simulations

Hanne S. Antila,[†] Tiago Ferreira,[‡] Matti Javanainen,[¶] O. H. Samuli Ollila,[§] and Markus S. Miettinen[*,†]

[†]*Department of Theory and Bio-Systems, Max Planck Institute of Colloids and Interfaces, 14476 Potsdam, Germany*
[‡]*Tiago's affiliation here*
[¶]*Add Matti to author list?*
[§]*Samuli's affiliation here*

E-mail: markus.miettinen@mpikg.mpg.de

Abstract

Brilliant abstract here

1 Introduction

Ever since the conception of the Protein Data Bank (PEB) and the NBCI GenBank, the access to open data has shaped the state of the art of research in life sciences. Not only has the development of new characterisation techniques (such as molecular replacement[1] in macromolecular x-ray crystallography and 3D electron microscopy) been aided by the existence of these databases,[2] but perhaps more importantly they have lead to entirely new ways of doing science in the form of bio- and cheminformatics, enabling data driven discovery of drugs,[3] materials[4] as well as identifying[5,6] and filling[7] gaps in the databanks themselves. All in all, open access to standardised and searchable pools of experimental data, constantly extending owing to a collaborative effort, has enabled scientific progress that is well beyond the resources of one single research group.

Much of the development of the PDB and other biomolecular databanks into the core resources they are today has been fueled by the push from scientific journals and funders towards public availability and conservation of data. In addition to experimental results, these principles have more recently extended to simulation trajectories of biomolecules, leading to databases of dynamic information. Inspired by the success of bio- and cheminformatics, we seek to exploit this and demonstrate, for the first time, the viability of creating new scientific knowledge analysis of pre-existing, open access simulation data.

More spesifically, we will analyze a wide set of publicly available phosphatidylcholine (PC) phospolipid bilayer molecular dynamics (MD) trajectories, including simulations of varying hydration, salt concentration, and cholesterol content in addition to one component bilayers under standard conditions, that has been produced by utilizing different MD models (force fields). We test whether different force fields reproduce the experimentally observed dynamics of PC lipids, and investigate if the dynamics extracted from various models share common features that can be used to draw general conclusions on the system and suggest future directions for experimental research as well as to avoid potential pitfalls in simulations of bilayers.

Our choice for the system of interest is inspired by the importance of phospolipids not only as the building blocks of cell membranes but also as emerging candidates for micro- and nanotechnology, such as the use of liposomes as microcapsules in targeted drug delivery.[8] These molecules are composed of a hydrophilic phosphate head group, which is connected to two hydrophopic fatty acid tails via a glycerol backbone. The ability of lipids to self assemble into bilayer membrane (and other) configurations is a direct consequence of this dual nature. Although biological membranes are complex mixtures of multiple different lipids as well as other molecules, lamellar phospholipid bilayers with one or few lipid types serve as an important model system, that have been successfully used to decipher, *eg.*, possible molecular mechanisms behind anesthetics,[9?] the effect of cholesterol on membrane structure,[10?] and the functioning of membrane proteins[11] **1.add more references**. In particular, MD simulations of these model systems have been widely used[9,10,12? –15] to provide an atomistic view on the biomembranes, and hold vast potential in making further connections between the structure and the function.

The significance for investigating the conformational dynamics of lipids in these bilayer simulations is two-fold. Firstly, when exploring static properties of the bilayers, it is crucial to assess how well the simulations have converged. In order to extract reliable statistics, the conformations sampled have to represent the equilibrium distribution with enough transitions between states. Indeed, simulations of a single (1,2-dioleoyl-sn-glycero-3-phosphocholine) DOPC lipid using the CHARMM32b2 force field indicated that the conformations sampled do not replicate the equilibrium distribution even after 500 ns[16] and the bond dynamics of the Berger model was shown[14] to be too slow at the glycerol region of 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) compared to correlation times extracted from NMR experiments.

Secondly, for complete picture of membrane functioning, knowledge on the bilayer dynamics in addition to equilibrium measurements are needed. The ability of the MD model to reproduce the relative abundance of different dynamical processes is crucial for the correct interpretation of pathways leading to, e.g., membrane deformation[17] and lipid-induced conformational[18,19] changes of membrane proteins. The availability of such model could also greatly guide the both the configuration and interpretation of NMR experiments used to extract dynamical information from lipid assemblies. **2.the following paragraph could be merged into methods, opinions?** Our analysis of the lipid dynamics is based on two quantities, the spin-lattice relaxation rates R_1 and the effective correlation time τ_e, both experimentally available through NMR measurements. Out of the two, the R_1 rates (or the corresponding T_1 times) have been traditionally used to assess both the conformational dynamics of lipids in experimental bilayers[20–24] and the dynamics produced by lipid MD models in bilayer simulations.[20,22,23,25] However, relying on R_1 only has several drawbacks **3.do all flavors (31P,13C,...) have the same problem?**. It builds on an underlying rotation-diffusion model, its sensitivity is typically limited to C–H bond reorientation with time scales ∼1-10 ns, and measurements at several temperatures and magnetic field strengths are required to fully characterize the dynamics. To address these deficiencies, two of us introduced a procedure[14] for quantifying the effective C–H correlation times (τ_e)—a model free quantity that encompasses conformational dynamics with time scales up to hundreds of nanoseconds—from bilayer systems. Most importantly, increasing τ_e always signals some type of slowdown in the C–H bond dynamics, making the interpretation less ambiguous than for R_1, where slowdown in the dynamics can lead to either an increase or a decrease of R_1 value.[14]

In summary, this work provides first comprehensive comparison of dynamics of different phosphatidylcholine MD models, where both pure bilayers and the model response to changing conditions and composition is explored. The study is conducted using data-driven exploration of pre-existing, publicly available simulation trajectories to demonstrate the power of open, well documented data in creating new knowledge at a lowered computational cost and

high potential for automation.

2 Methods

3 Theoretical background

The ^{13}C NMR experiments investigating the lipid conformational dynamics take advantage of the fact that the relaxation of ^{13}C magnetization dominantly happens via the dipolar coupling of the carbon with the magnetic moments of the protons bound to it, with the symmetry axis of the interaction aligning with the C–H bond. The spectral density depicting the ^{13}C relaxation rates (at frequency ω) is expressed as

$$\mathrm{j}(\omega) = 2\int_0^\infty \cos(\omega\tau)\mathrm{g}(\tau)d\tau, \qquad (1)$$

which is the Fourier transformation of the C–H bond second order autocorrelation function at time τ

$$\mathrm{g}(\tau) = \langle P_2\left[\vec{\mu}(t)\cdot\vec{\mu}(t+\tau)\right]\rangle, \qquad (2)$$

where $\vec{\mu}(t)$ is the unit vector in the direction of the C–H bond at time t and P_2 is the second order Legendre polynomial. The angular brackets depict averaging over time. The autocorrelation function can be expressed as the product of two functions

$$\mathrm{g}(\tau) = \mathrm{g_f}(\tau)\mathrm{g_s}(\tau), \qquad (3)$$

where $\mathrm{g_f}(\tau)$ characterizes fast decays owing to, for example, the molecular rotations, and $\mathrm{g_s}(\tau)$ describes slow decays that originate from, e.g., the lipid diffusion. The two components, along with the oscillation due to magic angle spinning at the ∼kHz region, are depicted in Fig. 1. Correlation time of 4.2 ms has been estimated for multilamellar POPC samples at 300 K for the slow modes, whereas in liquid crystalline lipid bilayers the faster $\mathrm{g_f}(\tau)$ decays to a plateau value S^2_{CH} within a few hundred nanoseconds.[14] The order parameters

$$S_{\mathrm{CH}} = \frac{1}{2}\langle 3\cos^2\theta - 1\rangle, \qquad (4)$$

where θ is the angle between the C–H bond and the bilayer normal, are measured in NMR experiments from this plateau. As S_{CH} describes the conformational ensemble of the molecule, the fast-decaying component of the rotational correlation function intuitively depicts the time needed to sample these conformations. The characteristic time can be quantified via the effective correlation time

$$\tau_e = \int_0^\infty \frac{\mathrm{g_f}(\tau) - S^2_{\mathrm{CH}}}{1 - S_{\mathrm{CH}}}d\tau \qquad (5)$$

which is defined via the area of the normalized correlation function ($\mathrm{g_f'}(\tau) = (\mathrm{g_f}(\tau) - S^2_{\mathrm{CH}})\,/\,(1 - S_{\mathrm{CH}})$) and graphically depicted in Fig. 1b. It is easily seen that in the presence of more long-lived correlations τ_e grows, signaling that more time is needed for full conformational sampling.

The bond correlation functions are readily available from MD simulations. In addition to extracting τ_e directly from integrating the area,

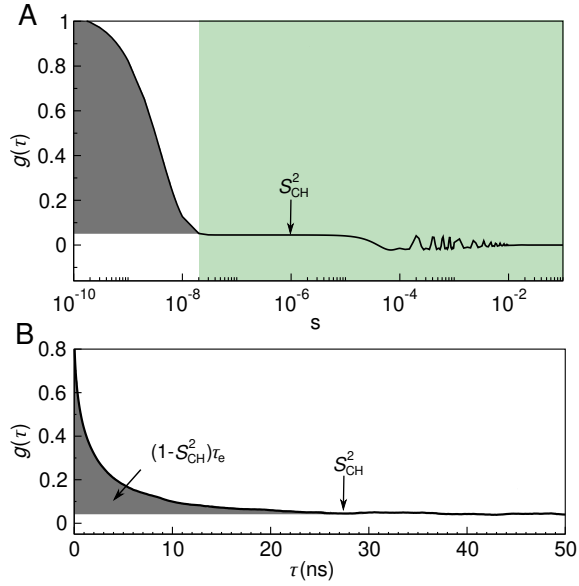

Figure 1: The autocorrelation function $g(\tau)$ a) The fast mode (white background) and the slow mode (shaded green) of the correlation function along with the oscillation owing to magic angle spinning. The fast mode decays to a plateau quantifying the S_{CH} while the slow mode gives the final descent to zero. b) Illustration of typical C–H bond autocorrelation function obtained from a MD simulation. The gray area under the curve gives a means of quantifying the τ_e.

fitting a set of N exponentials, each with their own correlation times τ_i to the normalized correlation function

$$g'_{\mathrm{MD}}(t) = \sum_{i=1}^{N} \alpha_i e^{-t/\tau_i} \tag{6}$$

provides an alternative way of quantifying τ_e

$$\tau_e = \sum_{i=1}^{N} \alpha_i \tau_i. \tag{7}$$

When the simulation trajectory is not long enough for the correlation function to reach the plateau, integrating the area gives a lower bound estimate for τ_e while the latter method includes also contribution from the longer-time components via the fitting process. However, in practice the fit is often highly unreliable in terms of depicting the long tails of the correlation function, and in this work we choose to quantify τ_e using the area.

The spin-lattice relaxation rate R_1 defines the time-scale on which ^{13}C longitudal magnetization equilibrates. It is defined as

$$\begin{aligned} R_1 = \frac{d_{\mathrm{CH}} N_{\mathrm{H}}}{20} \left[j(\omega_{\mathrm{H}} - \omega_{\mathrm{C}}) \right. \\ \left. + 3j(\omega_{\mathrm{C}}) + 6j(\omega_{\mathrm{H}} + \omega_{\mathrm{C}}) \right], \end{aligned} \tag{8}$$

where N_{H} is the number of bound hydrogens, ω_{H} and ω_{C} are the Larmor frequencies for ^{1}H and ^{13}C, and d_{CH} is the rigid dipolar coupling constant. For the methylene bond, $d_{\mathrm{CH}}/2\pi$ approximately equals to -22kHz.

The dependency of R_1 on the spectral densities at the Larmor frequencies means that the R_1 value depicts the relative amounts of relaxation processes with time-scales near the inverses of these frequencies. Since the Larmor frequencies depend on the field strength used in the NMR measurements, this typically makes R_1 sensitive to ∼1-10 ns time-scales. Importantly, a change in R_1 indicates a difference in the relative amounts of processes within the detection window, and therefore does not give information on the modulation of the total sampling rate.

Table 1: Simulations of POPC bilayers under full hydration. Number of lipids and number of water molecules are denoted with N_l and N_w, respectively, temperatures (T) are given in kelvins and t_{anal} is the length of the trajectory used for analysis. The column labeled "Files" citation numbers giving links for the downloadable simulation files.

force field	lipid	N_l	N_w	T (K)	t_{anal} (ns)	files
Berger-POPC-07[27]	POPC	128	7290	298	50	[28]
CHARMM36[29]	POPC	128	5120	303	140	[30]
CHARMM36[29]	POPC	17	510	300	140	[31]
MacRog[32]	POPC	128	6400	310	200	[33]
Lipid14[34]	POPC	72	2234	303	50	[35]
Slipids[36]	POPC	200	9000	310	500	[37]
ECC[38]	POPC	128	6400	300	300	[39]

3.1 Data aquisition and analysis

The simulation trajectories used in this work were collected from the Zenodo repository (`zenodo.org`) with majority of the data originating from the NMRlipids project[13,26] (`nmrlipids.blogspot.fi`). A List of the simulations, as well as the references to the data files are presented in Table 1 for trajectories near room temperature and in full hydration, Table 2 for simulations including cholesterol, Table 3 for data under varying hydration, and Table 4 for simulations in increasing NaCl concentration. Additional computational details of each of the simulations are available at the referred Zenodo entry. All the experimental quantities were collected from the literature **4.Except are they, or mostly from Tiago and re-analysed from raw data?** sources referred at the respective figures**5.How to refer to experimental data from Tiago?**.

The data was analysed using in-house scripts available in github[?] along with a python notebook outlining an example analysis run. First, the simulation files were downloaded from zenodo and the trajectory was processed using gromacs *trjconv* to make the molecules whole. The C–H correlation functions were calculated from the trajectory utilizing gromacs[?] *gmx rotac* tool, and normalized by order parameter values obtained from *calcOrderParameters.py*[?] script that uses the MDanalysis[?] python library. The

Table 2: Simulation data for cholesterol-containing POPC bilayers. Number of cholesterols is given by N_{chol} while C_{CHOL} denotes the percentage of cholesterol from all the lipids. Rest of the labels are as in Table 1.

force field	lipid	N_l	N_{chol}	C_{chol}	N_w	T (K)	t_{anal} (ns)	files
Berger-POPC-07[27]	POPC	128	0	0%	7290	298	50	[28]
/Höltje-CHOL-13[10,40]	POPC	64	64	50%	10314	298	60	[41]
CHARMM36[29,42]	POPC	128	0	0%	5120	303	140	[30]
	POPC	80	80	50%	4496	303	200	[43]
MacRog[32]	POPC	128	0	0%	6400	310	200	[33]
	POPC	64	64	50%	6400	310	200	[33]
Slipids[36,44]	POPC	200	0	0%	9000	310	500	[37]
	POPC	200	200	50%	18000	310	500	[37]

Table 3: Simulation data for bilayers under varying hydration level. The water to lipid ratio is denoted as W/L, and other labels are as in Table 1.

force field	lipid	W/L	N_l	N_w	T (K)	t_{anal} (ns)	files
Berger-POPC-07[27]	POPC	57	128	7290	298	50	[28]
	POPC	7	128	896	298	60	[45]
Berger-DLPC-13[46]	DLPC	24	72	1728	300	80	[47]
	DLPC	16	72	1152	300	80	[48]
	DLPC	12	72	864	300	80	[49]
	DLPC	4	72	288	300	80	[49]
CHARMM36[29]	POPC	40	128	5120	303	140	[30]
	POPC	15	72	1080	303	20	[50]
	POPC	7	72	504	303	20	[51]
MacRog[32]	POPC	50	288	14400	310	40	[52]
	POPC	15	288	4320	310	100	[52]
	POPC	10	288	2880	310	100	[52]

normalized correlation functions were then integrated until the time where $g_f'(\tau)$ falls to zero, or in case when the plateau $g_f'(\tau) = 0$ was not reached, to the end of the correlation function data, giving the lower bound estimate for τ_e.

6.Stuff on error estimates here

4 Results

The upper panels in Fig. 2 presents a comparison of experimental effective correlation times, obtained from POPC bilayers in room temperature and under full hydration, to those produced by five different MD force fields. In general, MD models capture the general shape of the τ_e profile (dynamics slows down when approaching the glycerol backbone either from the headgroup or the tails). That said, force fields exhibit a tendency towards slower dynamics than what is observed experimentally in the glycerol region, while the dynamics of the tail C–H bonds are well reproduced. The discrepancy observed around the glycerol backbone is consistent with previous results from the Berger model,[14] as well as with the insufficient conformational sampling of glycerol backbone torsions observed[16] in 500-ns-long CHARMMc32b2[62,63] simulations of a DOPC lipid. The best overall performance is obtained from CHARMM36 and Slipids force fields, although in Slipids dynamics exhibits a qualitatively wrong, decreasing, trend from g_3 to g_1. **7.Not extremely clear, though; would be good to know the errors to make this claim.**

Note that the temperature for simulation data varies across the models, and in general the τ_e values are expected to decrease with increasing temperature. That said, for CHARMM36 we present data for two system sizes and temperatures that is practically identical which leads us to conclude that small variations in both should not considerably affect the results. Furthermore, we emphasize that the values calculated from the simulations give the *lower* limit of τ_e, as we opted to quantify these from the area (see section 3). If trajectories were to be extended, the τ_e value may increase, and the overestimation of the effective correlation time in some models hence might be more severe than what is shown here.

Note that for an united atom model (such as the Berger model studied here) the hydrogens are added to the trajectory after the simulation and the dynamics are not preserved for the methyl C–H bonds by the protonation algorithm. Hence, the data for the γ, oleoyl C18, and palmitoyl C16' carbons from the Berger model are not presented.

In the lower row of panels in Fig. 2 we provide a comparison of experimental R_1 rates to those obtained from the simulations, under the same conditions as for τ_e.

For the glycerol and headgroup region the force fields overall reproduce the experimental R_1 data better**10.What is up with the gamma carbons?** than for the τ_e. For example, the MacRog model produces particularly good R_1 rates result although it systematically overestimates effective correlation times. As the R_1 data presented here was measured at 125 MHz,

Table 4: Simulation data for bilayers under varying concentration of NaCl. Number of Na^+ and Cl^- ions are denoted by N_{Na} and N_{Cl} while [salt] gives the NaCl concentration calculated as [salt] $=N_{Na}$[water]$/N_w$, where [water] = 55.5 M. Other labels are as in Table 1.

force field (lipid, ion)	lipid	[salt] mM	N_l	N_w	N_{Na}	N_{Cl}	T (K)	t_{anal} (ns)	files
CHARMM36[29]	POPC	0	128	5120	0	0	303	140	[30]
CHARMM36,[29] CHARMM36[53]	POPC	350	72	2085	13	13	303	80	[54]
CHARMM36,[29] CHARMM36[53]	POPC	690	72	2085	26	26	303	73	[55]
CHARMM36,[29] CHARMM36[53]	POPC	950	72	2168	37	37	303	60	[56]
MacRog[32]	POPC	0	128	6400	0	0	310	400	[33]
MacRog,[32] OPLS[57]	POPC	100	288	14554	27	27	310	90	[58]
MacRog,[32] OPLS[57]	POPC	210	288	14500	54	54	310	90	[58]
MacRog,[32] OPLS[57]	POPC	310	288	14446	81	81	310	80	[58]
MacRog,[32] OPLS[57]	POPC	420	288	14392	108	108	310	90	[58]
Slipids[36]	POPC	0	200	9000	0	0	310	500	[37]
Slipids,[36] AMBER[59]	POPC	130	200	9000	21	21	310	100	[60]
Slipids,[36] AMBER[59]	POPC	1.0	200	900	162	162	310	200	[61]

this means that the force fields tend to reproduce the rotational dynamics at $(2\pi \times 125\ MHz)^{-1}$ ~1 ns scales for the glycerol region. However, the discrepancies observed between experimental and computational MD τ_e data show that MD does not reproduce the longer-scale dynamics contained in the τ_e values in the glycerol and headgroup regions.

In the tail region, the MD models are also somewhat in agreement with R_1 rates, the Slipids and Berger model are usually underestimating the experimental value, whereas the Lipid14 and ECC mostly produce the largest rates of all the models. As the dynamics of the acyl tails is several times faster than the headgroup and glycerol regions, this too might indicate a more general tendency for MD to succeed in shorter scales whereas rotations with longer than 1 ns time scales are not represented as well.

We note that while some models, like CHARMM36, seem to provide a reasonable representation of the conformational dynamics, none of them produces a set of order parameters in full agreement with the experiments;[13] The effective correlation times therefore depict the time taken to sample a phase space which is either slightly (CHARMM36) or drastically (Slpids) differs from what is observed in the experiments. We will therefore avoid overly detailed discussion on the models and rather concentrate on detecting common trends.

11.look at the decomposition

12.How to motivate, if focus is shifted from the model validation? A comparison to experiments under one set of conditions is not sufficient to fully assess the quality of the dynamics obtained from an MD model: a more complete picture is acquired by evaluating whether the model responds correctly to a change in conditions. We therefore proceed to investigate how the dynamics change when cholesterol is added to the bilayer (Fig. 3), when the hydration level is reduced (Fig. 4), and when monovalent salt is added to the solution (Fig. 5).

The addition of cholesterol causes the conformational dynamics of certain regions of POPC to slow down, whereas others stay constant. This is most evident from the experimental τ_e values presented in the top panel Fig. 3, where a clear increase is detected for g_1, g_2, g_3 and $C2$ carbons accompanied by some evidence of a slow down near the oleanyl double bond. The experimental R_1 rates confirm that a change does occur near the double bond at the short time scale dynamics at $C9$ and $C11$ carbons. Importantly, neither experimental measure detects an effect on the headgroup β or α carbons.

All the force fields investigated in Fig. 3 qualitatively reproduce the slow down in dynamics with CHARMM36 giving the best estimate for the magnitude while Slipids clearly overestimates change, the especially at the glycerol carbons. Notably, some force fields, like Slipids, also predict a slow down for the *alpha* and *beta*

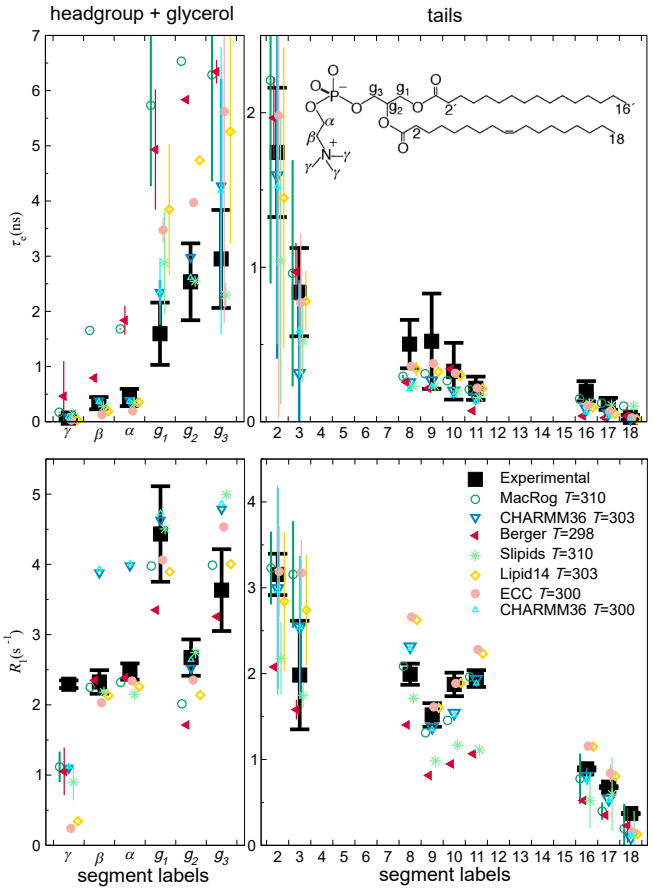

Figure 2: Comparison of experimental effective correlation times (τ_e, upper panels) and R_1 (lower panels) rates to those obtained from MD simulations of POPC bilayers under full hydration. The structure of POPC along with the the carbon labelling scheme is presented on the top. The experimental values are measure from fully hydrated POPC bilayers in L_α phase at 298 K. The data for segment 8–11 is from the sn-2 (oleoyl) chain, whereas non-resolved contributions from both acyl chains are included in segments 2–3 and 16–18 (14–16 for sn-1 chains). The error bars for experimental values reflect error estimate of XXX, whereas the bars for simulated data points give the minimum and maximum value observed at each segment while the symbol denotes the average. Further details on the simulations are provided in Table 1.

8.Error estimate in the experiment has changed since the paper in which these data were originally published; needs to be explained to the reader.

9.How to refer to the experiments?

carbons where no change is detected experimentally. Interestingly, the despite capturing the

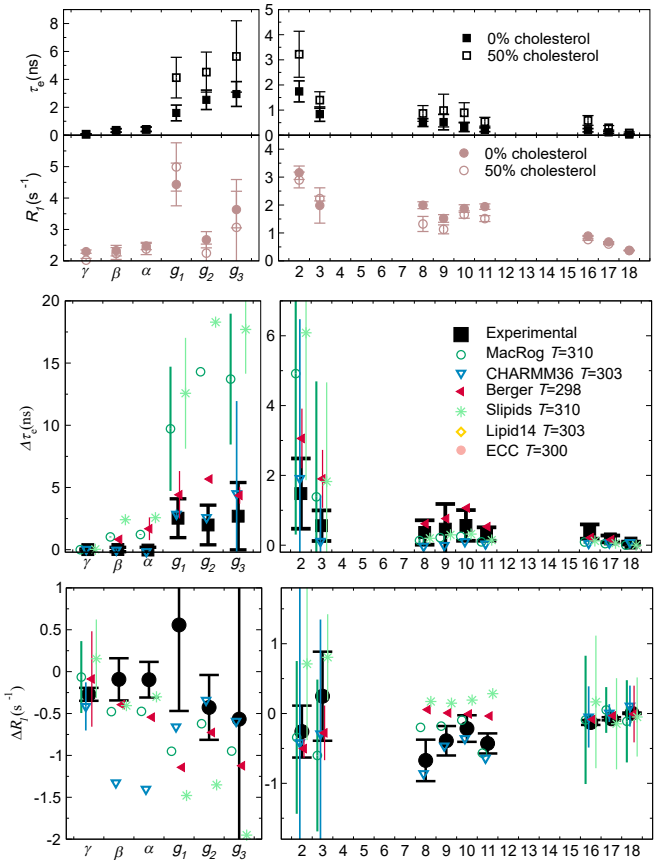

Figure 3: The effect of bilayer cholesterol content on the effective correlation times and R_1. Top panel presents the experimental τ_e and R_1 values both in pure POPC bilayer and when the bilayer contains 50% cholesterol. The experimental data is measured at XXXK and XXX. The middle and bottom panels present the change in τ_e (middle) and R_1 (bottom) when bilayer composition changes from pure POPC to containing 50% cholesterol, both experimentally and in MD simulations. Further details on the simulations are provided in Table 2.

experimentally observed $\Delta\tau_e = 0$, CHARMM36 still gives a non-zero ΔR_1 at α and β carbons which indicates some inaccuracies in the C–H dynamics at shorter time-scales in these carbons. The tail ΔR_1, on the other hand, seem to be well captured by CHARMM36. Along with the slow-down of dynamics at the tail region, all the models show an increase in the $|S_{\mathrm{CH}}|$ upon addition of cholesterol, reflecting the reduced available volume for the POPC tails.**13.is this a known effect/explanation?**

The change observed here, however, is particularly sensitive to the length of the trajectory as cholesterol-induced increase in effective correlation time is likely to lead to worse convergence of the correlation function within the limited simulation time, and more drastic underestimation of τ_e is expected than for simulations without cholesterol. This will, consequently, cause a tendency towards underestimation on the strength of the cholesterol-driven modulation of the effective correlation time. **14.Should we propose that longer simulations would be good for future work? SAMULI: Only for reasonable models, like CHARMM.**

One of the most popular models used to explain the effect of cholesterol on phopholipid bilayers is the umbrella model[64] where the phopholipid headgroup shields the bulky hydrophophic bodies of cholesterols from water. This leads to even distribution of cholesterol as covering single molecules is energetically less costly compared to cholesterol clusters. Recent study, using one of the most realistic MD models for cholesterol-phospholipid interactions (CHARMM36[13]), suggested that phospholipids cover the cholesterols simply by orienting their headgroups towards the them.[65] However, this effect was only detected up to 0.66 nm distance from a cholesterol molecule. Dynamically, we would expect this preferred orientation to manifest as a slow down in the headgroup C–H bonds. This is, however, detected neither experimentally nor with the most realistic MD models (Fig. 3), as stipulated above, which puts the validity of the umbrella model under question. It is important to note that simulations with Berger and MacRog models do show this slow-down, possibly leading to erroneous conclusions.

To investigate the effect of hydration on the C–H bond dynamics on the PC headgroup, we first present a comparison of experimental effective correlation times obtained from the POPC (measured in full hydration) and DMPC (1,2-dimyristoyl-sn-glycero-3-phosphocholine, measured in low hydration) in Fig. 4a. The values are the same within the experimental accuracy, which leads to two conclusions, 1) the motions of the headgroup bonds are unaffected by the chemical differences in the tails between the two molecules and 2) the decrease of hydration level down to 13 waters per lipid does not considerably alter the correlation times for the headgroup/glycerol region.

Figure 4b presents effective correlation times obtained from three different MD models as function of reducing hydration. All the force fields produce τ_es that are relatively unaffected by the hydration level above 15 waters per lipid (W/L), in line with the experimental observation. When the hydration is further reduced, the dynamics slows down. The effect is weakest with CHARMM36, while more pronounced increase in τ_e is observed with MacRog and Berger force fields. At these same levels of hydration, a change the lipid headgroup order parameters is detected,[13] owing to the tilt of the headgroup towards the membrane plane under low hydration conditions.[?] Deuterium NMR relaxation time (T_1) measurements from DOPC bilayers[66] have revealed a slow-down of the headgroup conformational dynamics below ≈10 W/L, which was attributed to the reduction in available volume for the tilted headgroup. This slow-down is in qualitatively in line with the increased τ_e observed here from the MD models.

15.SAMULI: Maybe we could have a quantitative discussion about how much dehydration slows down the dynamics and how this could affect, e.g., to membrane fusion.

16.SAMULI: It would be interesting to compare the acyl chains with dehydration to the ones with cholesterol.

In the tail region dehydration also causes a consistent decrease in τ_e when using the Berger and MacRog models (data not shown). This change is accompanied by an increase in the absolute value of the tail order parameters may serve as the early indicators of structural transition in the bilayer upon dehydration in the MD models. Experimentally, the gel-to-liquid crystalline phase transition occurs around 3 W/L at room temperature.[67] **17.validity of "structural transition" -statement**

The observed general slow down is of significance when studying not only single bilayers under low hydration but also intermembrane interactions, such as fusion which naturally occur

at dehydrated conditions. Slower dynamics imply that longer simulation times are needed for equilibration, reliably quantifying the properties of the bilayer, and for observing any dynamic events like the lipid tail flips from one membrane to another in case of the fusion.[?] The same applies to the slow-down of dynamics accompanying the addition of cholesterol. **18.not the smoothest text**

Finally, we study the response of the MD model dynamics to increasing amounts of monovalent salt. Experimentally, the modulation of α and β carbon order parameters upon increasing ion concentration have been used to quantify ion binding to lipid bilayers (the molecular electrometer[26,69]). The order parameters are constant for POPC bilayers under NaCl addition in experiments, indicating negligible ion binding. Based on this, we anticipate the effective correlation times also to be unaffected by monovalent salt, however, to our knowledge no experimental measurements have been conducted to quantify this.

The molecular electrometer has been used to show that most molecular dynamics force fields overestimate the binding of monovalent ions to PC bilayers:[26] In the simulations the modulation of the α and β carbon order parameters by increasing NaCl concentration was overestimated compared to the experiments, and accompanied by accumulation of ions at the bilayer surface. In Fig. 5 we compare three force fields, one that is known to exhibit overbinding[26] (MacRog) and two producing more realistic binding affinity (Slipids and CHARMM36). The accumulation of Na^+ ions near the bilayer is quantified in Fig. 5a whereas Fig. 5b shows the chance in τ_e for increasing salt concentration. Ion accumulation results in a slow down in the effective correlation time that is somewhat proportional to the strength of ion binding. Correlation times extracted from the CHARMM36 model vary only a little (low ion binding) when ion concentration is increased, whereas a slightly more pronounced change is observed with Slipids, and MacRog exhibits a clear slow-down (significant ion binding). This indicates that, similarly to the order parameters, τ_e may be useful in investigating the ion

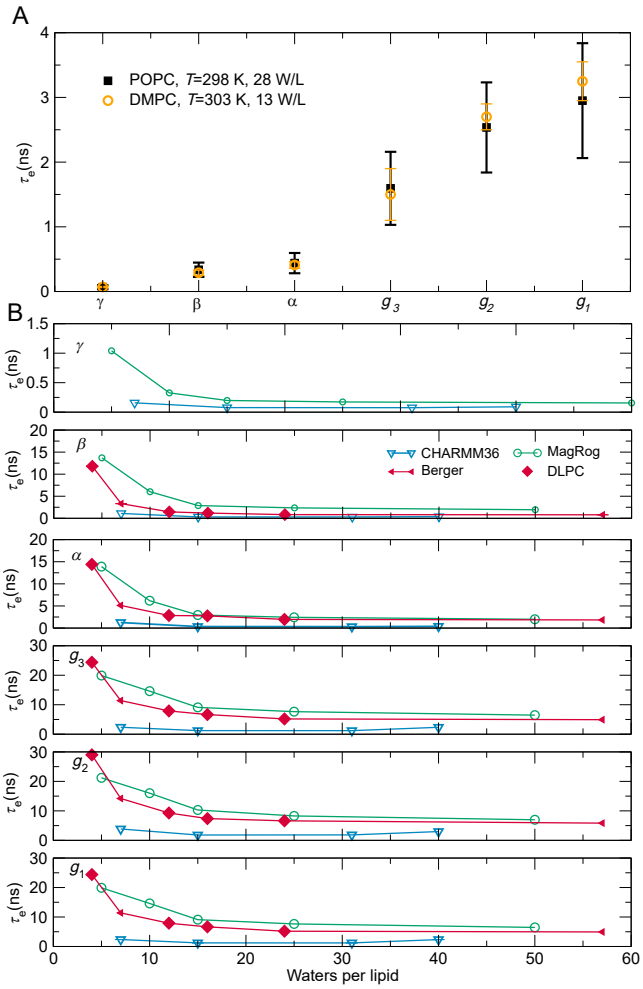

Figure 4: The effect of hydration on the effective correlation times. a) comparison of experimental effective correlation times from DMPC[68] in low hydration (13 W/L) and POPC in full hydration (28 W/L). Neither the different chemistry of the lipid tails, nor the hydration level has an effect within the experimental accuracy. b) The response of effective correlation times to changing hydration level from three MD models. The error bars give the minimum and maximum value observed at each carbon while the symbol denotes the average. Details on the simulations are given in Table 3. Note that three of the data points for the Berger model are from 1,2-didodecanoyl-sn-glycero-3-phosphocholine (DLPC) bilayers (diamonds).

19.how to refer to POPC data

binding affinity of lipid bilayers and experimental work exploring this avenue would be interesting.

20.validity of statement regarding Slipids

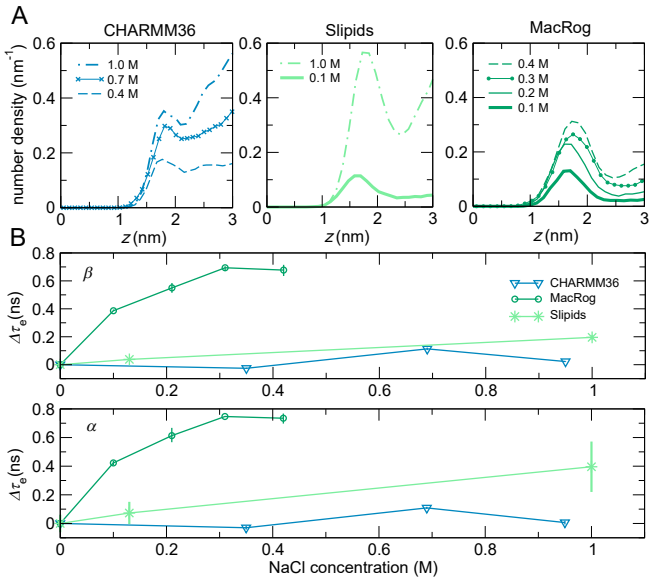

Figure 5: The impact of increasing ionic strength on effective correlation times. a) The density distribution (average over both leaflets) of Na^+ ions as function of distance z from the bilayer center. The plots for each force field are organized from left to right in the order of increasing ion accumulation. b) Effective correlation times for α and β C–H bonds in growing NaCl concentration from CHARMM36, Slipids, MacRog POPC simulations. Details on the simulation data are provided in Table 4.

5 Conclusions

Here, we have investigated the dynamics of phosphatidylcholine molecular dynamics models using publicly available MD trajectories. The MD models are able to qualitatively capture the correlation time profile of POPC—the slow glycerol backbone and the faster dynamics of the headgroup and tail regions— but most are prone to too slow dynamics of the glycerol C–H bonds. In general, the these force fields reproduce the experimentally detected R_1 values adequately, indicating that processes at time scales ∼1 ns are represented but problems arise at longer time-scales. While none of the force fields is able to reproduce all the experimental values, the CHARMM36 POPC model performs well when compared to the effective correlation times, while the Slipids and Lipid14 force field provide realistic R_1 in the PC headgroup and glycerol regions. However, since none of the current MD models reproduce the experimental order parameters, these timescales depict a sampling of a conformational space that does not fully represent the underlying reality.

In addition to the bilayers under standard conditions, we also explored how the dynamics react to the addition of cholesterol, salt, and to the reduction of hydration level. When cholesterol is mixed into the POPC bilayer, the conformational dynamics of the tails and the glycerol regions slows down. Again, the MD models are able to qualitatively capture this, but some also predict an increase in the correlation times for the headgroup carbons, possibly leading to erroneous conclusions. In increasing salt concentration a behaviour reminiscent of the molecular electrometer was observed: Amount of ion binding to the bilayer correlated with the magnitude of slowdown in the correlation times. This could open up the possibility of using effective correlation times in quantifying the ion binding to lipid bilayers. When reducing the water content, the MD models exhibited somewhat constant correlation times down to ∼15 waters per lipid in agreement with experimental data. After this, a slow down was observed.**21.hydration needs some kind of statement of significance.**

By gathering a set of experimental information on the phosphatidylcholine dynamics and underlining some of the typical features of the MD models, this work sets a foundation and a possible roadmap for further improvement of the current force fields. While work is still needed even in capturing the even correct order parameters, the dynamics is equally essential part of developing MD into a true computational microscope; after all, it is possible to obtain the correct order parameters just by freezing the system into a set of selected conformations.**22.not very smoothly put, help!**

Finally, this work demonstrates the power of open data in creating new knowledge out of existing trajectories at a reduced computational and labor cost. Although no new simulations were performed for the purpose of this work, we were able to conduct a comprehensive study on the dynamics of MD models under several conditions. An interesting extension would

be exploring other lipid headgroups individually as well as performing a comparison of MD model dynamics between headgroup types, as the available simulation data goes well beyond simulations of lipids with the phosphocholine headgroup. If the data are well indexed and documented, this process could be easily automated and has the potential to facilitate faster progress, eg., in the development of lipid (and other) MD models. Naturally, such database would provide a fruitful platform to other machine learning applications as well.

Acknowledgement

This material is based upon work supported by XXX under Grant No. XXX. The project is/isn't part of the NMRlipids open collaboration (nmrlipids.blogspot.com)

References

(1) Rossmann, M. G.; Blow, D. M. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica* **1962**, *15*, 24–31.

(2) Burley, S. K.; Berman, H. M.; Christie, C.; Duarte, J. M.; Feng, Z.; Westbrook, J.; Young, J.; Zardecki, C. RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Science* **2018**, *27*, 316–330.

(3) Kirchmair, J.; Markt, P.; Distinto, S.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Langer, T.; Wolber, G. The Protein Data Bank (PDB), Its Related Services and Software Tools as Key Components for In Silico Guided Drug Discovery. *Journal of Medicinal Chemistry* **2008**, *51*, 7021–7040.

(4) Huang, P.-S.; Boyken, S. E.; Baker, D. The coming of age of de novo protein design. *Nature* **2016**, *537*, 320.

(5) Hobohm, U.; Scharf, M.; Schneider, R.; Sander, C. Selection of representative protein data sets. *Protein Science* **1992**, *1*, 409–417.

(6) Levitt, M. Growth of novel protein structural data. *Proceedings of the National Academy of Sciences* **2007**, *104*, 3183–3188.

(7) Mészáros, B.; Dosztányi, Z.; Fichó, E.; Magyar, C.; Simon, I. In *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes: From Bioinformatics to Molecular Quantum Mechanics*; Liwo, A., Ed.; Springer International Publishing: Cham, 2019; pp 561–596.

(8) Sercombe, L.; Veerati, T.; Moheimani, F.; Wu, S. Y.; Sood, A. K.; Hua, S. Advances and Challenges of Liposome Assisted Drug Delivery. *Frontiers in Pharmacology* **2015**, *6*, 286.

(9) Chau, P.-L.; Hoang, P. N.; Picaud, S.; Jedlovszky, P. A possible mechanism for pressure reversal of general anaesthetics from molecular simulations. *Chemical Physics Letters* **2007**, *438*, 294 – 297.

(10) Ferreira, T. M.; Coreta-Gomes, F.; Ollila, O. H. S.; Moreno, M. J.; Vaz, W. L. C.; Topgaard, D. Cholesterol and POPC segmental order parameters in lipid membranes: solid state 1H13C NMR and MD simulation studies. *Phys. Chem. Chem. Phys.* **2013**, *15*, 1976–1989.

(11) Lindahl, E.; Sansom, M. S. Membrane proteins: molecular dynamics simulations. *Current Opinion in Structural Biology* **2008**, *18*, 425 – 431, Membranes / Engineering and design.

(12) Lyubartsev, A. P.; Rabinovich, A. L. Recent development in computer simulations of lipid bilayers. *Soft Matter* **2011**, *7*, 25–39.

(13) Botan, A.; Favela-Rosales, F.; Fuchs, P. F. J.; Javanainen, M.; Kandu, M.;

Kulig, W.; Lamberg, A.; Loison, C.; Lyubartsev, A.; Miettinen, M. S. et al. Toward Atomistic Resolution Structure of Phosphatidylcholine Headgroup and Glycerol Backbone at Different Ambient Conditions. *The Journal of Physical Chemistry B* **2015**, *119*, 15075–15088, PMID: 26509669.

(14) Ferreira, T. M.; Ollila, O. H. S.; Pigliapochi, R.; Dabkowska, A. P.; Topgaard, D. Model-free estimation of the effective correlation time for CH bond reorientation in amphiphilic bilayers: 1H13C solid-state NMR and MD simulations. *The Journal of Chemical Physics* **2015**, *142*, 044905.

(15) Miettinen, M. S.; Lipowsky, R. Bilayer membranes with frequent flip-flops have tensionless leaflets. *Nano letters* **2019**, *?*, ?–?

(16) Vogel, A.; Feller, S. E. Headgroup Conformations of Phospholipids from Molecular Dynamics Simulation: Sampling Challenges and Comparison to Experiment. *The Journal of Membrane Biology* **2012**, *245*, 23–28.

(17) Chernomordik, L. V.; Kozlov, M. M. Mechanics of membrane fusion. *Nature structural & molecular biology* **2008**, *15*, 675.

(18) Gibson, N. J.; Brown, M. F. Lipid headgroup and acyl chain composition modulate the MI-MII equilibrium of rhodopsin in recombinant membranes. *Biochemistry* **1993**, *32*, 2438–2454, PMID: 8443184.

(19) Phillips, R.; Ursell, T.; Wiggins, P.; Sens, P. Emerging roles for lipids in shaping membrane-protein function. *Nature* **2009**, *459*, 379.

(20) Feller, S. E.; Gawrisch, K.; MacKerell, A. D. Polyunsaturated Fatty Acids in Lipid Bilayers: Intrinsic and Environmental Contributions to Their Unique Physical Properties. *Journal of the American Chemical Society* **2002**, *124*, 318–326, PMID: 11782184.

(21) Eldho, N. V.; Feller, S. E.; Tristram-Nagle, S.; Polozov, I. V.; Gawrisch, K. Polyunsaturated Docosahexaenoic vs Docosapentaenoic AcidDifferences in Lipid Matrix Properties from the Loss of One Double Bond. *Journal of the American Chemical Society* **2003**, *125*, 6409–6421, PMID: 12785780.

(22) Wohlert, J.; Edholm, O. Dynamics in atomistic simulations of phospholipid membranes: Nuclear magnetic resonance relaxation rates and lateral diffusion. *The Journal of Chemical Physics* **2006**, *125*, 204703.

(23) Klauda, J. B.; Roberts, M. F.; Redfield, A. G.; Brooks, B. R.; Pastor, R. W. Rotation of Lipids in Membranes: Molecular Dynamics Simulation, 31P Spin-Lattice Relaxation, and Rigid-Body Dynamics. *Biophysical Journal* **2008**, *94*, 3074–3083.

(24) Leftin, A.; Brown, M. F. An NMR database for simulations of membrane dynamics. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2011**, *1808*, 818 – 839, Including the Special Section: Protein translocation across or insertion into membranes.

(25) Klauda, J. B.; Eldho, N. V.; Gawrisch, K.; Brooks, B. R.; Pastor, R. W. Collective and Noncollective Models of NMR Relaxation in Lipid Vesicles and Multilayers. *The Journal of Physical Chemistry B* **2008**, *112*, 5924–5929, PMID: 18179193.

(26) Catte, A.; Girych, M.; Javanainen, M.; Loison, C.; Melcr, J.; Miettinen, M. S.; Monticelli, L.; Mtt, J.; Oganesyan, V. S.; Ollila, O. H. S. et al. Molecular electrometer and binding of cations to phospholipid bilayers. *Phys. Chem. Chem. Phys.* **2016**, *18*, 32560–32569.

(27) Ollila, S.; Hyvönen, M. T.; Vattulainen, I. Polyunsaturation in Lipid Membranes: Dynamic Properties and Lateral Pressure

Profiles. *J. Phys. Chem. B* **2007**, *111*, 3139–3150.

(28) Ollila, O. H. S.; Ferreira, T.; Topgaard, D. MD simulation trajectory and related files for POPC bilayer (Berger model delivered by Tieleman, Gromacs 4.5). 2014; {http://dx.doi.org/10.5281/zenodo.13279}.

(29) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; Jr, A. D. M.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.

(30) Santuz, H. MD simulation trajectory and related files for POPC bilayer (CHARMM36, Gromacs 4.5). 2015; http://dx.doi.org/10.5281/zenodo.14066, DOI: 10.5281/zenodo.14066.

(31) Antila, H. . 2018; http://dx.doi.org/10.5281/zenodo.148560, DOI: 10.5281/zenodo.1468560.

(32) Kulig, W.; Jurkiewicz, P.; Olżyńska, A.; Tynkkynen, J.; Javanainen, M.; Manna, M.; Rog, T.; Hof, M.; Vattulainen, I.; Jungwirth, P. Experimental determination and computational interpretation of biophysical properties of lipid bilayers enriched by cholesteryl hemisuccinate. *Biochim. Biophys. Acta* **2015**, *1848*, 422 – 432.

(33) Javanainen, M. POPC/Cholesterol @ 310K. 0, 10, 40, 50 and 60 mol-cholesterol. Model by Maciejewski and Rog. **2015**,

(34) Dickson, C. J.; Madej, B. D.; Skjevik, A. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 865–879.

(35) Ollila, O. H. S.; Retegan, M. MD simulation trajectory and related files for POPC bilayer (Lipid14, Gromacs 4.5). 2014; DOI: 10.5281/zenodo.12767.

(36) Jämbeck, J. P. M.; Lyubartsev, A. P. An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes. *J. Chem. Theory Comput.* **2012**, *8*, 2938–2948.

(37) Javanainen, M. POPC with 0, 10, 20, and 30 mol-Slipids force field. 2016; http://dx.doi.org/10.5281/zenodo.3243328.

(38) Melcr, J.; Martinez-Seara, H.; Nencini, R.; Kolafa, J.; Jungwirth, P.; Ollila, O. H. S. Accurate Binding of Sodium and Calcium to a POPC Bilayer by Effective Inclusion of Electronic Polarization. *The Journal of Physical Chemistry B* **2018**, *122*, 4546–4557.

(39) Melcr, J. Simulations of POPC lipid bilayer in water solution at various NaCl, KCl and CaCl2 concentrations using ECC-POPC force field. **2019**,

(40) Höltje, M.; Förster, T.; Brandt, B.; Engels, T.; von Rybinski, W.; Höltje, H.-D. Molecular dynamics simulations of stratum corneum lipid models: fatty acids and cholesterol. *Biochim. Biophys. Acta* **2001**, *1511*, 156 – 167.

(41) Ollila, O. H. S. MD simulation trajectory and related files for POPC/cholesterol (50 molmodified Hltje, Gromacs 4.5). **2014**,

(42) Lim, J. B.; Rogaski, B.; Klauda, J. B. Update of the Cholesterol Force Field Parameters in CHARMM. *J. Phys. Chem. B* **2012**, *116*, 203–210.

(43) Santuz, H. MD simulation trajectory for POPC/50% Chol bilayer (CHARMM36, Gromacs 4.5). 2015; http://dx.doi.org/10.5281/zenodo.14068, DOI: 10.5281/zenodo.14068.

(44) Jämbeck, J. P. M.; Lyubartsev, A. P. Another Piece of the Membrane Puzzle: Extending Slipids Further. *Journal of Chemical Theory and Computation* **2013**, *9*, 774–784, PMID: 26589070.

(45) Ollila, O. H. S. MD simulation trajectory and related files for POPC bilayer in low hydration (Berger model delivered by Tieleman, Gromacs 4.5). **2015**,

(46) Kanduc, M.; Schneck, E.; Netz, R. R. Hydration Interaction between Phospholipid Membranes: Insight into Different Measurement Ensembles from Atomistic Molecular Dynamics Simulations. *Langmuir* **2013**, *29*, 9126–9137.

(47) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=24 w/l. 2015; DOI: 10.5281/zenodo.16289.

(48) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=16 w/l. 2015; DOI: 10.5281/zenodo.16292.

(49) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=12 w/l. 2015; DOI: 10.5281/zenodo.16293.

(50) Ollila, O. H. S.; Miettinen, M. MD simulation trajectory and related files for POPC bilayer in medium low hydration (CHARMM36, Gromacs 4.5). 2015; {http://dx.doi.org/10.5281/zenodo.13946}, DOI: 10.5281/zenodo.13946.

(51) Ollila, O. H. S.; Miettinen, M. MD simulation trajectory and related files for POPC bilayer in low hydration (CHARMM36, Gromacs 4.5). 2015; {http://dx.doi.org/10.5281/zenodo.13945}, DOI: 10.5281/zenodo.13945.

(52) Javanainen, M. POPC @ 310K, varying water-to-lipid ratio. Model by Maciejewski and Rog. 2014; {http://dx.doi.org/10.5281/zenodo.13498}, DOI: 10.5281/zenodo.13498.

(53) Venable, R. M.; Luo, Y.; Gawrisch, K.; Roux, B.; Pastor, R. W. Simulations of Anionic Lipid Membranes: Development of Interaction-Specific Ion Parameters and Validation Using NMR Data. *J. Phys. Chem. B* **2013**, *117*, 10183–10192.

(54) Ollila, O. H. S. MD simulation trajectory and related files for POPC bilayer with 350mM NaCl (CHARMM36, Gromacs 4.5). 2015; http://dx.doi.org/10.5281/zenodo.32496.

(55) Ollila, O. H. S. MD simulation trajectory and related files for POPC bilayer with 690mM NaCl (CHARMM36, Gromacs 4.5). 2015; http://dx.doi.org/10.5281/zenodo.32497.

(56) Ollila, O. H. S. MD simulation trajectory and related files for POPC bilayer with 950mM NaCl (CHARMM36, Gromacs 4.5). 2015; http://dx.doi.org/10.5281/zenodo.32498.

(57) Åqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* **1990**, *94*, 8021–8024.

(58) Javanainen, M.; Tynkkynen, J. POPC @ 310K, varying amounts of NaCl. Model by Maciejewski and Rog. 2015; http://dx.doi.org/10.5281/zenodo.14976.

(59) Smith, D. E.; Dang, L. X. Computer simulations of NaCl association in polarizable water. *J. Chem. Phys* **1994**, *100*, 3757–3766.

(60) Javanainen, M. POPC @ 310K, 130 mM of NaCl. Slipids with ions by Smith & Dang. 2015; http://dx.doi.org/10.5281/zenodo.35275.

(61) Javanainen, M. POPC with varying amounts of cholesterol, 1 M of NaCl. Slipids with ions by Smith & Dang. 2015; http://dx.doi.org/10.5281/zenodo.259341.

(62) Schlenkrich, M.; Brickmann, J.; MacKerell, A. D.; Karplus, M. *Biological Membranes*; Springer, 1996; pp 31–81.

(63) Feller, S. E.; MacKerell, A. D. An improved empirical potential energy function for molecular simulations of phospholipids. *The Journal of Physical Chemistry B* **2000**, *104*, 7510–7515.

(64) Huang, J.; Feigenson, G. W. A Microscopic Interaction Model of Maximum Solubility of Cholesterol in Lipid Bilayers. *Biophysical Journal* **1999**, *76*, 2142 – 2157.

(65) Leeb, F.; Maibaum, L. Spatially Resolving the Condensing Effect of Cholesterol in Lipid Bilayers. *Biophysical Journal* **2018**, *115*, 2179 – 2188.

(66) Ulrich, A.; Watts, A. Molecular response of the lipid headgroup to bilayer hydration monitored by 2H-NMR. *Biophys. J.* **1994**, *66*, 1441 – 1449.

(67) Lynch, D. V.; Steponkus, P. L. Lyotropic phase behavior of unsaturated phosphatidylcholine species: relevance to the mechanism of plasma membrane destabilization and freezing injury. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1989**, *984*, 267 – 272.

(68) Pham, Q. D.; Topgaard, D.; Sparr, E. Cyclic and Linear Monoterpenes in Phospholipid Membranes: Phase Behavior, Bilayer Structure, and Molecular Dynamics. *Langmuir* **2015**, *31*, 11067–11077, PMID: 26375869.

(69) Seelig, J.; MacDonald, P. M.; Scherer, P. G. Phospholipid head groups as sensors of electric charge in membranes. *Biochemistry* **1987**, *26*, 7535–7541, PMID: 3322401.

Graphical TOC Entry

TOC here if needed