Using open data to rapidly benchmark biomolecular simulations: Phospholipid internal dynamics

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, 14424 Potsdam, Germany*

[‡]*NMR Group — Institute for Physics, Martin-Luther University Halle–Wittenberg, 06120 Halle (Saale), Germany*

[¶]*Add Matti to author list?*

[§]*Institute of Biotechnology, University of Helsinki, 00014 Helsinki, Finland*

E-mail: markus.miettinen@iki.fi

Abstract

Molecular dynamics (MD) simulations are widely used to study the atomistic structure and dynamics of biomembranes. It remains unknown, however, how well the conformational dynamics observed in MD simulations correspond to those occurring in real life phospholipids. The accuracy of such time scales in MD can be assessed by comparing against the effective correlation times τ_e of the C-H bonds measured in nuclear magnetic resonance experiments (J. Chem. Phys. 142 044905 (2015)).

Here, we extend this previous analysis by considering carefully the error estimation of MD-determined τ_e, and analyze the conformational dynamics of phospholipids as produced by several commonly used MD models (force fields). None of the tested force fields reproduced all the effective correlation times within experimental error, much like they do not provide accurate conformational ensemble (J. Phys. Chem. B 119 15075 (2015)). However, the dynamics observed in CHARMM36 and Slipids were more realistic than those seen in the Amber Lipid14, OPLS-based MacRog, and GROMOS-based Berger force fields, where dynamics of the glycerol backbone was unrealistically slow.

1 Introduction

Ever since the conception of Protein Data Bank (PDB)[1,2] and GenBank,[3,4] open access to standardised and searchable pools of experimental data has revolutionized research in life sciences. The databanks,[5] constantly growing and improving in fidelity[6–8] due to collaborative effort, enable scientific progress that is well beyond the resources of one single research group, giving rise to entirely new ways of doing science in the form of bio- and cheminformatics, and enabling data-driven development of characterisation techniques,[9] drugs,[10] and materials.[11] ~~In addition to experimental results, the push from funders towards~~ The idea of public availability and conservation of data has recently extended also to molecular dynamics (MD) simulation trajectories of biomolecules, and discussion on how and by whom these databases for dynamic information would be set up is currently active.[12–16] ~~Notably,~~ **~~1.Mention GPCRmd[17]?~~**

~~SAMULI: I think yes.~~

Since 2013 the NRMlipids Project (~~nmrlipids.blogspot.fi~~) ~~has since 2013 accumulated such a databank[18] for~~

nmrlipids.blogspot.fi) has promoted a open collaboration approach, where the whole research process, from the initial ideas and discussions to the analysis methods, data and publications, are publicly available all the time [22]. While the main focus of the NMRlipids project has been in conformational ensembles of lipid headgroups and ion binding to lipid membranes [22,38?], it has also accumulated a collection of atomistic MD simulations of lipid membranes containing hundreds of trajectories (zenodo.org/communities/nmrlipids). ~~Using this freely available resource we demonstrate here, for the first time, the viability of creating new scientific knowledge solely through analysis of pre-existing, open access MD simulation data.~~

~~Our system of interest, lipids, in their biologically relevant state as the core components of the cell's membranes, are intrinsically unstructured. To properly describe such molecules, a whole ensemble of conformations~~ This data is also partially indexed at www.nmrlipids.fi. Such databanks are particularly relevant for disordered molecules, such as biological lipids composing cellular membranes, which cannot be descibed by coordinates of single structure in contrast to folded proteins or DNA strands. Realistic MD simulations can provide conformational ensemble and dynamics of such molecules as well as ~~the dynamics linking them is needed. To obtain such description, MD simulations of lamellar phospholipid bilayer systems are widely used [19? –24], and hold vast potential to decipher, e.g., molecular mechanisms behind anesthetics [20?], the effect of cholesterol on membrane structure [21?], and the functioning of membrane proteins [25].~~ enable studies of their biological functions in complex biomolecules assemblies. However, current MD simulation force fields largely fail to capture conformational ensembles of lipid headgroups and disordered proteins [22? ?]. Therefore, the quality of MD simulations in databanks and other applications must be carefully assessed against experimental data. For lipid bilayers, such evaluation is possible against NMR and scatterign data [?].

~~However, to be truly useful MD should give the right 1) equilibrium statistics and 2) dynamics. To extract reliable statistics, it is crucial to assess that the simulations have converged: The conformations sampled have to represent the equilibrium distribution with enough transitions between states. Indeed, for lipids even 500 ns simulations might be insufficient [23,26].~~ Here, we demonstrate how publicly available set of MD simulation data can be utilized to rapidly evaluate how fast individual lipid molecules sample their conformational ensemble against experimental data in different force fields with unprecedented extent. MD simulations with correct lipid dynamics are desired, for example, for the interpretation of NMR or other experiments detecing molecular dynamics and to understand dynamics of biological processes where lipid deformations have rate limiting role such as membrane fusion [?]. In addition, information on dynamics is crucial to assess if simulations have converged.~~Along with equilibrium statistics, the ability of MD to reproduce the bilayer dynamics is equally crucial for an accurate picture of membrane function. The correct relative abundance of different dynamical processes is needed for reliable interpretation of pathways leading to, e.g., membrane deformation [27] and lipid-induced conformational [28,29] changes of membrane proteins. Notably, the availability of such a realistic MD model could greatly guide both the configuration and the interpretation of NMR experiments used to extract dynamical information from lipid assemblies.~~

~~By analyzing a wide set of publicly available phosphatidylcholine (PC) lipid bilayer MD trajectories, we test whether different MD models (force fields) reproduce the experimentally observed internal dynamics of PC lipids, and investigate if the dynamics of various models share common features. Such features can be used to draw general conclusions on the system, to avoid potential pitfalls in future simulations of bilayers, and to suggest future directions for experimental research. In addition to simulations of one component bilayers under standard conditions,~~

~~we study the effects of varying hydration, cholesterol content, and NaCl concentration.~~

~~We analyze lipid dynamics based on two quantities available from published ^{13}C-NMR experiments: [23,30,31] The effective C–H bond correlation time τ_e and the spin-lattice relaxation rate R_1, both directly quantifiable from atomistic MD simulations. The τ_e are effectively an average over all the time scales relevant for the lipid internal dynamics, and respond intuitively to changes in these: Increasing τ_e always signals some type of slowdown in the C–H bond dynamics. [23] The R_1 rates (or the corresponding T_1 times) have been traditionally used to assess both the conformational dynamics of lipids in experiments [32–36] and the dynamics produced by MD models in simulations [32,34,35,37]. In contrast to τ_e, the R_1 are sensitive to processes within a rather narrow time scale window set by the magnet frequency, and changes in R_1 are not intuitively related to changes in process speeds: A decrease in R_1 tells that the amount of processes in the sensitive time window decreases, but not if the corresponding processes become faster or slower.~~

~~In summary, our work provides first~~ Our 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 pre-existing, publicly available simulation trajectories to~~ between different MD models for phosphatidylcholine lipids with varying biologically relevant compositions and conditions paves the way for the development of more realistic lipid force fields. Furthermore, the analysis of extensive set of data from different models shed light on the complex dynamics lipid in their biological relevant disordered state. Our results demonstrate the power of ~~open, well documented MD data~~ publicly available simulation trajectories in creating new knowledge at a lowered computational cost and high potential for automation. We believe that our work paves the way for novel applications of publicly available MD simulations databanks, as well as demonstrates their usefullness not only for lipid bilayers but also for other biomolecular systems.

2 Methods

3 ~~Theoretical Background~~

2.1 Evaluation of conformational dynamics of lipids against NMR data

~~^{13}C NMR experiments investigating 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~~

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

~~which is the Fourier transformation of the C–H bond~~ In a lipid bilayer in liquid crystalline phase, each individual lipid molecule samples an internal conformational ensemble and rotates around the membrane normal direction. The conformational dynamics of a lipid molecule can be characterized using the second order autocorrelation ~~function at time τ~~ functions of C–H bonds

$$g(\tau) = \langle P_2\left(\vec{\mu}(t)\cdot\vec{\mu}(t+\tau)\right)\rangle, \qquad (1)$$

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~~ For lipids in a bilayer, the internal dynamics and rotation around membrane normal have timescales below μs, leading to the decay of the correlation function to a non-zero plateau value (Fig. 1). This plateau is the square of the C-H bond

order parameter, S^2_{CH}.

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

where $\theta(t)$ is the angle between the bond and the bilayer normal. This order parameter can be measured using dipolar coupling in ^{13}C NMR or quadrupolar coupling in ^{2}H NMR, and is highly useful in order to evaluate conformational ensembles of lipids [?].

In order to analyze the internal dynamics of lipids, the C-H bond autocorrelation function is often expressed as

$$g(\tau) = g_f(\tau)g_s(\tau), \quad (3)$$

where $g_f(\tau)$ ~~characterizes fast decays owing to, for example, the molecular rotations,~~ represents the fast dynamics below $\sim \mu$s timescales, and $g_s(\tau)$ describes slow ~~decays that originate~~ decay from, e.g., lipid diffusion ~~. The two components, along with the oscillation due to~~ between lamellae with different orientations and magic angle spinning ~~at the ~kHz region, are depicted in~~ in solid state NMR experiments (Fig. 1~~. Correlation time of 4.2 ms has been estimated for~~). In multilamellar POPC samples at 300 K ~~for the slow modes, whereas~~ in liquid crystalline ~~lipid bilayers the faster~~ state, the correlation time of 4.2 ms was estimated for $g_s(\tau)$, whereas $g_f(\tau)$ decays to a plateau value ~~S^2_{CH}~~ within a few hundred nanoseconds.[23] ~~The C-H bond order parameters~~

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

~~where $\theta(t)$ is the angle between the 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~~ The internal dynamics containing multiple timescales can be quantified ~~via~~ using the effective correlation time

$$\tau_e = \int_0^\infty \frac{g_f(\tau) - S^2_{CH}}{1 - S^2_{CH}} d\tau, \quad (4)$$

~~The integrand can be viewed as a reduced and normalized correlation function~~

$$g'_f(\tau) = \frac{g_f(\tau) - S^2_{CH}}{1 - S^2_{CH}}.$$

~~That is, τ_e is defined as the area under $g'_f(\tau)$, as graphically depicted~~ which is related to the integrated area below the correlation function in Fig. 1 ~~b.~~ B. The integrand is Eq. 4 defines the reduced and normalized correlation function

$$g'_f(\tau) = \frac{g_f(\tau) - S^2_{CH}}{1 - S^2_{CH}}. \quad (5)$$

2.Maybe also add 1C that explicitly shows g'_f? ~~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 ~~spin-lattice relaxation rate R_1 defines the time-scale on which~~ internal dynamics of lipids in MD simulations has been previously evaluated by comparing the ^{2}H or ^{13}C ~~longitudal magnetization equilibrates. It is defined as~~ spin relaxation times or effective correlation times calculated from simulation trajectory with the experimental data [23,32,34,35,37?]. However, lipids exhibit complex internal dynamics with multiple timescales that cannot be fully captured with a single parameter. Therefore, several experimental parameters, detected for example with different magnetic fields or temperatures, are required to evaluate dynamics in simulations or to interpret dynamics from experiments [?].

Here, we use use two parameters: the effective correlation times, τ_e, and R_1 spin relaxation rates from ^{13}C-NMR experiments [23,30,31]. The effective correlation times detect essentially an average over the time scales relevant for the lipid internal dynamics, and have intuitive relation to dynamics as larger values always

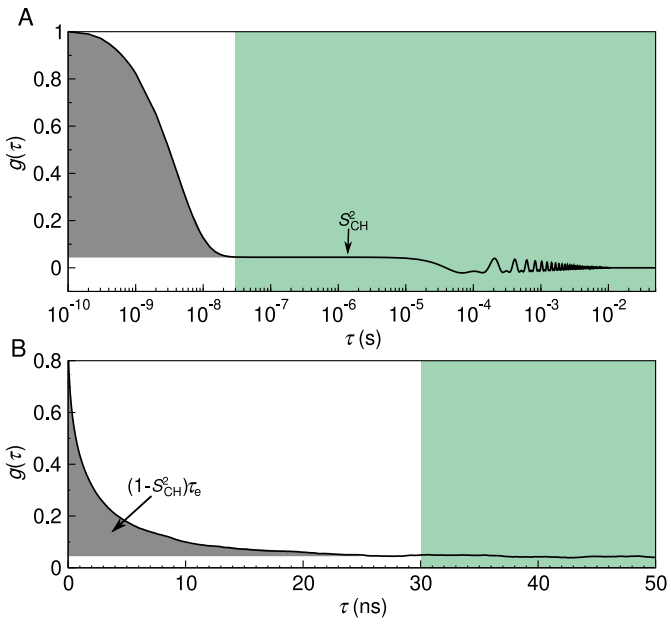

Figure 1: The autocorrelation function $g(\tau)$ a) The fast mode (white background) and the slow mode (shaded green) of the correlation function along with the oscillation owing to magic angle spinning. The fast mode decays to a plateau quantifying the S_{CH} while the slow mode gives the final descent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

indicate slower dynamics. On the other hand, R_1 values are most sensitive to processes with the timescales around ~1–10 ns with typical magnetic field strengths. R_1 is related to molecular dynamics through equation

$$R_1 = \frac{d_{CH}^2 N_H}{20}\left[j(\omega_H - \omega_C) + 3j(\omega_C) + 6j(\omega_H + \omega_C)\right], \quad (6)$$

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, $d_{CH}/2\pi$ approximately equals to -22 kHz. **3.why there is a minus sign above?**

~~The dependency of R_1 on the spectral densities j at the Larmor frequencies means that the R_1 value depicts the relative amounts of relaxation processes.~~ The spectral density is the Fourier transformation of the rotational correlation function

$$j(\omega) = 2\int_0^\infty \cos(\omega\tau) g(\tau) d\tau. \quad (7)$$

The connection between R_1 and molecular dynamics is not straightforward, but magnitude of R_1 reflects the relative significance of the 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~~ inverse of proton and ^{13}C Larmor frequencies (typically ~1–10 ns~~time-scales. Importantly, a change in R_1 thus 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.~~).

2.2 Experimental data acquisition and analysis

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?**.

2.3 Simulational data acquisition 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[22,38] (`nmrlipids.blogspot.fi`). **6.SAMULI: I think that we should define the criteria for selecting the analyzed simulations here.** Table 1 details, with references to the trajectory files, the simulations of pure POPC bilayers at/near room temperature and at full hydration, whereas Table 2 lists simulations including cholesterol; Table 3 simulations with varying hydration; and Table 4 at increasing NaCl concentration. Additional computational details for each of the simulations are available at the cited Zenodo entry.

Table 1: Analyzed simulations of POPC lipid bilayers at standard conditions.

force field	N_l[a]	N_w[b]	T[c](K)	t_{anal}[d](ns)	files[e]
Berger-POPC-07[39]	128	7290	298	50	[40]
CHARMM36[41]	128	5120	303	140	[42]
	34	1020	300	140	[43]
MacRog[44]	128	6400	310	200	[45]
Lipid14[46]	72	2234	303	50	[47]
Slipids[48]	200	9000	310	500	[49]
ECC[50]	128	6400	300	300	[51]

[a]Number of POPC molecules.
[b]Number of water molecules.
[c]Simulation temperature.
[d]Trajectory length used for analysis.
[e]Reference for the openly available simulation files.

Table 2: Analyzed simulations of cholesterol-containing POPC bilayers.

force field POPC/cholesterol	c_{chol}[a]	N_{chol}[b]	N_l[c]	N_w[d]	T[e](K)	t_{anal}[f](ns)	files[g]
Berger-POPC-07[39]	0%	0	128	7290	298	50	[40]
/Höltje-CHOL-13[21,52]	50%	64	64	10314	298	60	[53]
CHARMM36[41]	0%	0	128	5120	303	140	[42]
/CHARMM36[54]	50%	80	80	4496	303	200	[55]
MacRog[44]	0%	0	128	6400	310	200	[45]
/MacRog[44]	50%	64	64	6400	310	200	[45]
Slipids[48]	0%	0	200	9000	310	500	[49]
/Slipids[56]	50%	200	200	18000	310	500	[49]

[a]Bilayer cholesterol content (mol %).
[b]Number of cholesterol molecules.
[c]Number of POPC molecules.
[d]Number of water molecules.
[e]Simulation temperature.
[f]Trajectory length used for analysis.
[g]Reference for the openly available simulation files.

9.SAMULI: I would emphasize here more the automatization of the analysis, and also mention that the mapping file was used to enable automatic analysis of simulations with different force fields having different atom names (if this was the case?). The simulation data were analyzed using in-house scripts. These are available on GitHub[?] along with a Python notebook outlining an example analysis run. After downloading the necessary files from Zenodo, the trajectory was processed with Gromacs `gmx trjconv` to make the molecules whole. The C–H bond order parameters S_{CH}, see Eq. (2), were then calculated with the `calcOrderParameters.py`[?] script that uses the MDanalysis[?] Python library. The C–H bond correlation functions $g(\tau)$, see Eq. (1), were calculated with Gro-

Table 3: Analyzed simulations of lipid bilayers under varying hydration level.

force field	lipid	$n_{w/l}$[a]	N_l[b]	N_w[c]	T[d](K)	t_{anal}[e](ns)	files[f]
Berger-POPC-07[39]	POPC	57	128	7290	298	50	[40]
Berger-DLPC-13[57]	DLPC[g]	24	72	1728	300	80	[58]
	DLPC[g]	16	72	1152	300	80	[59]
	DLPC[g]	12	72	864	300	80	[60]
Berger-POPC-07[39]	POPC	7	128	896	298	60	[61]
Berger-DLPC-13[57]	DLPC[g]	4	72	288	300	80	[62]
CHARMM36[41]	POPC	40	128	5120	303	140	[42]
	POPC	15	72	1080	303	20	[63]
	POPC	7	72	504	303	20	[64]
MacRog[44]	POPC	50	288	14400	310	40	[65]
	POPC	15	288	4320	310	100	[65]
	POPC	10	288	2880	310	100	[65]

[a]Water/lipid molar ratio.
[b]Number of lipid molecules.
[c]Number of water molecules.
[d]Simulation temperature.
[e]Trajectory length used for analysis.
[f]Reference for the openly available simulation files.
[g]1,2-didodecanoyl-sn-glycero-3-phosphocholine.

7. The data points here do not match those in Fig. 5B. MacRog in Fig. 5B: 50, 25, 10, 5 w/l, and C36 in Fig. 5B: 40, 31, 15, 7 w/l.

8.The t_{anal} for MacRog here do not match Ref. 65 (100 ns → 50 ns)?

Table 4: Analyzed simulations of POPC lipid bilayers at varying NaCl concentration.

force field POPC/ions	[NaCl][a] (mM)	N_{Na}[b]	N_l[c]	N_w[d]	T[e](K)	t_{anal}[f](ns)	files[g]
CHARMM36[41]	0	0	128	5120	303	140	[42]
/CHARMM36[66]	346	13	72	2085	303	80	[67]
	692	26	72	2085	303	73	[68]
	947	37	72	2168	303	60	[69]
MacRog[44]	0	0	128	6400	310	400	[45]
/OPLS[70]	103	27	288	14554	310	90	[71]
	207	54	288	14500	310	90	[71]
	311	81	288	14446	310	80	[71]
	416	108	288	14392	310	90	[71]
Slipids[48]	0	0	200	9000	310	500	[49]
/AMBER[72]	130	21	200	9000	310	100	[73]
	999	162	200	9000	310	200	[74]

[a]NaCl concentration, calculated as [NaCl]= $N_{Na}\times$[water]/N_w, where [water] = 55.5 M.
[b]Number of Na^+ ions, equal to number of Cl^- ions.
[c]Number of POPC molecules.
[d]Number of water molecules.
[e]Simulation temperature.
[f]Trajectory length used for analysis.
[g]Reference for the openly available simulation files.

macs5.1.4? `gmx rotacf`; **10.SAMULI: I do not understand this:** note that on simulational (fast) time scales $g = g_s g_f = g_f$. To obtain the g'_f, the S_{CH} were used to normalize the g_f following Eq. (5).

The effective correlation times τ_e were then calculated ~~by integrating $g'_f(\tau)$, see Eqs~~from Eq. (4) ~~and , over time~~ integrating from $\tau = 0$ until $\tau = t_0$~~. Here $t_0 = \min\{t \,|\, g'(t) = 0\}$, that is, t_0~~ , where $t_0 = \min\{t \,|\, g'(t) = 0\}$ is the first time point at which g'_f reached zero. If g'_f did not reach zero within $t_{anal}/2$, the τ_e was not determined, and we report only its upper and lower error estimates.

To quantify the error on τ_e, we first estimate the error on $g'_f(\tau)$,where we account for two sources of uncertainty, $g_f(\tau)$ and S^2_{CH}. Performing linear error propagation on Eq. (5) gives

$$\Delta g'_f(\tau) = \left|\frac{1}{1 - S^2_{CH}}\right| \Delta g_f(\tau) + \left|\frac{2\,(g_f(\tau) - 1)\, S_{CH}}{(1 - S^2_{CH})^2}\right| \Delta S_{CH}. \quad (8)$$

Here the ΔS_{CH} was determined ~~as in the NMR-lipids Project [22]:~~ as the standard error of the mean of the S_{CH} ~~of all the~~ when averaged over N_l individual lipids in the system [22]. Similarly, we quantified the error on $g_f(\tau)$ by first determining ~~an individual correlation function~~the correlation function, $g^m_f(\tau)$~~for each~~ , for each individual lipid m~~over the whole trajectory, ~~, and then obtaining the error estimate $\Delta g_f(\tau)$ as the standard error of the mean over the N_l lipids ~~. Importantly, this gives an uncertainty estimate at~~ for each time point τ.

To obtain the lower bound on τ_e, we integrate the function $g'_f(\tau) - \Delta g'_f(\tau)$ over time from $\tau = 0$ until $\tau = t_l$. Here

$$t_l = \min\left\{\{t \,|\, g'_f(t) - \Delta g'_f(t) = 0\}, \frac{t_{anal}}{2}\right\}. \quad (9)$$

That is, t_l equals the first time point at which the lower error estimate of g'_f reached zero; or $t_l = t_{anal}/2$, if zero was not reached by that point.

To obtain the upper error estimate on τ_e, we first integrate the function $g'_f(\tau) + \Delta g'_f(\tau)$ over time from $\tau = 0$ until $t_u = \min\{t_0, t_{anal}/2\}$. Note, however, that this is not yet sufficient, because there could be slow processes that our simulation was not able to see. Although these would contribute to τ_e with a low weight, their contribution over long times could still add up to a sizable effect on τ_e. That said, it seems feasible **11.SAMULI: I think that we can say stronger than "seems feasible to assume" because experiments in Tiago's 2015 paper indicate that these, or maybe even shorter, timescales are not present. Maybe Tiago can comment this?** to assume (see Fig. 1A) that there are no longer-time contributions to g_f than something that decays with a time constant of 10^{-6} s. We use this as our worst case estimate to assess the upper bound for τ_e, and assume that all the decay from the time point $t_u = \min\{t_0, t_{anal}/2\}$ onwards comes solely from this slowest process. **12.SAMULI: I do not understand this:** The additional contribution to the upper bound for τ_e then reads $\Delta g'_f(t_u) \times (\exp(-t_u/10^{-6}\,\text{s}) - \exp(-1)) \times 10^{-6}\,\text{s}$. **13.Discuss the possibility of skewed error distributions?**

The R_1 rates were calculated using Eq. (6). The spectral density $j(\omega)$ was obtained from the normalized correlation function g'_f by fitting it with a sum of $N = 71$ exponentials

$$g'_f(\tau) \approx \sum_{i=1}^{N} \alpha_i e^{-\tau/\tau_i}, \quad (10)$$

with logarithmically spaced time-scales τ_i ranging from 0.1 ps to 1 μs, and then calculating the spectral density of this fit based on the Fourier transformation [23]

$$j(\omega) = 2(1 - S_{CH}) \sum_{i=1}^{N} \alpha_i \frac{\tau_i}{1 + \omega\tau_i}. \quad (11)$$

14.SAMULI: I think that below text may sound weird for NMR people. I think that bonds do not have R_1 rates, nucleus has. I think that this should be rewritten in the way that we talk about 13C spin relaxation from C-H coupling in the way that it still looks like that we understand that this is actually a spin system. This is not easy, and right now I cannot figure out an easy way to it. I can think this later if needed. The R_1 rate of a given C–H bond was first calculated

separately for each lipid m (using Eq. (6) with $N_H = 1$, and $j^m(\omega)$ obtained for the normalized correlation function $g_f'^m$). The resulting N_l measurements per bond were then assumed independent: Their mean gave the R_1 rate of the bond, and standard error of the mean its uncertainty. The total R_1 rate of a given carbon was obtained as a sum of the R_1 rates of its C–H bonds. **15.SAMULI: When does this happen:** When several carbons contribute to ~~the experimental R_1 rate of a carbon segment, the carbon-wise R_1 rates were averaged to obtain the segment-wise R_1 rate.~~ a single experimental R1 rate due to the overlapping peaks (for example in C_2 carbon in acyl chains and γ carbons), the R1 from simulations was averaged over carbons with overlapping peaks. The segment-wise error estimates were obtained by standard error propagation, starting from the uncertainties of the R_1 rates of the C–H bonds.

To gain ~~some qualitative~~ insight on the time scales at which the main contributions to the ~~(headgroup)~~ R_1 rates arise, we ~~also looked at~~ calculated 'cumulative' R_1 rates, $R_1(\tau)$~~. These contained just those contributions in the sum of~~ , which contained the contributions of the sum in Eq. (11) for which $\tau_i < \tau$. Note that here the g_f' averaged over lipids was used; therefore, the 'cumulative' $R_1(\tau \to \infty)$ does not necessarily have exactly the same numerical value as the actual R_1.

16.SAMULI: Do we need this paragraph at all? Finally, we note that the fit of Eq. (10) provides an alternative to estimating τ_e, because

$$\tau_e = \int_0^\infty g_f'(\tau)\,d\tau \approx \sum_{i=1}^{N} \alpha_i \tau_i. \qquad (12)$$

When the simulation trajectory is not long enough for the correlation function to reach the plateau, integrating g_f' gives a lower bound estimate for τ_e, while the sum of Eq. (12) includes also (some) contribution from the longer-time components via the fitting process. However, in practice the fit is often highly unreliable in depicting the long tails of the correlation function, and thus we chose to quantify τ_e using the area under g_f', and estimate its uncertainty as detailed above.

3 Results and Discussion

~~In the following, we discuss phospholipid conformational dynamics in six different MD force fields. We do this first for standard conditions (pure POPC bilayers, full hydration, no salt; see Table 1 for simulation details and~~

Internal lipid dynamics in POPC bilayer compared with NMR experiments

The internal dynamics of POPC lamellar liquid crystalline state is compared between MD simulations and NMR experimentsin Fig. 2 ~~for results) and then proceed to cover a wider range of experimentally, biologically, and computationally relevant conditions. We investigate how the dynamics change when cholesterol is added to the bilayer (Table 2 and Fig. 4), when hydration level is reduced (Table 3 and Fig. 5), and when monovalent salt is added to the solution (Table 4 and Fig. 7).~~

~~One should keep in mind that none of the force fields we study produces all the C–H bond order parameters , S_{CH}, within experimental accuracy [22]. This means that the structural ensembles simulated do not exactly match the structural ensemble occurring in reality.Consequently, the τ_e times and R_1 rates depict the dynamics of sampling a somewhat different phase space for each model. To this end, we avoid overly detailed discussion on the models and rather concentrate on common and qualitative trends.~~

~~Effective correlation times τ_e at standard conditions.~~

~~The left panels of Fig. 2 compare the τ_e obtained for fully hydrated POPC bilayers in experiments (black) and in the six different MD force fields (color).~~

~~Qualitatively, every force field captures the general shape of the τ_e profile: Dynamics slows down towards the glycerol backbone in both~~ using the effective correlation time τ_e and R_1 spin relaxation rate. In line with the

comparisons of conformational ensembles using order parameters [?], the effective correlation times indicate that the average dynamics of hydrophopic acyl chain region is better reproduced in MD simulation models than of glycerol backbone and headgroup. All force fields qualitatively capture the slower dynamics in glycerol backbone than in the headgroup and the tails~~. Quantitatively, MD has~~, but MD simulations have a tendency towards slightly too fast dynamics in the membrane core ~~, but~~ and too slow dynamics at the water-facing interface~~MD is typically too slow. CHARMM36 and Slipids show the best overall performance—although the τ_e in Slipids exhibit a qualitatively wrong (decreasing) trend from g_3 to g_1.~~

.The detected slow glycerol backbone dynamics in MD is consistent with previous results for the Berger model[23] ~~. It also agrees with the insufficient~~ **17.SAMULI: I am not sure if we should be mention this work here, because we do not have data for CHARMMc32b2:** and with very slow conformational sampling of glycerol backbone torsions observed in 500-ns-long CHARMMc32b2[75,76] simulations of a DOPC lipid.[26] Such general differences between MD simulations and experiments are not visible in the R_1 rates that mainly detect dynamics at $\sim 1-10$ ns timescales.

~~Note that the temperature varied across these openly available simulation data. However, it was in no case lower than in the experiment. Were the simulations done at the experimental 298 K, the overestimation of τ_e at the glycerol backbone by MD would get worse as τ_e increases at decreasing temperature—as indicated by the CHARMM36 data covering several temperatures.~~

~~R_1 rates at standard conditions.~~

~~The panels on the right side of Fig. 2 compare experimental and simulated R_1 rates under the same conditions as for~~

CHARMM36 and Slipids show the best overall performance—although the τ_e ~~on the left.~~

~~There are certain qualitative features that all force fields predict correctly (for example that g_2 has the smallest R_1 among the glycerol and C9 among the oleoyl double bond segments),~~ in Slipids exhibit a qualitatively wrong (decreasing) trend from g_3 to g_1. This is not surprising because CHARMM36 reproduces also the most realistic conformational ensembles, while Slipids does not correctly capture glycerol backbone structures [22?]. It is important to note that conformational ensembles of glycerol backbone and headgroup greatly differ between MD simulation force fields and ~~certain that they all miss (that~~ are not exactly correct in ay of them [22?]. Consequently, the τ_e times and R_1 rates ~~for the oleoyl segments C8, C10, and C11 are all roughly equal)~~depict the dynamics of sampling a somewhat different and incorrect phase space for each model. To this end, we avoid overly detailed discussion on the models and rather concentrate on common and qualitative trends.

~~Quantitatively,~~ However, there are a few ~~cases where both~~ carbons segments in the data for which the experimental order parameters, R_1 and τ_e ~~(almost) match experiments, suggesting (almost) correct rotational dynamics at all relevant time scales.~~ are approximately reproduced by simulations, suggesting that the conformational ensemble and dynamics is correctly captured by MD simulatios in these cases. For example, Slipids performs well at the β and α segments; CHARMM36 for the g_3, g_2, C2 and C3; Lipid14 and ECC for the oleoyl double bond; and MacRog for the tail end segments.

~~Notably, there are also instances where the R_1 comparison distinctly differs from what is seen for τ_e: Some models that do very well for~~ There are also cases where order parameters and τ_e ~~, do rather poorly for~~ agree with experiments at least approximately, but R_1 ~~. Conversely, a matching R_1 can be accompanied by a larger-than-experimental~~ does not such as ~~—~~ β and α carbons in CHARMM36 force field. Therein a cancellation of error occurs in τ_e~~. To appreciate such differences, recall that in order to capture our experimental R_1 rates (measured at 125 MHz) a force fieldhas to have correct rotational~~ : The overestimated relative dy-

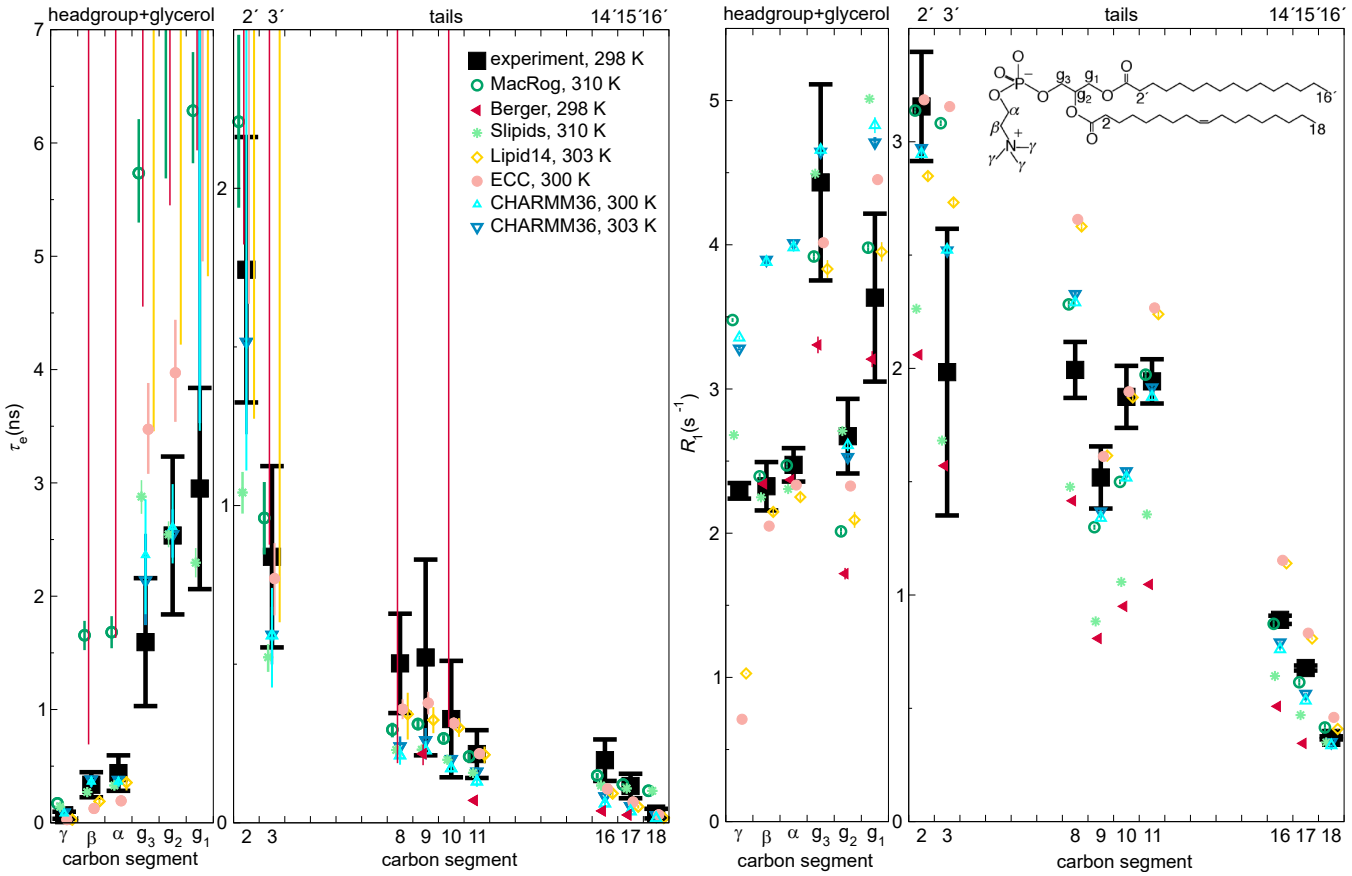

Figure 2: Effective correlation times (τ_e, left panels) and R_1 rates (right panels) in experiments (black) and MD simulations (colored) of POPC bilayers in L_α phase under full hydration. Inset on the right shows the POPC structure and carbon segment labeling. Each plotted value contains contributions from all the hydrogens within its carbon segment; the data for segments 8–11 are only from the sn-2 (oleoyl) chain, whereas the (experimentally non-resolved) contributions of both tails are included for segments 2–3 (2'–3' in the sn-1 chain) and 16–18 (14'–16'). Simulation data are only shown for the segments for which there exists experimental data. For τ_e, a simulation data point indicates the average over C–H bonds; however, if τ_e could not be determined for all bonds, only the error bar (extending from the mean of the lower to the mean of the upper error estimates) is shown. The Berger data for methyl segments (γ, C18, and C16') are left out, because the protonation algorithm used to construct the hydrogens post-simulation in united atom models does not preserve the methyl C–H bond dynamics. Table 1 provides further simulation details. Error bars for the experimental values reflect error estimate of XXX.

18.Experimental error estimate changed since the data were originally published; needs to be explained to the reader.

19.How to refer to the experiments? Not really from previous publication because of re-analysis.

namics at the ~~$(2\pi \times 125\ \text{MHz})^{-1} \approx 1$~~1 ns ~~time scale , whereas~~ scale are compensated by wrong dynamics at the other time scales. As CHARMM36 overall performs rather well for both R_1 and τ_e~~reflects all the sub-μs time scales(Fig. 1).~~

~~MacRog~~, we proceed to study this shortcoming on the headgroup R_1 rates in more detail in the next section. The opposite situation is observed for the β, α, g_3, and g_1 segments ~~provides a prominent example~~ in MacRog simulation where the R_1 rates are well reproduced, but τ_e times systematically overestimated. Such a combination suggests that MD ~~does well at the~~ correct relative weight of 1 ns scale dynamics, but has too slow long-time dynamics.

~~The opposite—where τ_e matches experiments,~~

but R_1 does not is demonstrated by all five all-atom force fields for the γ segment, and by CHARMM36 for β and α. Therein a cancellation of error occurs in τ_e: The wrong dynamics at the 1 ns scale are compensated by wrong dynamics at the other time scales. As CHARMM36 overall performs rather well for both R_1 and τ_e, we proceed to study this shortcoming on the headgroup R_1 rates in some more detail.

Dynamics of headgroup segments in CHARMM36.

20.SAMULI: I would move this to methods. Note that the temperature varied across these openly available simulation data. However, it was in no case lower than in the experiment. Were the simulations done at the experimental 298 K, the overestimation of τ_e at the glycerol backbone by MD would get worse as τ_e increases at decreasing temperature —as indicated by the CHARMM36 data covering several temperatures. **21.HA: add new CHARMM36 data to plot**

Internal motions of POPC headgroup in MD simulations

Because lipid headgroups facing toward water phase are the first membrane parts that interact with approaching proteins, understanding their conformations and dynamics is crucial to explain biological processes where proteins binding to membranes plays critical role. Yet, the analysis of lipid headgroup dynamics from MD simulations and NMR experiments has gained much less attention than acyl chains ? ? . Therefore, we focus here to discuss lessons learned on lipid headgroup dynamics from the analysis of the databank.

Figure 3A zooms in on the headgroup (γ, β, α) segments, whose τ_e were not clearly visible on the scale of Fig. 2. For all three, CHARMM36 matches the experimental τ_e, but overestimates R_1. , No other force field does any better for γ, but while Slipids provides almost perfect dynamics for β and α Slipids provides almost perfect dynamics carbons. Therefore, Slipids simulation gives a realistic model for the conformational dynamics in this region, but relative weight of timescales around ∼1 ns are overestimated in CHARMM36 simulations.

(A) Zoom on the headgroup τ_e (left panel) and R_1 (right). (B) 'Cumulative' R_1 (see Methods for definition) of the γ (top panel), β (middle), and α (bottom) segments. (C) Prefactor weighs α_i from Eq. of γ (top), β (middle), and α (bottom). In B and C, a sliding average over 5 neighboring data points is shown.

To investigate where the differences between force fields arise, we visualize the To further investigate the internal timescales of lipids in simulations, we plot the distribution of relative weights, α_i in Eq. (10), resulting from the fitting to the rotational correlation functions from simulatons in Fig. 3C, and their 'cumulative' contribution to $R_1(\tau)$ in Fig. 3B. **22.SAMULI: maybe change the order of B and C in figure 3?**

23.SAMULI: Maybe this to methods: It is obtained, as detailed in Methods, by including in the sum of Eq. (11) only terms with $\tau_i < \tau$. Consequently, at $\tau \to \infty$ the 'cumulative' $R_1(\tau)$ approaches the actual R_1. Ranges of steepest increase therefore indicate time scales that most strongly contribute to R_1 rates.

Figure 3B shows that for models that overestimate the R_1 rate of γ (MacRog, CHARMM36, and Slipids, see Fig. 3A) the major contribution to R_1 arises at $\tau > 50$ ps, whereas those underestimating the R_1 (Lipid14 and ECC, see Fig. 2) the major contribution comes from $\tau < 50$ ps . This also manifests in the distribution of fitting weights (α_i in Eq.) in Fig. 3C: The earlier the non-zero weights occur, the smaller is the resulting R_1.

For the β and α segments, Fig. 3B shows that the main contribution to R_1 rates arises from processes The dominant relative timescales with the largest weight factors, α_i, are below 100 ps in all models, except in CHARMM36. The large relative weights between 200 ps and 2 ns . As CHARMM36 has the largest weights of all models in this

window (explain the overestimated R_1 rates for the β and α segments in CHARMM36 in Fig. 3C), it overestimates R_1. . In contrast, Slipids, which has simultaneously All the other models reproduce R_1 and τ_e correct , has its largest weights at $\tau < 200$ ps. rates closer to experiments for these segments, while only Slipids simultaneously gives the correct τ_e. Indeed, considerable weights at short time scales (< 10The overestimated τ_e values in MacRog and Berger simulations can be explained by the relative contributions from long time scales around $\sim$?? ps in α for ns. **24.SAMULI: I think we could mention here the approximative location of the peak maximum, but it is hard to see now because not all x-axis values are there.** On the other hand, the underestimated τ_e values in Lipid14, ECC, Berger) and at long time scales (> 10simulations can be explained by significant relative contributions from short time scale dynamics below ~ 20 ns in both β and α for MacRog and Berger) do not manifest at all in the ps.Such short and long time scale contributions are not present in Slipids that correctly reproduces the experimental values for both τ_e and R_1rates. , suggesting the dynamics of α and β segments in POPC headgroup occurs between $\sim$20 ps - 10 ns with the dominant timescales between 40-70 ps. However, the latter contribute heavily on τ_e, which is thus considerably overestimated by MacRog and Berger (Fig. 2).

What are the motions in It would be highly interesting to identify the origin of observed artificial timescales, particularly for the 0.2–2 ns window that are over-presented in CHARMM36? Identifying them and speeding them up would improve the model dynamics. , and propose how those could be corrected in the simulation models. However, the connection between the fitted correlation times and the correlation times of distinct motional processes, such as dihedral rotations and lipid wobbling, turns out to be highly non-trivial; we thus refrain from further analysis here.

Effect of cholesterol.

The experimental effective correlation times τ_e (Fig. 4A, top panels) show that when cholesterol is added, the glycerol region conformational dynamics slow down markedly. The tail segments slow down too, the effect increasing towards the backbone.

In stark contrast, however, the τ_e of headgroup segments (γ, β, α) are unaffected by cholesterol. Furthermore, cholesterol induces no measurable change in the headgroup β and α segment dynamics at short ($\sim$1 ns) time scales, as demonstrated by the experimental R_1 rates (Fig. 4A, lower panels). That said, there is a small but measurable impact on R_1 at γ.

All the force fields investigated qualitatively reproduce the increase in τ_e (see Fig. 4B): Slipids gives the best magnitude estimates, while CHARMM36 and MacRog clearly overestimate the changes at the glycerol, C2, and C3 carbons. Notably, MacRog **25.and Berger?** erroneously predict slow down also for the β and α carbons, for which experiments detect no change. Note that, while CHARMM36 correctly shows no chance in τ_e of the γ, β, and α carbons, it predicts a non-zero ΔR_1 for all three, indicating some inaccuracies in the headgroup rotational dynamics. Such inaccuracies might be reflected in the recent findings[77] (obtained using CHARMM36) that the headgroups of PCs neighbouring a cholesterol (within 6.6 Å) spend more time on top of the cholesterol than elsewhere; such arrested rotations could manifest on τ_e and R_1. Interestingly, the tail ΔR_1 seem to be pretty well reproduced by all three all-atom force fields, whereas Berger fails to capture the change at the oleoyl double bond.

Effect of drying.

Figure 5A shows how a mild dehydration affects C–H bond dynamics in the PC headgroup and glycerol backbone; the plot compares the experimental effective correlation times τ_e measured for POPC at full hydration and for DMPC (1,2-dimyristoyl-sn-glycero-3-phosphocholine) at 13 waters per lipid.

The τ_e are the same within experimental ac-

curacy, which suggests two conclusions. Firstly, the headgroup (γ, β, α) τ_e are unaffected by structural differences in the tails. This is analogous to what was seen experimentally when adding cholesterol (Fig. 4): Changes in the tail and glycerol regions do not reflect to the headgroup. Secondly, a mild dehydration does not alter the τ_e in the headgroup and glycerol regions.

Figure 5B shows the effects of dehydration in three MD models. Combination of the unrealistically slow dynamics, especially in the glycerol backbone, (Fig. 2) and the relatively short lengths of the openly available trajectories (Table 3) led to large uncertainty estimates. However, in the γ segment there is clearly no effect above 13 w/l in CHARMM36 and MacRog, in agreement with the experiments; reducing water content further induces a slow down, especially in MacRog below 10 w/l. Similarly, the β and α segments show no detectable change above 13 w/l for CHARMM36 and Berger, in agreement with the experiments; below 10 w/l Berger exhibits a slowdown, and in CHARMM36 the slowdown manifests as an abrupt increase of the uncertainty estimate.

Owing to the large uncertainties, we only point out the qualitative trends of the lower error estimates on the glycerol segments. For CHARMM36 it stays almost constant all the way until 7 w/l, for Berger and MacRog the lower error estimate suggests a retardation of the dynamics starting already from ∼20 w/l.

These simulational findings suggest that experiments reducing hydration levels below 10 w/l would also show an increase in τ_e. This prediction is in line with the exponential slowdown of the headgroup conformational dynamics upon dehydration that was indicated by ^{2}H-NMR R_1 measurements of DOPC bilayers: $R_1 \sim \exp(-n_{w/l}/4)$.[78] The slowdown was attributed to the reduction in the effective volume available for the headgroup[78] owing to its tilt towards the membrane upon dehydration; the tilt is observed via changes of the lipid headgroup order parameters,[79] and is qualitatively reproduced by all the simulation models.[22]

Figure 6 shows a collection of experimental ^{13}C-NMR R_1 rates measured at 125 MHz for the headgroup segments at different water contents; in addition to the full hydration POPC data from Fig. 2, DMPC at 13 w/l,[30] and POPC at 20 and 5 w/l[31] are shown. An increasing trend with decreasing hydration is observed for all the segments, indicating changes of headgroup dynamics at short (∼1 ns) time scales. Interestingly, only CHARMM36 captures this, whereas Berger and MacRog give decreasing R_1 rates for β and α.

The slow down discussed here is of significance not only when simulating a bilayer (stack) under low hydration, but also for studies of intermembrane interactions, such as membrane fusion, because these naturally lead to dehydrated conditions when the lipid assemblies approach. Slower dynamics imply that longer simulation times are needed for equilibration, for reliably quantifying the properties of the bilayers, and for observing rare events.

Effect of cation binding.

32.MARKUS: I have started to think that we maybe should drop this section, because we do not have any experimental data to compare against. However, if we do decide to keep it, to me it seems that our main point here is not on the effects monovalent salt, but rather on the effects of cation binding. Therefore, it would be better to plot $\Delta\tau_e$ as a function of the bound cation charge, as we did in Fig. 3 of Ref. 38; then we could also include calcium data.

Finally, we comment on the response of the MD model dynamics to increasing amounts of monovalent salt. To our knowledge, no experimental ^{13}C-NMR R_1 or τ_e data exists as a function of monovalent salt concentration; therefore, the following discussion is kept qualitative. 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[38,80]). 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.

The molecular electrometer has been used to show that most molecular dynamics force fields

overestimate the binding of monovalent ions to PC bilayers:[38] 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. 7 we compare three force fields, one that is known to exhibit pronounced overbinding[38] (MacRog) and two producing more realistic binding affinity (Slipids and CHARMM36). The lateral distribution of Na^+ ions near the bilayer is quantified in Fig. 7A whereas Fig. 7B shows the chance in τ_e for increasing salt concentration. Ion accumulation results in a slow down in the effective correlation time. Correlation times extracted from CHARMM36 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 binding affinity of lipid bilayers and experimental work exploring this avenue would be interesting.

33.validity of statement regarding Slipids

Correlation time of S_{CH} versus τ_e.

To determine the C–H bond order parameter S_{CH} (Eq. (2)) in MD, one first calculates an instantaneous order parameter

$$S_{CH}(t) = \frac{1}{2}\left(3\cos^2\theta(t) - 1\right), \qquad (13)$$

where $\theta(t)$ is the angle between the C–H bond and the membrane normal at time t. As this quantity is sampled along the trajectory, its average $\langle S_{CH}(t)\rangle$ approaches S_{CH}. For reliable determination of S_{CH}, it would be of interest to know the correlation time of this relaxation, because it determines the minimum simulation length required.

To this end, it is also of interest to know how this correlation time relates to τ_e. As discussed (see Fig. 1), in a bilayer the C–H bond's second order rotational correlation function $g(\tau)$, see Eq. (1), approaches S_{CH}^2 with time. The speed of this approach tells how fast the C–H bond orientations are sampled. However, the correlation time of S_{CH}, which is calculated using the *a priori* knowledge of the membrane normal direction (Eq. (13)), does not need to equal τ_e. Rather, one would intuitively expect it to be shorter than τ_e, because the rotational averaging around the membrane normal direction is already implicitly taken into account in Eq. (13).

A further complication is that the relaxational process of the C–H bond direction (used to determine τ_e) can be single or multi-exponential. If the relaxation is single-exponential, τ_e is the relaxation time of this exponential process. If the relaxation is multi-exponential, τ_e is the weighted mean of the corresponding set of relaxation times, and it is a bit hard to say just based on τ_e how long one needs to sample to reach the S_{CH}, because this depends also on the weights of the processes.

Figure **??** shows this correlation for systems studied in this work; we see... **34.Laske bilayerissa S_{CH}:n korrelaatioaika (yksittaisessa lipidissa) vs τ_e. Tee scatter plot.**

4 Conclusions

Open access databanks of MD trajectories enables the creation new scientific information without running a single new simulation. Here, we demonstrated this by investigating the dynamics of a wide range of phosphatidylcholine molecular dynamics models using the existing trajectories from the NMRlipids databank.

We found that MD qualitatively captures the ^{13}C-NMR effective correlation time (τ_e) profile of POPC—the slow glycerol backbone and the faster motions of the headgroup and tail regions—but most MD force fields are prone to too slow dynamics of the glycerol C–H bonds (Fig. 2). While no force field reproduces all the experimental data, CHARMM36 and Slipids have an overall impressive τ_e. This is particularly true for CHARMM36, as it is also known to well reproduce the experimental conformational ensemble.[22] That said, we find that CHARMM36 struggles with the balance of dy-

namics in the headgroup region: The R_1 rates, sensitive for ∼1-ns processes, are too high for the γ, β, and α segments (Fig. 3).

35.Make the point that the 500-ns simulations indicated by Vogel[26] are not needed for sufficient sampling?

In addition to standard conditions, we explored how the dynamics react to addition of cholesterol or NaCl, or to removal of water. MD qualitatively captures that when cholesterol is mixed into a POPC bilayer, the conformational dynamics in the tail and glycerol regions slows down; however, some force fields predict an (erroneous) slowdown also for the headgroup (Fig. 4). With increasing NaCl concentration, a behaviour reminiscent of the molecular electrometer was observed: Amount of ion binding to the bilayer correlated with the magnitude increase in τ_e; this could open up the possibility of using τ_e in quantifying cation binding to lipid bilayers. When reducing the water content, MD exhibits slowdown of headgroup and backbone dynamics below ∼10 waters per lipid in qualitative agreement with experimental data. **36. Hydration needs some kind of statement of significance.**

By gathering a set of ^{13}C-NMR data on the phosphatidylcholine dynamics and charting the typical features of the existing MD models against it, this study lays the foundation for further improvement of the force fields. While work is still needed in capturing even the correct conformations,[22] realistic dynamics will be an essential part of developing MD into a true computational microscope.

Importantly, this work demonstrates the power of open data in creating new knowledge out of existing trajectories at a reduced computational and labor cost. If the data are well indexed and documented, this process could be automated and has the potential to facilitate faster progress, e.g., in the development of MD force fields, for example through machine learning approaches.

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.fi)

References

(1) Crystallography: Protein Data Bank. *Nat. New Biol.* **1971**, *233*, 223.

(2) wwPDB consortium, Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **2019**, *47*, D520–D528.

(3) Jordan, E.; Carrico, C. DNA Database. *Science* **1982**, *218*, 108.

(4) Sayers, E. W.; Cavanaugh, M.; Clark, K.; Ostell, J.; Pruitt, K. D.; Karsch-Mizrachi, I. GenBank. *Nucleic. Acids Res.* **2020**, *48*, D84–D86.

(5) Gaber, Y.; Rashad, B.; Fathy, E. In *Biological 3D Structural Databases*; Shaik, N. A., Hakeem, K. R., Banaganapalli, B., Elango, R., Eds.; Springer International Publishing: Cham, 2019; pp 47–73.

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

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

(8) 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.

(9) 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.

(10) 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.

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

(12) Hildebrand, P. W.; Rose, A. S.; Tiemann, J. K. S. Bringing Molecular Dynamics Simulation Data into View. *Trends in Biochemical Sciences* **2019**, *44*, 902–913.

(13) Abraham, M.; Apostolov, R.; Barnoud, J.; Bauer, P.; Blau, C.; Bonvin, A. M. J. J.; Chavent, M.; Chodera, J.; Čondić-Jurkić, K.; Delemotte, L. et al. Sharing Data from Molecular Simulations. *Journal of Chemical Information and Modeling* **2019**, *59*, 4093–4099.

(14) Gygli, G.; Pleiss, J. Simulation Foundry: Automated and F.A.I.R. Molecular Modeling. *Journal of Chemical Information and Modeling* **2020**, *60*, 1922–1927.

(15) Abriata, L. A.; Lepore, R.; Dal Peraro, M. About the need to make computational models of biological macromolecules available and discoverable. *Bioinformatics* **2020**, *36*, 2952–2954.

(16) Hospital, A.; Battistini, F.; Soliva, R.; Gelpí, J. L.; Orozco, M. Surviving the deluge of biosimulation data. *WIREs Computational Molecular Science* **2020**, *10*, e1449.

(17) Rodríguez-Espigares, I.; Torrens-Fontanals, M.; Tiemann, J. K. S.; Aranda-García, D.; Ramírez-Anguita, J. M.; Stepniewski, T. M.; Worp, N.; Varela-Rial, A.; Morales-Pastor, A.; Lacruz, B. M. et al. GPCRmd uncovers the dynamics of the 3D-GPCRome. *bioRxiv* **2019**, 839597.

(18) Miettinen, M. S.; NMRlipids Collaboration,; Ollila, O. H. S. LDB: Lipid Databank from the NMRlipids Project. *Biophysical Journal* **2019**, *116*, 91a.

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

(20) 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.

(21) 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.

(22) 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.

(23) 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.

(24) Miettinen, M. S.; Lipowsky, R. Bilayer Membranes with Frequent Flip-Flops Have Tensionless Leaflets. *Nano Lett.* **2019**, *19*, 5011–5016.

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

(26) 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.

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

(28) 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.

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

(30) 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.

(31) Volke, F.; Pampel, A. Membrane Hydration and Structure on a Subnanometer Scale as Seen by High Resolution Solid State Nuclear Magnetic Resonance: POPC and POPC/$C_{12}E0_4$ Model Membranes. *Biophys. J.* **1995**, *68*, 1960–1965.

(32) 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.

(33) 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.

(34) 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.

(35) 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.

(36) 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.

(37) 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.

(38) 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.

(39) 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.

(40) 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}.

(41) 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.

(42) 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.

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

(44) 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.

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

(46) 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.

(47) 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.

(48) 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.

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

(50) 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.

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

(52) 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.

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

(54) 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.

(55) 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.

(56) 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.

(57) 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.

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

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

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

(61) 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**,

(62) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=4 w/l. 2015; DOI: 10.5281/zenodo.16295.

(63) 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.

(64) 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.

(65) 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.

(66) 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.

(67) 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.

(68) 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.

(69) 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.

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

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

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

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

(74) 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.

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

(76) 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.

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

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

(79) Bechinger, B.; Seelig, J. Conformational changes of the phosphatidylcholine headgroup due to membrane dehydration. A 2H-NMR study. *Chemistry and Physics of Lipids* **1991**, *58*, 1 – 5.

(80) 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.

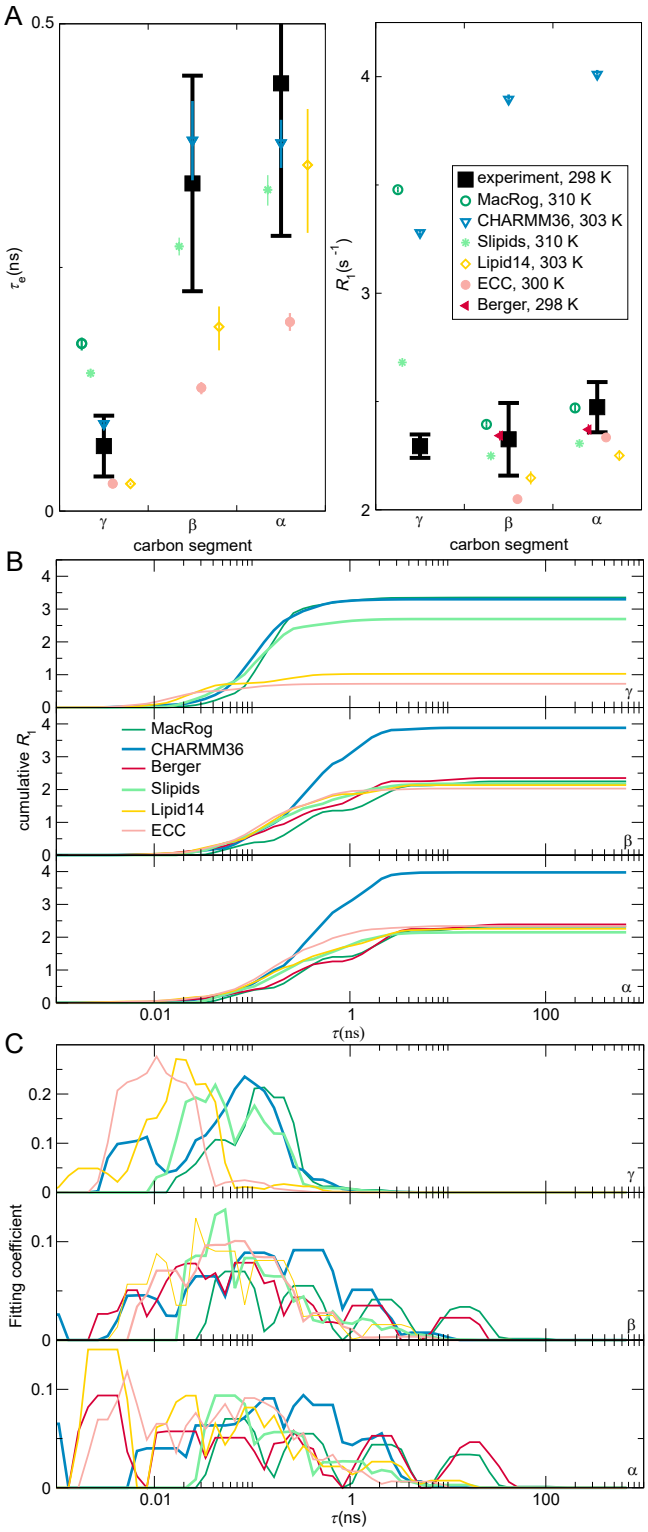

Figure 3: (A) Zoom on the headgroup τ_e (left panel) and R_1 (right). (B) 'Cumulative' R_1 (see Methods for definition) of the γ (top panel), β (middle), and α (bottom) segments. (C) Prefactor weighs α_i from Eq. (10) of γ (top), β (middle), and α (bottom). In B and C, a sliding average over 5 neighboring data points is shown.

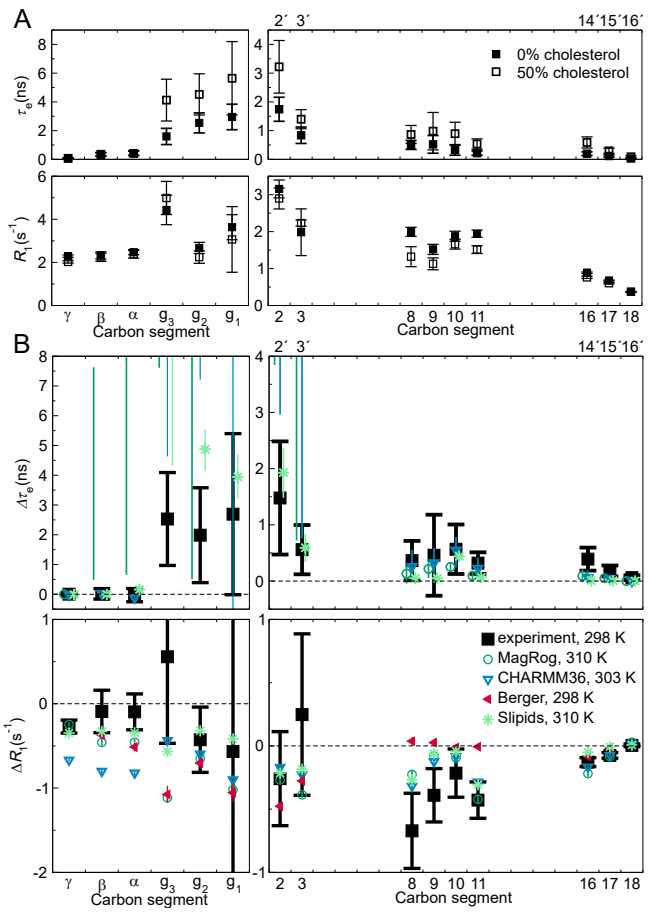

Figure 4: Effect of bilayer cholesterol content. (A) The experimental effective correlation times τ_e (top panels) and R_1 rates (bottom) in a pure POPC bilayer and in a bilayer containing 50% cholesterol. The data were measured at 298 K and full hydration. (B) The change in τ_e ($\Delta\tau_e$, top panels) and R_1 (ΔR_1, bottom), both in experiments and in MD simulations, when bilayer composition changes from pure POPC to 50% cholesterol. Berger not shown for $\Delta\tau_e$, because the open data available were insufficient to determine meaningful error estimates. Error estimates for the simulated $\Delta\tau_e$ are the maximal possible based on the errors at 0% and 50% cholesterol; for other data regular error propagation is used. Table 2 provides further simulation details; for segment labeling, see Fig. 2.

26.@Hanne: Double check that the calculation of errors in (B) was as the caption describes. 27.Check if cholesterol data is in full hydration

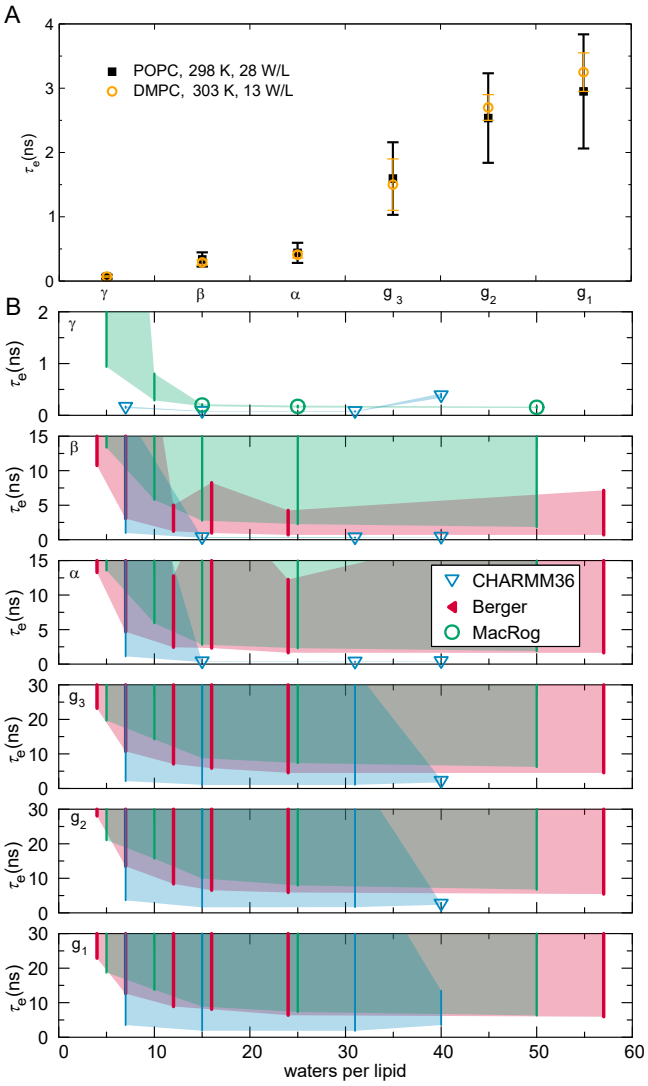

Figure 5: Effect of drying on effective correlation times in headgroup and glycerol backbone. (A) Experimental τ_e for DMPC (from Ref. 30) at low hydration do not significantly differ from the τ_e for POPC at full hydration. (B) Calculated τ_e for POPC at decreasing hydration in three MD models. Note that three Berger data points are from DLPC bilayers dashed. Symbols give the mean of segment hydrogens, if τ_e could be determined for all hydrogens; else only the error bar (extending from the mean of the lower to the mean of the upper uncertainty estimates) is shown; the area delimited by the error bars is shaded for visualization. See Table 3 for simulation details.

28.How to refer to full hydration POPC data? 29.Add also the black and orange (i.e. experimental) data points to B. 30.DLPC Berger points dashed or just in caption

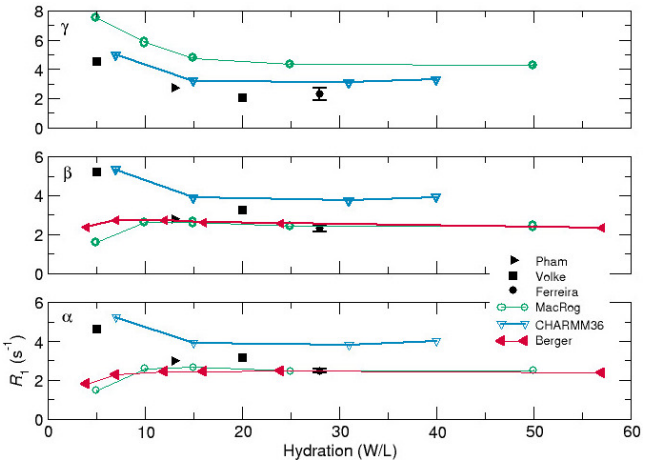

Figure 6: Effect of drying on ^{13}C-NMR R_1 rates of the headgroup segments (at 125 MHz) in experiments and simulations.

31.HA: Redo & merge this with the Fig. 5

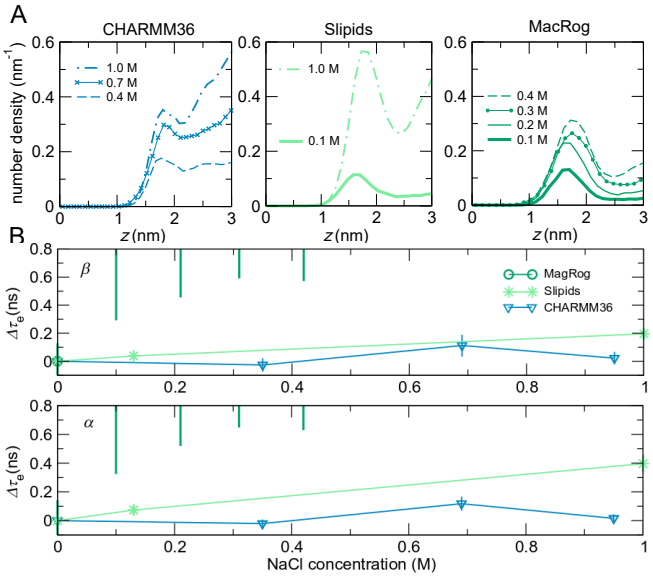

Figure 7: The impact of increasing ionic strength on effective correlation times. a) The density distribution (average over both leaflets) of Na^+ ions as function of distance z from the bilayer center. The plots for each force field are presented from left to right in the order of increasing ion accumulation. 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.

Graphical TOC Entry

TOC here if needed