

# Using open data to rapidly benchmark biomolecular simulations: Phospholipid conformational dynamics

Hanne S. Antila,<sup>\*,†</sup> Tiago M. Ferreira,<sup>‡</sup> O. H. Samuli Ollila,<sup>¶</sup> and Markus S. Miettinen<sup>\*,†</sup>

<sup>†</sup>*Department of Theory and Bio-Systems, Max Planck Institute of Colloids and Interfaces, 14424 Potsdam, Germany*

<sup>‡</sup>*NMR Group — Institute for Physics, Martin-Luther University Halle-Wittenberg, 06120 Halle (Saale), Germany*

<sup>¶</sup>*Institute of Biotechnology, University of Helsinki, 00014 Helsinki, Finland*

E-mail: hanne.antila@mpikg.mpg.de; 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  $\tau_e$  of the C-H bonds measured in nuclear magnetic resonance experiments (J. Chem. Phys. 142 044905 (2015)).

Here, we use a large set of open data trajectories made public by the NMRlipids project ([nmrlipids.blogspot.fi](http://nmrlipids.blogspot.fi)) to perform an unprecedented test on the conformational dynamics of phospholipids as produced by several commonly used MD models (force fields). We find that none of the tested force fields to reproduce all the effective correlation times within experimental error, much like they do not provide accurate conformational ensembles (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.Tiago: We need a punchline highlighting that the results and findings obtained are extremely important for the lipid MD simulation community and that they were only possible due to the open data**

## 1 Introduction

Ever since the conception of Protein Data Bank (PDB)<sup>1,2</sup> and GenBank,<sup>3,4</sup> open access to standardised and searchable pools of experimental data has revolutionized scientific research. Constantly growing and improving in fidelity due to collaborative effort,<sup>5–8</sup> the now hundreds of databanks<sup>9</sup> fuel the data-driven development of biomolecular structure determination,<sup>10</sup> refinement,<sup>11</sup> prediction,<sup>12</sup> and design<sup>13</sup> approaches, as well as development of drugs,<sup>14,15</sup> materials,<sup>16,17</sup> and more.<sup>18,19</sup> It is clear that open data enables scientific progress that is far beyond the resources of a single research group or institute. Consequently, the call for public availability and conservation of data has extended to molecular dynamics (MD) simulation trajectories of biomolecules,<sup>20–22</sup> and the discussion on how and by whom such databanks for dynamic structures would be set up is currently active.<sup>23–26</sup> While no general MD tra-

jectory databank currently operates, individual databanks are accepting contributions on nucleic acid,<sup>27</sup> protein/DNA/RNA,<sup>28</sup> cyclodextrin,<sup>29</sup> G-protein-coupled receptor,<sup>30</sup> and lipid bilayer<sup>31</sup> simulations.

Since 2013, the NRMlipids Project ([nmrlipids.blogspot.fi](http://nmrlipids.blogspot.fi)) has promoted a fully open collaboration approach, where the whole scientific research process—from initial ideas and discussions to analysis methods, data, and publications—is all the time publicly available.<sup>32</sup> While its main focus has been on conformational ensembles of different lipid headgroups and on ion binding to lipid membranes,<sup>32–34</sup> the NRMlipids Project has also built a databank<sup>31</sup> ([zenodo.org/communities/nmrlipids](http://zenodo.org/communities/nmrlipids)) containing hundreds of atomistic MD trajectories of lipid bilayers and indexed at [nmrlipids.fi](http://nmrlipids.fi).

MD databanks are expected to be particularly relevant for disordered biomolecules, such as biological lipids composing cellular membranes or intrinsically disordered proteins. These, in contrast to folded proteins or DNA strands, cannot be meaningfully described by the coordinates of a single structure alone. Realistic MD simulations, however, can provide the complete conformational ensemble and dynamics of such molecules, as well as enable studies of their biological functions in complex biomolecular assemblies. Unfortunately, the current MD force fields largely fail to capture the conformational ensembles of lipid headgroups and disordered proteins.<sup>32,34–37</sup> Therefore, before they can be used to draw conclusions, the quality of MD simulations must always be carefully assessed against structurally sensitive experiments. For lipid bilayers, such evaluation is possible against NMR and scattering data.<sup>38</sup>

Here, we demonstrate the use of a pre-existing, publicly available set of MD trajectories to rapidly evaluate the fidelity of phospholipid conformational dynamics in state-of-the-art force fields. The speed at which individual molecules sample their conformational ensemble is traditionally used to assess if a given MD simulation has converged. Going beyond such practicalities, realistic dynamics are particularly desired for the intuitive interpretation

of NMR experiments sensitive to molecular motions,<sup>39</sup> as well as to understand the dynamics of biological processes where molecular deformations play a rate-limiting role, such as membrane fusion.<sup>40</sup> The here presented comprehensive comparison of dynamics between experiments and different MD models at various biologically relevant compositions and conditions is thus likely to facilitate the development of increasingly realistic phospholipid force fields.

Above all, our results demonstrate the power of publicly available MD trajectories in creating new knowledge at a lowered computational cost and high potential for automation. We believe that this paves the way for novel applications of MD trajectory databanks, as well as underlines their usefulness—not only for lipid membranes, but for all biomolecular systems.

## 2 Methods

### Evaluation of lipid conformational dynamics in MD against NMR data.

We analyzed the veracity of lipid dynamics in MD based on two quantities that are readily available from published<sup>39,41–43</sup>  $^{13}\text{C}$ -NMR experiments and directly quantifiable from atomistic MD simulations: The effective C–H bond correlation times  $\tau_e$ , and the spin-lattice relaxation rates  $R_1$ .

#### Effective C–H bond correlation times $\tau_e$ .

In a lipid bilayer in liquid crystalline phase, each individual lipid samples its internal conformational ensemble and rotates around the axis normal to the membrane. Lipid conformational dynamics are reflected in the second order autocorrelation functions of its C–H bonds

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

where the angular brackets depict time average,  $\vec{\mu}(t)$  is the unit vector in the direction of the C–H bond at time  $t$ , and  $P_2$  is the second order Legendre polynomial  $P_2(x) = \frac{1}{2}(3x^2 - 1)$ . To analyze the internal dynamics of lipids, the C–H bond autocorrelation function is often written as a product

$$g(\tau) = g_f(\tau)g_s(\tau), \quad (2)$$

where  $g_f(\tau)$  characterizes the fast decays owing to, e.g., the internal dynamics and rotation around membrane normal, and  $g_s(\tau)$  the slow decays that originate from, e.g., lipid diffusion between lamellae with different orientations, and periodic motions due to the use of magic angle spinning conditions (Fig. 1). Ferreira et al.<sup>41</sup> have experimentally demonstrated that for all phospholipid carbons the motion correlation times contributing to  $g_f$  are well below  $\mu\text{s}$  and to  $g_s$  well above  $100 \mu\text{s}$ . This separation of time-scales gives rise to the plateau  $g(1 \mu\text{s}) \lesssim \tau \lesssim 100 \mu\text{s}) = S_{\text{CH}}^2$  illustrated in

Fig. 1.  $S_{\text{CH}}$  is the C–H bond order parameter

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

where  $\theta(t)$  is the angle between the C–H bond and the bilayer normal.  $S_{\text{CH}}$  can be independently measured using dipolar coupling in  $^{13}\text{C}$  or quadrupolar coupling in  $^2\text{H}$ -NMR experiments, and it is highly useful in order to evaluate conformational ensembles of lipids.<sup>38</sup>

Since  $S_{\text{CH}}$  describe the conformational ensemble of the lipid, the fast-decaying component  $g_f$  of the C–H bond autocorrelation function intuitively reflects the time needed to sample these conformations. The complex internal dynamics containing multiple timescales can be conveniently summarized using the effective correlation time

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

which is related to the gray-shaded area below the correlation function in Fig. 1. The  $\tau_e$  detect essentially an average over all the time scales relevant for the lipid internal dynamics, and have an intuitive relation to dynamics: In the presence of more long-lived correlations  $\tau_e$  grows. The integrand in Eq. (4) defines the reduced and normalized correlation function

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

#### Spin-lattice relaxation rates $R_1$ .

C–H bond dynamics is related to the spin-lattice relaxation rate through

$$R_1 = \frac{d_{\text{CH}}^2 N_{\text{H}}}{20} [j(\omega_{\text{H}} - \omega_{\text{C}}) + 3j(\omega_{\text{C}}) + 6j(\omega_{\text{H}} + \omega_{\text{C}})], \quad (6)$$

where  $\omega_{\text{H}}$  is the  $^1\text{H}$  and  $\omega_{\text{C}}$  the  $^{13}\text{C}$ -NMR Larmor frequency,  $N_{\text{H}}$  the number of bound hydrogens, and  $d_{\text{CH}}$  the rigid dipolar coupling constant. For the methylene bond,  $d_{\text{CH}}/2\pi$  approximately equals to  $-22 \text{ kHz}$ . The spectral density

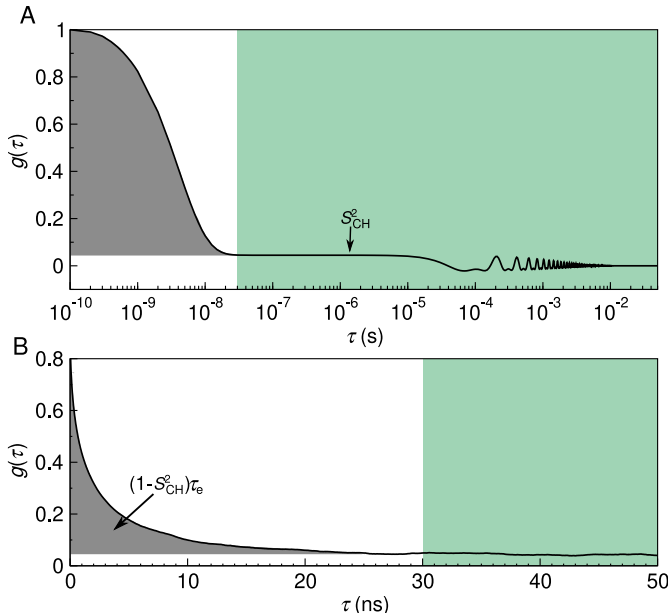


Figure 1: The C–H bond autocorrelation function  $g(\tau)$ . (A) Illustration of the fast (white background) and the slow (green) mode of the correlation function. The fast mode decays to a plateau on which  $g(\tau) = S_{\text{CH}}^2$ , while the slow mode gives the final descent to zero. Oscillations at the slow mode region are due to magic angle spinning. (B) Typical C–H bond autocorrelation function obtained from a MD simulation showing a decent towards  $g(\tau) = S_{\text{CH}}^2$ . The gray area under the curve is equal to  $(1 - S_{\text{CH}}^2)\tau_e$ .

$j(\omega)$  is given by the Fourier transformation

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

of the C–H bond autocorrelation function  $g(\tau)$  (Eq. (1)). Clearly the connection between  $R_1$  and molecular dynamics is not straightforward; the magnitude of  $R_1$  does, however, reflect the relative significance of processes with timescales near the inverse of  $\omega_{\text{H}}$  and  $\omega_{\text{C}}$ . These two frequencies depend on the field strength used in the NMR experiments: Typically  $R_1$  is most sensitive to motions with time scales  $\sim 1$ – $10$  ns. A change in  $R_1$ , therefore, indicates a change in the relative amount of processes occurring in a window around the sensitive timescale, but does not give information on the direction (speedup/slowdown) to which the processes changed.

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

force field	$N_1^a$	$N_w^b$	$T^c(\text{K})$	$t_{\text{anal}}^d(\text{ns})$	files <sup>e</sup>
Berger <sup>44,45</sup>	256	10240	300	300	[46]
CHARMM36 <sup>47</sup>	256	8704	300	300	[48]
MacRog <sup>49</sup>	128	5120	300	500	[50]
Lipid14 <sup>51</sup>	72	2234	303	50	[52]
Slipids <sup>53</sup>	200	9000	310	500	[54]
ECC <sup>55</sup>	128	6400	300	300	[56]

<sup>a</sup>Number of POPC molecules.

<sup>b</sup>Number of water molecules.

<sup>c</sup>Simulation temperature.

<sup>d</sup>Trajectory length used for analysis.

<sup>e</sup>Reference for the openly available simulation files.

## Experimental data acquisition and analysis.

All the experimental quantities were collected from the literature sources referred at the respective figures.

## Simulational data acquisition and analysis.

The simulation trajectories used in this work were collected from the Zenodo repository ([zenodo.org](https://zenodo.org)) with majority of the data originating from the NMRlipids Project<sup>32,33</sup> ([nmrlipids.blogspot.fi](https://nmrlipids.blogspot.fi)). The trajectories were chosen based on how well the simulation conditions matched the available experimental data (temperature, cholesterol content, hydration), and how precisely one could extract the quantities of interest from the trajectory (length of simulation, system size).

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, and Table 3 simulations with varying hydration. Additional computational details for the simulations are available at the cited Zenodo entry.

The simulation data were analyzed using in-house scripts. These are available on GitHub ([https://github.com/hsantila/Corrtimes/tree/master/teff\\_analysis](https://github.com/hsantila/Corrtimes/tree/master/teff_analysis)) along with a

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

force field POPC/cholesterol	$c_{\text{chol}}^a$	$N_{\text{chol}}^b$	$N_l^c$	$N_w^d$	$T^e$ (K)	$t_{\text{anal}}^f$ (ns)	files <sup>g</sup>
Berger-POPC-07 <sup>57</sup>	0%	0	128	7290	298	50	[58]
/Höltje-CHOL-13 <sup>59,60</sup>	50%	64	64	10314	298	50	[61]
CHARMM36 <sup>47</sup>	0%	0	200	9000	310	500	[62]
/CHARMM36 <sup>63</sup>	50%	200	200	18000	310	500	[64]
MacRog <sup>49</sup>	0%	0	128	6400	310	500	[65]
/MacRog <sup>49</sup>	50%	64	64	6400	310	500	[65]
Slipids <sup>53</sup>	0%	0	200	9000	310	500	[54]
/Slipids <sup>66</sup>	50%	200	200	18000	310	500	[54]

<sup>a</sup>Bilayer cholesterol content (mol %).

<sup>b</sup>Number of cholesterol molecules.

<sup>c</sup>Number of POPC molecules.

<sup>d</sup>Number of water molecules.

<sup>e</sup>Simulation temperature.

<sup>f</sup>Trajectory length used for analysis.

<sup>g</sup>Reference for the openly available simulation files.

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)	$t_{\text{anal}}^e$ (ns)	files <sup>f</sup>
Berger-POPC-07 <sup>57</sup>	POPC	57	128	7290	298	50	[58]
	POPC	7	128	896	298	60	[67]
Berger <sup>44,45</sup>	POPC	40	256	10240	300	300	[46]
Berger-DLPC-13 <sup>68</sup>	DLPC <sup>g</sup>	24	72	1728	300	80	[69]
	DLPC <sup>g</sup>	16	72	1152	300	80	[70]
	DLPC <sup>g</sup>	12	72	864	300	80	[71]
	DLPC <sup>g</sup>	4	72	288	300	80	[72]
CHARMM36 <sup>47</sup>	POPC	40	128	5120	303	140	[73]
	POPC	34	128	5120	300	500	[50]
	POPC	31	72	2232	303	20	[74]
	POPC	15	72	1080	303	20	[75]
MacRog <sup>49</sup>	POPC	7	72	504	303	20	[76]
	POPC	50	288	14400	310	40	[77]
	POPC	25	288	7200	310	50	[77]
	POPC	15	288	4320	310	50	[77]
	POPC	10	288	2880	310	50	[77]
	POPC	5	288	1440	310	50	[77]

<sup>a</sup>Water/lipid molar ratio.

<sup>b</sup>Number of lipid molecules.

<sup>c</sup>Number of water molecules.

<sup>d</sup>Simulation temperature.

<sup>e</sup>Trajectory length used for analysis.

<sup>f</sup>Reference for the openly available simulation files.

<sup>g</sup>1,2-didodecanoyl-sn-glycero-3-phosphocholine.

Python notebook outlining an example analysis run. To enable automated analysis of several force fields with different atom naming conventions, we employ mapping files and the related definition format developed within the NMR-lipids project to recognize the atoms and bonds of interest when analyzing the trajectory.

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  $S_{\text{CH}}$ , see Eq. (3), were then calculated with the `OrderParameter.py` script that uses the MD-analysis<sup>78,79</sup> Python library. The C–H bond correlation functions  $g(\tau)$  (see Eq. (1)) were calculated with Gromacs5.1.4<sup>80</sup> `gmx rotacf` (note that on simulational (fast) time scales  $g = g_s g_f = g_f$ ) after which the  $S_{\text{CH}}$  were used to normalize the  $g_f$  to obtain the  $g'_f$ , following Eq. (5).

The effective correlation times  $\tau_e$  were then calculated by integrating  $g'_f(\tau)$ , (Eq. (5)), from  $\tau = 0$  until  $\tau = t_0$ . Here,  $t_0$  is the first time point at which  $g'_f$  reached zero,  $t_0 = \min\{t | g'(t) = 0\}$ . If  $g'_f$  did not reach zero within  $t_{\text{anal}}/2$ , the  $\tau_e$  was not determined, and we report only its upper and lower estimates.

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

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

Here the  $\Delta S_{\text{CH}}$  was determined as the standard error of the mean of the  $S_{\text{CH}}$  over the  $N_l$  individual lipids in the system.<sup>32</sup> Similarly, we quantified the error on  $g_f(\tau)$  by first determining the correlation function  $g_f^m(\tau)$  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 for  $g_f(\tau)$  at each time point  $\tau$ .

To obtain the lower bound on  $\tau_e$ , we integrate the function  $g'_f(\tau) - \Delta g'_f(\tau)$  over time from  $\tau = 0$  until  $\tau = t_1$ . Here

$$t_1 = \min \left\{ \{t \mid g'_f(t) - \Delta g'_f(t) = 0\}, \frac{t_{\text{anal}}}{2} \right\}. \quad (9)$$

That is,  $t_1$  equals the first time point at which the lower error estimate of  $g'_f$  reached zero; or  $t_1 = t_{\text{anal}}/2$ , if zero was not reached by that point.

To obtain the upper error estimate on  $\tau_e$ , we first integrate the function  $g'_f(\tau) + \Delta g'_f(\tau)$  over time from  $\tau = 0$  until  $t_u = \min \{t_0, t_{\text{anal}}/2\}$ . 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  $\tau_e$  with a low weight, their contribution over long times could still add up to a sizable effect on  $\tau_e$ . That said, it is feasible 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  $\tau_e$ , and assume that all the decay from the time point  $t_u = \min \{t_0, t_{\text{anal}}/2\}$  onwards comes solely from this slowest process.

The additional contribution to the upper bound for  $\tau_e$  then reads

$$\begin{aligned} & (g'_f(t_u) + \Delta g'_f(t_u)) \int_{t_u}^T \exp(-(t - t_u)/T) dt = \\ & (g'_f(t_u) + \Delta g'_f(t_u)) (1 - \exp(-(T - t_u)/T)) T \end{aligned} \quad (10)$$

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  $N = 61$  exponentials

$$g'_f(\tau) \approx \sum_{i=1}^N \alpha_i e^{-\tau/\tau_i}, \quad (11)$$

with logarithmically spaced time-scales  $\tau_i$  ranging from 1 ps to 1  $\mu$ s, and then calculating the spectral density of this fit based on the Fourier

transformation<sup>41</sup>

$$j(\omega) = 2(1 - S_{\text{CH}}) \sum_{i=1}^N \alpha_i \frac{\tau_i}{1 + \omega^2 \tau_i^2}. \quad (12)$$

The  $R_1$  rate of a given C-H pair 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  $N_1$  measurements per pair were then assumed independent: Their mean gave the  $R_1$  rate of the C-H pair, 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 pairs. When several carbons contribute to a single experimental  $R_1$  rate due to the overlapping peaks (for example in  $\text{C}_2$  carbon in acyl chains and  $\gamma$  carbons), the  $R_1$  from simulations was obtained as an average over all overlapping carbons. The segment-wise error estimates were obtained by standard error propagation, starting from the uncertainties of the  $R_1$  rates of the C-H pairs.

To gain some qualitative insight on the time scales at which the main contributions to the  $R_1$  rates arise, we also calculated 'cumulative'  $R_1$  rates,  $R_1(\tau)$ , which contained contributions of the sum in Eq. (12) for which  $\tau_i < \tau$ . Note that here the  $g'_f$  averaged over lipids was used; therefore, the 'cumulative'  $R_1(\tau \rightarrow \infty)$  does not necessarily have exactly the same numerical value as the actual  $R_1$ .

Finally, we note that the fit of Eq. (11) provides an alternative to estimating  $\tau_e$ , because

$$\tau_e = \int_0^\infty g'_f(\tau) d\tau \approx \sum_{i=1}^N \alpha_i \tau_i. \quad (13)$$

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  $\tau_e$ , while the sum of Eq. (13) 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  $\tau_e$  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 Fig. 2 for results) and then proceed to cover a wider range of experimentally, biologically, and computationally relevant conditions, where we explore how the dynamics change when cholesterol is added to the bilayer (Table 2 and Fig. 4) and when hydration level is reduced (Table 3 and Fig. 5).

One should keep in mind that none of the force fields we study produces all the C–H bond order parameters,  $S_{\text{CH}}$ , within experimental accuracy.<sup>32</sup> This means that the structural ensembles simulated do not exactly match the structural ensemble occurring in reality. Consequently, the  $\tau_e$  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 $\tau_e$ at standard conditions.

The top panels of Fig. 2 compare the  $\tau_e$  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  $\tau_e$  profile: Dynamics slows down towards the glycerol backbone in both the headgroup and the tails. Quantitatively, MD has a tendency to produce too slow dynamics in the glycerol region. CHARMM36 and Slipids show the best overall performance—although the  $\tau_e$  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.<sup>41</sup> It also is in line with the insufficient conformational sampling of glycerol backbone torsions observed in 500-ns-long CHARMMc32b2<sup>87,88</sup> simulations of a DOPC

lipid.<sup>89</sup>

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  $\tau_e$  at the glycerol backbone by MD would get worse as  $\tau_e$  increases at decreasing temperature.

#### $R_1$ rates at standard conditions.

The lower on the right side of Fig. 2 compare experimental and simulated  $R_1$  rates under the same conditions as for the  $\tau_e$  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), and certain that they all miss (that  $R_1$  rates for the oleoyl segments C8, C10, and C11 are all roughly equal).

Quantitatively, there are a few cases where both  $R_1$  and  $\tau_e$  (almost) match experiments, suggesting (almost) correct rotational dynamics at all relevant time scales. For example, Slipids performs well at the  $\beta$  and  $\alpha$  segments; CHARMM36 for the  $g_3$ ,  $g_2$ , C2 and C3; Lipid14, ECC, and MacRog for the oleoyl double bond.

Notably, there are also instances where the  $R_1$  comparison distinctly differs from what is seen for  $\tau_e$ : Some models that do very well for  $\tau_e$ , do rather poorly for  $R_1$ . Conversely, a matching  $R_1$  can be accompanied by a larger-than-experimental  $\tau_e$ . To appreciate such differences, recall that in order to capture our experimental  $R_1$  rates (measured at 125 MHz) a force field has to have correct rotational dynamics at the  $(2\pi \times 125 \text{ MHz})^{-1} \approx 1 \text{ ns}$  time scale, whereas  $\tau_e$  reflects all the sub- $\mu\text{s}$  time scales (Fig. 1).

MacRog for the  $\beta$ ,  $\alpha$ , and  $g_1$  segments provides a prominent example where the  $R_1$  rates are well reproduced, but  $\tau_e$  times systematically overestimated. Such a combination suggests that MD does well at the 1 ns scale, but has too slow long-time dynamics.

The opposite—where  $\tau_e$  matches experiments, but  $R_1$  does not—is demonstrated by CHARMM36 for  $\beta$  and  $\alpha$ . Therein a cancel-

lation of error occurs in  $\tau_e$ : 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  $\tau_e$ , we proceed to study this shortcoming on the headgroup  $R_1$  rates in some more detail.

## Dynamics of headgroup segments in CHARMM36.

Figure 3A zooms in on the headgroup ( $\gamma$ ,  $\beta$ ,  $\alpha$ ) segments, whose  $\tau_e$  were not clearly visible on the scale of Fig. 2. For  $\beta$ ,  $\alpha$ , CHARMM36 matches the experimental  $\tau_e$ , but overestimates  $R_1$ , while Slipids captures both measurables near perfectly. No force field provides both  $\tau_e$  and  $R_1$  for  $\gamma$ .

To investigate where the differences between force fields arise, we visualize the 'cumulative'  $R_1(\tau)$  in Fig. 3B. It is obtained, as detailed in Methods, by including in the sum of Eq. (12) only terms with  $\tau_i < \tau$ . Consequently, at  $\tau \rightarrow \infty$  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  $\gamma$  (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 ( $\alpha_i$  in Eq. (11)) in Fig. 3C: The earlier the non-zero weights occur, the smaller is the resulting  $R_1$ .

For the  $\beta$  and  $\alpha$  segments, Fig. 3B shows that the main contribution to  $R_1$  rates arises from processes between 200 ps and 2 ns. As CHARMM36 has the largest weights of all models in this window (Fig. 3C), it overestimates  $R_1$ . In contrast, Slipids, which has simultaneously  $R_1$  and  $\tau_e$  correct, has its largest weights at  $\tau < 200$  ps. Indeed, considerable weights at short time scales ( $< 10$  ps in  $\alpha$  for Lipid14, ECC, Berger, CHARMM36) and at long time scales ( $> 10$  ns in both  $\beta$  and  $\alpha$  for MacRog and Berger) do not manifest at all in the  $R_1$

rates. However, the latter contribute heavily on  $\tau_e$ , which is thus considerably overestimated by MacRog and Berger (Fig. 2).

What are the motions in the 0.2–2 ns window that are over-presented in CHARMM36? Identifying them and speeding them up would improve the model dynamics. 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.

Cholesterol is essential component in cell membranes with various biological functions. While cholesterol is well known to order the acyl chains in cell membranes, its effect on headgroup is more controversial.<sup>60,90</sup> Lipid headgroups are proposed to reorganize to shield cholesterol from interaction with water.<sup>90</sup> However, while acyl chains exhibit substantial ordering, no significant conformational changes in headgroup are observed in NMR experiments upon addition of even 50% of cholesterol, suggesting that acyl chain and headgroup regions behave essentially independently.<sup>32,60</sup> On the other hand, the headgroups could shield water-cholesterol interactions without changes in internal conformational ensemble by reorienting headgroups laterally on top of cholesterol. In this case, one would expect the dynamics of headgroup carbons to be affected by cholesterol.

Figure 4A (top panels) depicts the experimental  $\tau_e$  in pure POPC bilayers as well as ones containing 50% cholesterol. The effective correlation times at the glycerol backbone slow down markedly when cholesterol is added. Tail segment dynamics slow down too, with most detectable effect close to the glycerol backbone. In stark contrast, however, the  $\tau_e$  of headgroup segments ( $\gamma$ ,  $\beta$ ,  $\alpha$ ) are unaffected by cholesterol. Furthermore, cholesterol induces no measurable change in the headgroup  $\beta$  and  $\alpha$  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  $\gamma$ .



All the force fields investigated qualitatively reproduce the increase in  $\tau_e$  (see Fig. 4B): Slipids and CHARMM36 give the decent magnitude estimates, while MacRog clearly overestimates the changes at the glycerol, C2, and C3 carbons. Notably, MacRog erroneously predicts slow down also for the headgroup  $\beta$  and  $\alpha$  carbons, for which experiments detect no change.

**3.Markus: see new error bars for Macrog, would you say the sentence above is still true?** Note that, while CHARMM36 correctly shows no change in  $\tau_e$  of the  $\gamma$ ,  $\beta$ , and  $\alpha$  carbons, it predicts an erroneous  $\Delta R_1$  for all three, indicating some inaccuracies in the headgroup rotational dynamics. Such inaccuracies might be reflected in the recent findings<sup>91</sup> (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; **4.Samuli wants to remove this, why? such arrested rotations could manifest on  $\tau_e$  and  $R_1$ .** Interestingly, the tail  $\Delta R_1$  seem to be qualitatively reproduced by all three all-atom force fields, whereas Berger fails to capture the trend at the oleoyl double bond.

In summary, the experiments suggest that acyl chain ordering upon cholesterol addition is accompanied with slower internal dynamics in hydrophobic core and glycerol backbone region, while headgroup dynamics is almost intact even with 50% of cholesterol, supporting the idea<sup>7</sup> that acyl chains and headgroup can respond almost independently to change in conditions and composition. In line with general picture from order parameters,<sup>7</sup> MD simulations capture the changes in acyl chain region rather well, but changes on and near glycerol backbone region can be overestimated.

## Effect of dehydration.

Understanding the impact of dehydration on the structure and dynamics of lipid bilayers is of considerable biological interest. Most prominently, membrane fusion is always preceded by removal of water between the opposing membrane and dehydration therefore may considerably affect the fusion characteristics such as the rate. Lipid bilayers in dehydrated states are also found, e.g., in skin tissue.

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  $\tau_e$  measured for POPC at full hydration and for DMPC (1,2-dimyristoyl-sn-glycero-3-phosphocholine) at 13 waters per lipid.

The  $\tau_e$  are the same within experimental accuracy, which suggests two conclusions. Firstly, the headgroup ( $\gamma$ ,  $\beta$ ,  $\alpha$ )  $\tau_e$  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  $\tau_e$  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.

Owing to the uncertainties, we only point out the qualitative trends. For all carbons in the headgroup and glycerol segments, the simulated  $\tau_e$  indicates slow down upon dehydration. This is manifested in the increase in the magnitude of the error estimate (cf. the Berger data for  $\beta$  and  $\alpha$ ) as well as in the increase of the lower limit of the error. For CHARMM36 the lower error estimates stay almost constant all the way until 7 w/l, whereas for Berger and MacRog they indicate a retardation of the dynamics starting already from  $\sim 20$  w/l.

These simulational findings suggest that experiments reducing hydration levels below 10 w/l would also show an increase in  $\tau_e$ . This prediction is in line with the exponential slowdown of the headgroup conformational dynamics upon dehydration that was indicated by <sup>2</sup>H-NMR  $R_1$  measurements of DOPC bilayers:  $R_1 \sim \exp(-n_{w/l}/4)$ .<sup>92</sup> The slowdown was attributed to the reduction in the effective volume available for the headgroup<sup>92</sup> owing to its tilt towards the membrane upon dehydration; the tilt is observed via changes of the lipid headgroup order parameters,<sup>93</sup> and is qualitatively reproduced by all the simulation models.<sup>32</sup>

Figure 5C shows a collection of experimental  $^{13}\text{C}$ -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,<sup>42</sup> and POPC at 20 and 5 w/l<sup>43</sup> are shown. Experimentally, an increasing trend with decreasing hydration is observed for all the segments, indicating changes of headgroup dynamics at short ( $\sim 1$  ns) time scales. Interestingly, only CHARMM36 captures this, whereas Berger and MacRog give decreasing  $R_1$  rates for  $\beta$  and  $\alpha$ .

The characteristics of the slow down discussed here are of significance not only for computational studies of intermembrane interactions, such as membrane fusion, but also when simulating a bilayer (stack) under low hydration. Slower dynamics imply that longer simulation times are needed for equilibration, for reliably quantifying the properties of the bilayers, and for observing rare events.

## 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  $^{13}\text{C}$ -NMR effective correlation time ( $\tau_e$ ) 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 perfectly reproduces all the experimental data, CHARMM36 and Slipids have an overall impressive  $\tau_e$ . This is particularly true for CHARMM36, as it is also known to well represent the experimental conformational ensemble.<sup>32</sup> That said, we find that CHARMM36 struggles with the balance of dynamics in the headgroup region: The  $R_1$  rates, sensitive for  $\sim 1$ -ns processes, are too high for the  $\gamma$ ,  $\beta$ , and  $\alpha$  segments (Fig. 3).

In addition to standard conditions, we ex-

plored how the dynamics react to addition of cholesterol and removal of water. MD qualitatively captures the slow-down of conformational dynamics in the tail and glycerol regions when cholesterol is mixed into a POPC bilayer; however, some force fields erroneously predict changes also in the short-scale dynamics of the headgroup carbons (Fig. 4). Upon reducing the water content, MD exhibits slow-down of headgroup and backbone dynamics below  $\sim 10$  waters per lipid in qualitative agreement with experimental data. That said, only CHARMM36 qualitatively captures the experimentally detected increase of  $R_1$  rates upon dehydration.

By gathering a set of  $^{13}\text{C}$ -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,<sup>32</sup> realistic dynamics will be an essential part of developing MD into a true computational microscope.

Importantly, this work demonstrates the potential of open data in achieving this at a reduced computational and labor cost. A pool of well indexed and documented data provides an ideal platform for automation, and thus has the potential to facilitate faster progress in pinpointing the typical failures of existing models, in identifying key differences in models describing chemical variations under the same molecule type (eg. different lipid headgroups), and developing better models through machine learning approaches.

## Acknowledgement

H.A gratefully acknowledges the support from Osk. Huttunen’s foundation, Finnish Academy of Science and Letters (Foundations’ Post Doc Pool), Instrumentarium Science Foundation, and the AvH Foundation.

## 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) Levitt, M. Growth of novel protein structural data. *Proceedings of the National Academy of Sciences* **2007**, *104*, 3183–3188.
- (6) Brzezinski, D.; Dauter, Z.; Minor, W.; Jaskolski, M. On the evolution of the quality of macromolecular models in the PDB. *The FEBS Journal* **2020**, *287*, 2685–2698.
- (7) Harris, D. J. Can you bank on GenBank? *Trends in Ecology & Evolution* **2003**, *18*, 317–319.
- (8) Steinegger, M.; Salzberg, S. L. Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biology* **2020**, *21*, 115.
- (9) Rigden, D. J.; Fernández, X. M. The 27th annual Nucleic Acids Research database issue and molecular biology database collection. *Nucleic Acids Research* **2020**, *48*, D1–D8.
- (10) Simpkin, A. J.; Thomas, J. M. H.; Simkovic, F.; Keegan, R. M.; Rigden, D. J. Molecular replacement using structure predictions from databases. *Acta Crystallographica Section D Structural Biology* **2019**, *D75*, 1051–1062.
- (11) Leelananda, S. P.; Lindert, S. Using NMR Chemical Shifts and Cryo-EM Density Restraints in Iterative Rosetta-MD Protein Structure Refinement. *Journal of Chemical Information and Modeling* **2020**, *60*, 2522–2532.
- (12) Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710.
- (13) Huang, P.-S.; Boyken, S. E.; Baker, D. The coming of age of de novo protein design. *Nature* **2016**, *537*, 320.
- (14) Westbrook, J. D.; Burley, S. K. How Structural Biologists and the Protein Data Bank Contributed to Recent FDA New Drug Approvals. *Structure* **2019**, *27*, 211–217.
- (15) Martinez-Mayorga, K.; Madariaga-Mazon, A.; Medina-Franco, J.; Maggiora, G. The impact of chemoinformatics on drug discovery in the pharmaceutical industry. *Expert Opinion on Drug Discovery* **2020**, *15*, 293–306.
- (16) Senderowitz, H.; Tropsha, A. Materials Informatics. *J. Chem. Inf. Model.* **2018**, *58*, 2377–2379.
- (17) Wan, X.; Feng, W.; Wang, Y.; Wang, H.; Zhang, X.; Deng, C.; Yang, N. Materials Discovery and Properties Prediction in Thermal Transport via Materials Informatics: A Mini Review. *Nano Letters* **2019**, *19*, 3387–3395.
- (18) Perez-Riverol, Y.; Zorin, A.; Dass, G.; Vu, M.-T.; Xu, P.; Glont, M.; Vizcaíno, J. A.; Jarnuczak, A. F.; Petryszak, R.; Ping, P.; Hermjakob, H. Quantifying the impact of public omics data. *Nature Communications* **2019**, *10*.
- (19) Feng, Z.; Verdigué, N.; Costanzo, L. D.; Goodsell, D. S.; Westbrook, J. D.; Burley, S. K.; Zardecki, C. Impact of the Protein Data Bank Across Scientific Disciplines. *Data Science Journal* **2020**, *19*, 25.
- (20) Feig, M.; Abdullah, M.; Johnsson, L.; Pettitt, B. M. Large scale distributed data repository: design of a molecular dynamics trajectory database. *Future Generation Computer Systems* **1999**, *16*, 101–110.

- (21) Tai, K.; Murdock, S.; Wu, B.; Ng, M. H.; Johnston, S.; Fangohr, H.; Cox, S. J.; Jeffreys, P.; Essex, J. W.; P. Sansom, M. S. BioSimGrid: towards a worldwide repository for biomolecular simulations. *Organic & Biomolecular Chemistry* **2004**, *2*, 3219–3221.
- (22) Silva, C. G.; Ostropeytsky, V.; Loureiro-Ferreira, N.; Berrar, D.; Swain, M.; Dubitzky, W.; Brito, R. M. M. P-found: The Protein Folding and Unfolding Simulation Repository. 2006 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology. 2006; pp 1–8.
- (23) Hildebrand, P. W.; Rose, A. S.; Tieermann, J. K. S. Bringing Molecular Dynamics Simulation Data into View. *Trends in Biochemical Sciences* **2019**, *44*, 902–913.
- (24) Abraham, M. et al. Sharing Data from Molecular Simulations. *Journal of Chemical Information and Modeling* **2019**, *59*, 4093–4099.
- (25) 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.
- (26) Hospital, A.; Battistini, F.; Soliva, R.; Gelpí, J. L.; Orozco, M. Surviving the deluge of biosimulation data. *WIREs Computational Molecular Science* **2020**, *10*, e1449.
- (27) Hospital, A.; Andrio, P.; Cugnasco, C.; Codo, L.; Becerra, Y.; Dans, P. D.; Battistini, F.; Torres, J.; Goñi, R.; Orozco, M.; Gelpí, J. L. BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Research* **2016**, *44*, D272–D278.
- (28) Bekker, G.-J.; Kawabata, T.; Kurisu, G. The Biological Structure Model Archive (BSM-Arc): an archive for in silico models and simulations. *Biophysical Reviews* **2020**, *12*, 371–375.
- (29) Mixcoha, E.; Rosende, R.; Garcia-Fandino, R.; Piñeiro, Á. Cyclo-lib: a database of computational molecular dynamics simulations of cyclodextrins. *Bioinformatics* **2016**, *32*, 3371–3373.
- (30) Rodríguez-Espigares, I. et al. GPCRmd uncovers the dynamics of the 3D-GPCRome. *bioRxiv* **2019**, 839597.
- (31) Miettinen, M. S.; NMRlipids Collaboration; Ollila, O. H. S. LDB: Lipid Database from the NMRlipids Project. *Biophysical Journal* **2019**, *116*, 91a.
- (32) Botan, A. 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.
- (33) Catte, A.; Girysh, M.; Javanainen, M.; Loison, C.; Melcr, J.; Miettinen, M. S.; Monticelli, L.; Määttä, J.; Oganessian, V. S.; Ollila, O. H. S.; Tynkkynen, J.; Vilov, S. Molecular electrometer and binding of cations to phospholipid bilayers. *Phys. Chem. Chem. Phys.* **2016**, *18*, 32560–32569.
- (34) Antila, H.; Buslaev, P.; Favela-Rosales, F.; Ferreira, T. M.; Gushchin, I.; Javanainen, M.; Kav, B.; Madsen, J. J.; Melcr, J.; Miettinen, M. S.; Määttä, J.; Nencini, R.; Ollila, O. H. S.; Piggot, T. J. Headgroup Structure and Cation Binding in Phosphatidylserine Lipid Bilayers. *J. Phys. Chem. B* **2019**, *123*, 9066–9079.
- (35) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci.* **2018**, *115*, E4758–E4766.
- (36) Henriques, J.; Arleth, L.; Lindorff-Larsen, K.; Skepö, M. On the Calculation

- of SAXS Profiles of Folded and Intrinsically Disordered Proteins from Computer Simulations. *J. Mol. Biol.* **2018**, *430*, 2521 – 2539.
- (37) Virtanen, S. I.; Kiirikki, A. M.; Mikula, K. M.; Iwai, H.; Ollila, O. H. S. Heterogeneous dynamics in partially disordered proteins. *Phys. Chem. Chem. Phys.* **2020**, –.
- (38) Ollila, S.; Pabst, G. Atomistic resolution structure and dynamics of lipid bilayers in simulations and experiments. *BBA - Biomembranes* **2016**, *1858*, 2512–2528.
- (39) Antila, H. S.; Wurl, A.; Ollila, O. H. S.; Miettinen, M. S.; Ferreira, T. M. Quasi-uncoupled rotational diffusion of phospholipid headgroups from the main molecular frame. *arXiv preprint* **2020**, arXiv:2009.06774v1.
- (40) Han, J.; Pluhackova, K.; Böckmann, R. A. The Multifaceted Role of SNARE Proteins in Membrane Fusion. *Frontiers in Physiology* **2017**, *8*, 5.
- (41) Ferreira, T. M.; Ollila, O. H. S.; Pigliapochi, R.; Dabkowska, A. P.; Topgaard, D. Model-free estimation of the effective correlation time for C–H bond reorientation in amphiphilic bilayers:  $^1\text{H}$ – $^{13}\text{C}$  solid-state NMR and MD simulations. *The Journal of Chemical Physics* **2015**, *142*, 044905.
- (42) 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.
- (43) 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/ $\text{C}_{12}\text{EO}_4$  Model Membranes. *Biophys. J.* **1995**, *68*, 1960–1965.
- (44) Berger, O.; Edholm, O.; Jähnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophysical Journal* **1997**, *72*, 2002 – 2013.
- (45) Bachar, M.; Brunelle, P.; Tieleman, D. P.; Rauk, A. Molecular Dynamics Simulation of a Polyunsaturated Lipid Bilayer Susceptible to Lipid Peroxidation. *The Journal of Physical Chemistry B* **2004**, *108*, 7170–7179.
- (46) Bacle, A.; Fuchs, P. F. Berger pure POPC MD simulation (300 K - 300ns - 1 bar). 2018; <https://doi.org/10.5281/zenodo.1402417>.
- (47) 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.
- (48) Papadopoulos, C.; Fuchs, P. F. CHARMM36 pure POPC MD simulation (300 K - 300ns - 1 bar). 2018; <http://dx.doi.org/10.5281/zenodo.1306800>, DOI: 10.5281/zenodo.1306800.
- (49) 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 cholesterol hemisuccinate. *Biochim. Biophys. Acta* **2015**, *1848*, 422 – 432.
- (50) Milan Rodriguez, P.; Fuchs, P. F. MacRog pure POPC MD simulation (300 K - 500ns - 1 bar). **2020**,
- (51) 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.

- (52) 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.
- (53) 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.
- (54) Javanainen, M. POPC with 0, 10, 20, and 30 mol-Slipids force field. 2016; <http://dx.doi.org/10.5281/zenodo.3243328>.
- (55) 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.
- (56) Melcr, J. Simulations of POPC lipid bilayer in water solution at various NaCl, KCl and CaCl<sub>2</sub> concentrations using ECC-POPC force field. **2019**,
- (57) 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.
- (58) 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>.
- (59) 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.
- (60) 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 <sup>1</sup>H–<sup>13</sup>C NMR and MD simulation studies. *Phys. Chem. Chem. Phys.* **2013**, *15*, 1976–1989.
- (61) Ollila, O. H. S. MD simulation trajectory and related files for POPC/cholesterol (50 molmodified Höltje, Gromacs 4.5). **2014**,
- (62) Javanainen, M. POPC with 0, 10, 20, and 30 mol-Charmm36 force field. 2016; <https://doi.org/10.5281/zenodo.3237420>, DOI: 10.5281/zenodo.3237420.
- (63) 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.
- (64) Javanainen, M. POPC with 40 and 50 mol-force field. 2016; <https://doi.org/10.5281/zenodo.3238157>, DOI: 10.5281/zenodo.3238157.
- (65) Javanainen, M.; Kulig, W. POPC/Cholesterol @ 310K. 0, 10, 40, 50 and 60 mol-cholesterol. Model by Maciejewski and Rog. **2015**,
- (66) 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.
- (67) 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**,
- (68) 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.
- (69) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=24 w/l. 2015; DOI: 10.5281/zenodo.16289.
- (70) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=16 w/l. 2015; DOI: 10.5281/zenodo.16292.

- (71) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=12 w/l. 2015; DOI: 10.5281/zenodo.16293.
- (72) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=4 w/l. 2015; DOI: 10.5281/zenodo.16295.
- (73) 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.
- (74) Ollila, O. H. S.; Miettinen, M. MD simulation trajectory and related files for POPC bilayer (CHARMM36, Gromacs 4.5). 2015; <http://dx.doi.org/10.5281/zenodo.13944>, DOI: 10.5281/zenodo.13944.
- (75) 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.
- (76) 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.
- (77) 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.
- (78) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **2011**, *32*, 2319–2327.
- (79) Richard J. Gowers;; Max Linke;; Jonathan Barnoud;; Tyler J. E. Reddy;; Manuel N. Melo;; Sean L. Seyler;; Jan Domański;; David L. Dotson;; Sébastien Buchoux;; Ian M. Kenney;; Oliver Beckstein, MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference. 2016; pp 98 – 105.
- (80) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.
- (81) Schlenkrich, M.; Brickmann, J.; MacKerell, A. D.; Karplus, M. *Biological Membranes*; Springer, 1996; pp 31–81.
- (82) 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.
- (83) 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.
- (84) Huang, J.; Feigenson, G. W. A Microscopic Interaction Model of Maximum Solubility of Cholesterol in Lipid Bilayers. *Biophysical Journal* **1999**, *76*, 2142 – 2157.
- (85) Leeb, F.; Maibaum, L. Spatially Resolving the Condensing Effect of Cholesterol in Lipid Bilayers. *Biophysical Journal* **2018**, *115*, 2179 – 2188.
- (86) Ulrich, A.; Watts, A. Molecular response of the lipid headgroup to bilayer hydration monitored by 2H-NMR. *Biophys. J.* **1994**, *66*, 1441 – 1449.
- (87) 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.

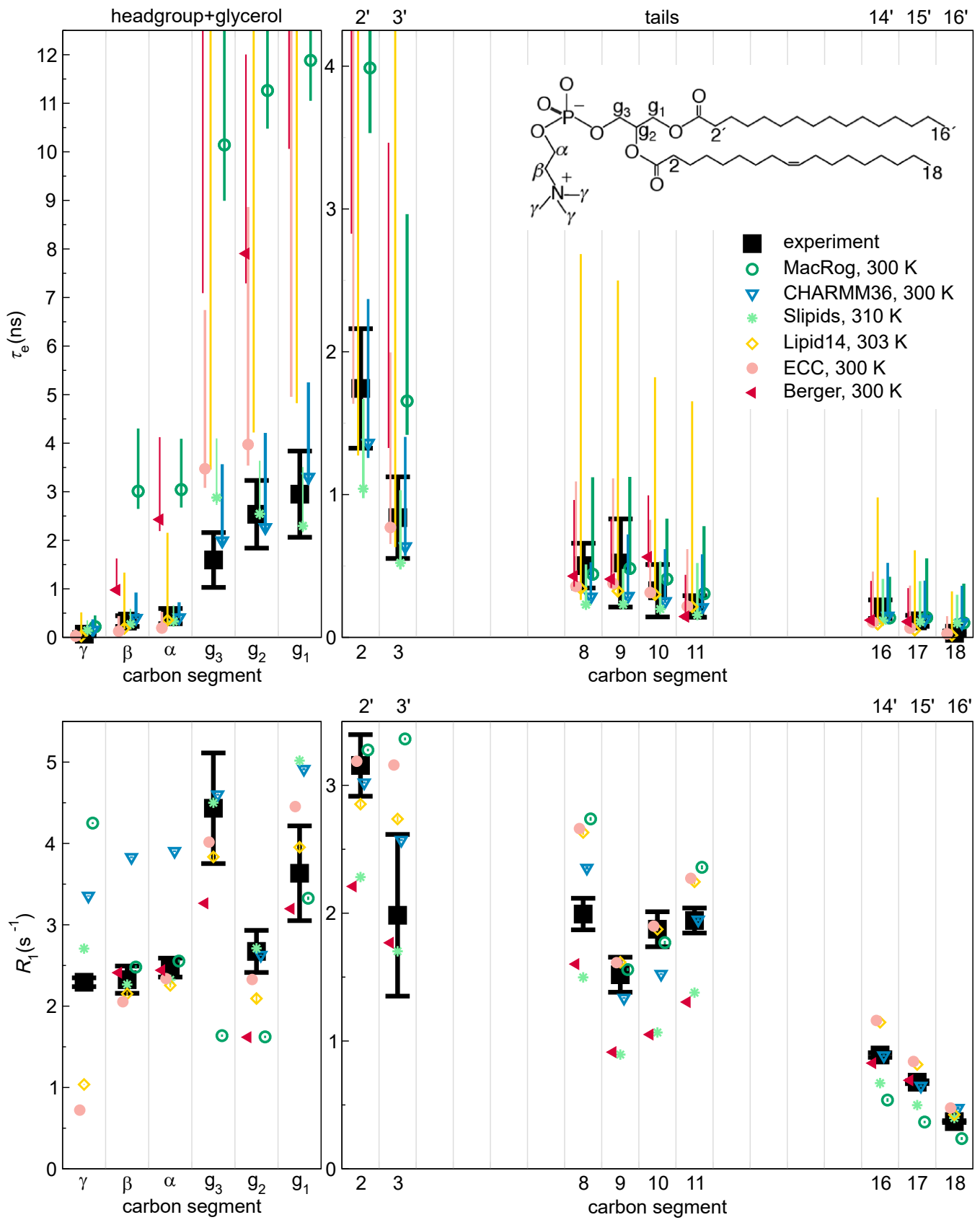


Figure 2: Effective correlation times ( $\tau_e$ , top panels) and  $R_1$  rates (bottom panels) in experiments (black) and MD simulations (colored) of POPC bilayers in  $L_\alpha$  phase under full hydration. Inset on the top right shows the POPC structure and carbon segment labeling. Each plotted value 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 results are only shown for the segments where experimental data is available. For  $\tau_e$ , a simulation data point indicates the average over C–H bonds; however, if  $\tau_e$  could not be determined for all bonds,



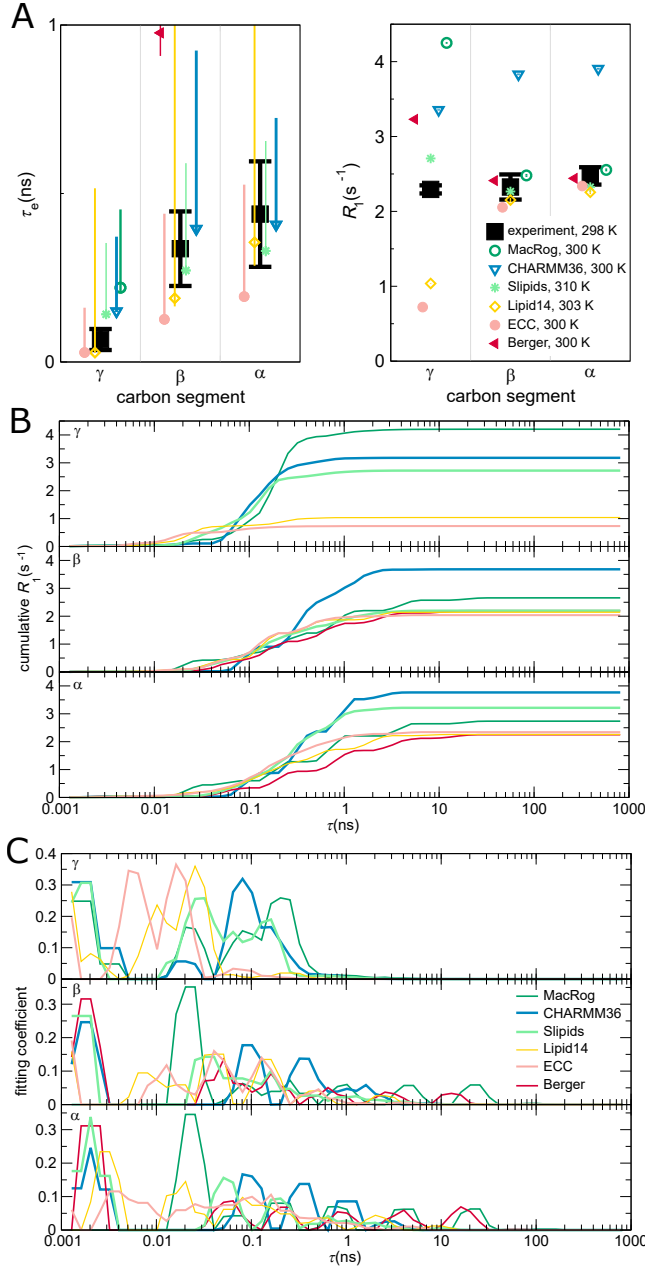


Figure 3: Contributions to the dynamics of the headgroup segments. (A) Zoom on the headgroup  $\tau_e$  (left panel) and  $R_1$  (right). (B) 'Cumulative'  $R_1$  (see Methods for definition) of the  $\gamma$  (top panel),  $\beta$  (middle), and  $\alpha$  (bottom) segments. (C) Prefactor weights  $\alpha_i$  from Eq. (11) of  $\gamma$  (top),  $\beta$  (middle), and  $\alpha$  (bottom). In B and C, a sliding average over 3 neighboring data points is shown.

**2.Note: samuli wanted to reverse B-C panel order**

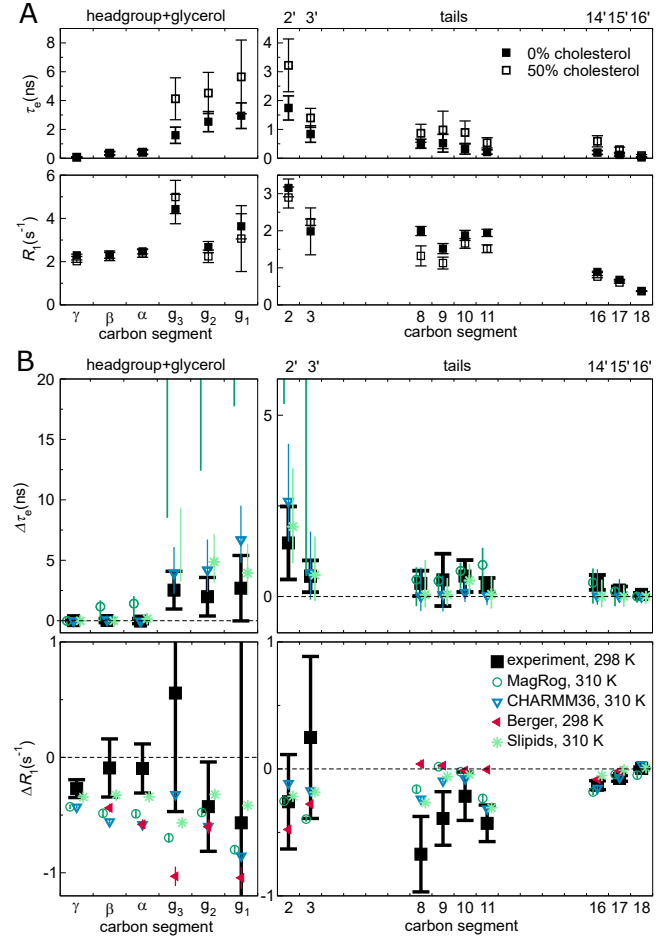


Figure 4: Effect of bilayer cholesterol content. (A) The experimental effective correlation times  $\tau_e$  (top panels) and  $R_1$  rates (bottom) in a pure POPC bilayer and in a bilayer containing 50% cholesterol. The data were measured at 298 K and full hydration, see Ref. 39 for further details. (B) The change in  $\tau_e$  ( $\Delta\tau_e$ , top panels) and  $R_1$  ( $\Delta R_1$ , 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.

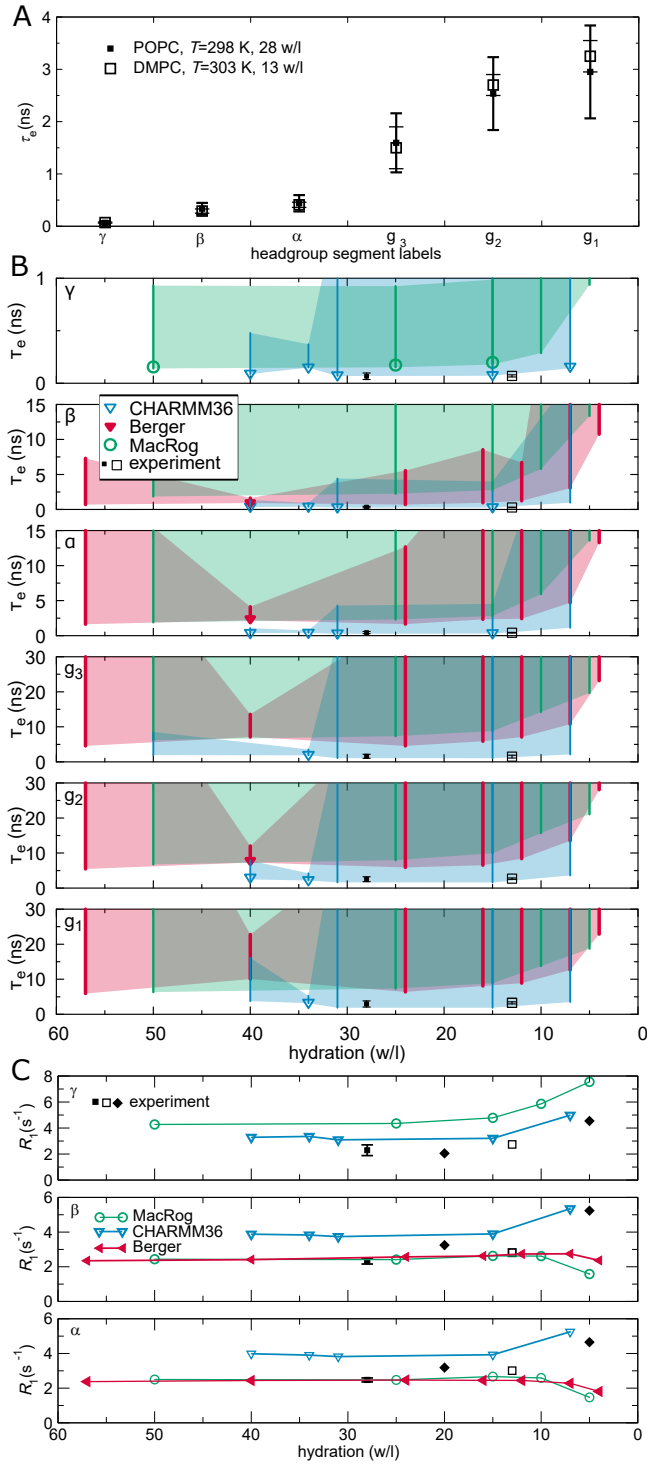


Figure 5: Effect of drying on effective correlation times in headgroup and glycerol backbone. (A) Experimental  $\tau_e$  for DMPC (from Ref. 42) at low hydration do not significantly differ from the  $\tau_e$  for POPC at full hydration (from Ref. 39). (B) Calculated  $\tau_e$  for POPC at decreasing hydration in three MD models. Note that three Berger data points (4, 12, 16, and 24 w/l) are from DLPC bilayers. Symbols give the mean of segment hydrogens, if  $\tau_e$  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

## Graphical TOC Entry

TOC here if needed