

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 fidelity⁶⁻⁸ 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 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

1.Mention GPCRmd¹⁷? SAMULI: I think yes.

Since 2013 the NRMLipids Project (nmrlipids.blogspot.fi) has promoted an 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.¹⁸ While the main focus of the NMRlipids Project has been in conformational ensembles of lipid headgroups and ion binding to lipid membranes,^{18–20} it has also accumulated a collection of atomistic MD simulations of lipid membranes containing hundreds of trajectories (zenodo.org/communities/nmrlipids). These data are 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 described 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 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.^{18,20?} 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 scattering data.[?]

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 detecting 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.

Our comprehensive comparison of dynamics 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 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 usefulness not only for lipid bilayers but also for other biomolecular systems.

2 Methods

2.1 Evaluation of conformational dynamics of lipids against NMR data

We analyze lipid dynamics based on two quantities available from published ¹³C-NMR experiments:^{21–23} 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.²¹ The R_1 rates (or the corresponding T_1 times) have been traditionally used to assess both the conformational dynamics of lipids in experiments^{24–28} and the dynamics produced by MD models in simulations.^{24,26,27,29} 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.

¹³C NMR experiments investigating lipid conformational dynamics take advantage of the fact that the relaxation of ¹³C 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

$$j(\omega) = 2 \int_0^\infty \cos(\omega\tau) g(\tau) d\tau, \quad (1)$$

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

$$g(\tau) = \langle P_2(\vec{\mu}(t) \cdot \vec{\mu}(t + \tau)) \rangle, \quad (2)$$

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

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

where $g_f(\tau)$ characterizes fast decays owing to, for example, the molecular rotations, and $g_s(\tau)$ describes slow decays that originate from, e.g., lipid diffusion. The two components, along with the oscillation due to magic angle spinning at the $\sim\text{kHz}$ region, are depicted in Fig. 1. Correlation time of 4.2 ms has been estimated for multilamellar POPC samples at 300 K for the slow modes, whereas in liquid crystalline lipid bilayers the faster $g_f(\tau)$ decays to a plateau value S_{CH}^2 within a few hundred nanoseconds.²¹ The C–H bond order parameters

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

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 can be quantified via the effective correlation time

$$\tau_e = \int_0^\infty \frac{g_f(\tau) - S_{\text{CH}}^2}{1 - S_{\text{CH}}^2} d\tau. \quad (5)$$

The integrand can be viewed as a reduced and

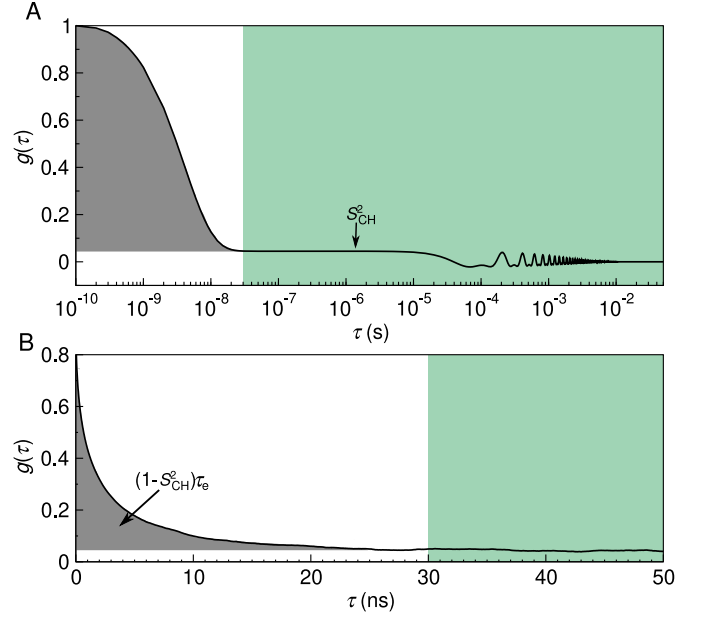


Figure 1: The autocorrelation function $g(\tau)$ a) The fast mode (white background) and the slow mode (shaded green) of the correlation function along with the oscillation owing to magic angle spinning. The fast mode decays to a plateau quantifying the S_{CH} while the slow mode gives the final descent 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 .

normalized correlation function

$$g'_f(\tau) = \frac{g_f(\tau) - S_{\text{CH}}^2}{1 - S_{\text{CH}}^2}. \quad (6)$$

That is, τ_e is defined as the area under $g'_f(\tau)$, as graphically depicted in Fig. 1b. **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 ^{13}C longitudinal magnetization equilibrates. It is defined as

$$R_1 = \frac{d_{\text{CH}}^2 N_{\text{H}}}{20} [j(\omega_{\text{H}} - \omega_{\text{C}}) + 3j(\omega_{\text{C}}) + 6j(\omega_{\text{H}} + \omega_{\text{C}})], \quad (7)$$

where N_{H} is the number of bound hydrogens, ω_{H} and ω_{C} are the Larmor frequencies for ^1H

and ^{13}C , and d_{CH} is the rigid dipolar coupling constant. For the methylene bond, $d_{\text{CH}}/2\pi$ 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 with time-scales near the inverses of these frequencies. Since the Larmor frequencies depend on the field strength used in the NMR measurements, this typically makes R_1 sensitive to ~ 1 – 10 ns time-scales. Importantly, a change in R_1 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 sources referred at the respective figures.

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^{18,19} (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; Table 3 simulations with varying hydration; and Table ?? 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	$T^c(\text{K})$	$t_{\text{anal}}^d(\text{ns})$	files ^e
Berger ^{30,31}	256	10240	300	300	[32]
CHARMM36 ³³	256	8704	300	300	[34]
MacRog ³⁵	128	5120	300	500	[36]
Lipid14 ³⁷	72	2234	303	50	[38]
Slipids ³⁹	200	9000	310	500	[40]
ECC ⁴¹	128	6400	300	300	[42]

^aNumber of POPC molecules.

^bNumber of water molecules.

^cSimulation temperature.

^dTrajectory length used for analysis.

^eReference 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	$T^e(\text{K})$	$t_{\text{anal}}^f(\text{ns})$	files ^g
Berger-POPC-07 ⁴³	0%	0	128	7290	298	50	[44]
/Höltje-CHOL-13 ^{45,46}	50%	64	64	10314	298	50	[47]
CHARMM36 ³³	0%	0	200	9000	310	500	[48]
/CHARMM36 ⁴⁹	50%	200	200	18000	310	500	[50]
MacRog ³⁵	0%	0	128	6400	310	500	[51]
/MacRog ³⁵	50%	64	64	6400	310	500	[51]
Slipids ³⁹	0%	0	200	9000	310	500	[40]
/Slipids ⁵²	50%	200	200	18000	310	500	[40]

^aBilayer cholesterol content (mol %).

^bNumber of cholesterol molecules.

^cNumber of POPC molecules.

^dNumber of water molecules.

^eSimulation temperature.

^fTrajectory length used for analysis.

^gReference for the openly available simulation files.

The simulation data were analyzed using in-house scripts. These are available on GitHub (https://github.com/hsantila/Corrtrimes/tree/master/teff_analysis) along with a 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 NMRlipids 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_{CH} , see Eq. (4), were then calculated with the `calcOrderParameters.py` script that uses the

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 ⁴³	POPC	57	128	7290	298	50	[44]
	POPC	7	128	896	298	60	[53]
Berger ^{30,31}	POPC	40	256	10240	300	300	[32]
Berger-DLPC-13 ⁵⁴	DLPC ^g	24	72	1728	300	80	[55]
	DLPC ^g	16	72	1152	300	80	[56]
	DLPC ^g	12	72	864	300	80	[57]
	DLPC ^g	4	72	288	300	80	[58]
CHARMM36 ³³	POPC	40	128	5120	303	140	[59]
	POPC	34	128	5120	300	500	[36]
	POPC	31	72	2232	303	20	[60]
	POPC	15	72	1080	303	20	[61]
	POPC	7	72	504	303	20	[62]
MacRog ³⁵	POPC	50	288	14400	310	40	[63]
	POPC	25	288	7200	310	50	[63]
	POPC	15	288	4320	310	50	[63]
	POPC	10	288	2880	310	50	[63]
	POPC	5	288	1440	310	50	[63]

^aWater/lipid molar ratio.

^bNumber of lipid molecules.

^cNumber of water molecules.

^dSimulation temperature.

^eTrajectory length used for analysis.

^fReference for the openly available simulation files.

^g1,2-didodecanoyl-sn-glycero-3-phosphocholine.

MDanalysis^{64,65} Python library. The C–H bond correlation functions $g(\tau)$, see Eq. (2), were calculated with Gromacs5.1.4⁶⁶ `gmx rotacf`; 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. (6).

The effective correlation times τ_e were then calculated by integrating $g'_f(\tau)$, ((6)), from $\tau = 0$ until $\tau = t_0$. Here, t_0 is the first time point at which g'_f reached zero, $t_0 = \min\{t \mid g'_f(t) = 0\}$. If g'_f did not reach zero within $t_{\text{anal}}/2$, 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_{CH}^2 . Performing linear error propagation on Eq. (6) 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 ΔS_{CH} was determined as the standard error of the mean of the S_{CH} over the N_l individ-

ual lipids in the system¹⁸ 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 at 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

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

That is, t_l equals the first time point at which the lower error estimate of g'_f reached zero; or $t_l = t_{\text{anal}}/2$, 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 $t_u = \min\{t_0, t_{\text{anal}}/2\}$. 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 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 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 τ_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 τ_e then reads

$$(g'_f(t_u) + \Delta g'_f(t_u)) \int_{t_u}^T \exp(-(t-t_u)/T) dt = (g'_f(t_u) + \Delta g'_f(t_u)) T. \quad (10)$$

The R_1 rates were calculated using Eq. (7). The spectral density $j(\omega)$ was obtained from the normalized correlation function g'_f by fitting it with a sum of $N = 71$ exponentials

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

with logarithmically spaced time-scales τ_i ranging from 0.1 ps to 1 μ s, and then calculating the spectral density of this fit based on the Fourier transformation²¹

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

The R_1 rate of a given C–H pair was first calculated separately for each lipid m (using Eq. (7) with $N_{\text{H}} = 1$, 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 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 pairs. When several carbons contribute to a single experimental R_1 rate due to the overlapping peaks (for example in C_2 carbon in acyl chains and γ carbons), the R_1 from simulations was then 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 bonds.

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 τ_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 τ_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 func-

tion, 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 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 ?? 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.¹⁸ 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 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 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.²¹ It also in line with the insufficient conformational sampling of glycerol backbone torsions observed in 500-ns-long CHARMMc32b2^{67,68} simulations of a DOPC lipid.⁶⁹

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.

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 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), 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 τ_e (almost) match experiments, suggesting (almost) correct rotational dynamics at all relevant time scales. For example, Slipids performs well at the β and α 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 τ_e : Some models that do very well for τ_e , do rather poorly for R_1 . Conversely, a matching R_1 can be accompanied by a larger-than-experimental τ_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 τ_e reflects all the sub- μs time scales (Fig. 1).

MacRog for the β , α , and g_1 segments provides a prominent example where the R_1 rates are well reproduced, but τ_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 τ_e matches experiments, but R_1 does not—is demonstrated 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.

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 β , α , CHARMM36 matches the experimental τ_e , but overestimates R_1 , while Slipids captures both measurables near perfectly. No force field provides both τ_e and R_1 for γ .

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 γ (MacRog, CHARMM36, and Slipids, see Fig. 3A) the major contribution to R_1 arises at $\tau > 50 \text{ ps}$, whereas those underestimating the R_1 (Lipid14 and ECC, see Fig. 2) the major contribution comes from $\tau < 50 \text{ ps}$. This also manifests in the distribution of fitting weights (α_i in Eq. (11)) 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 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 simultane-

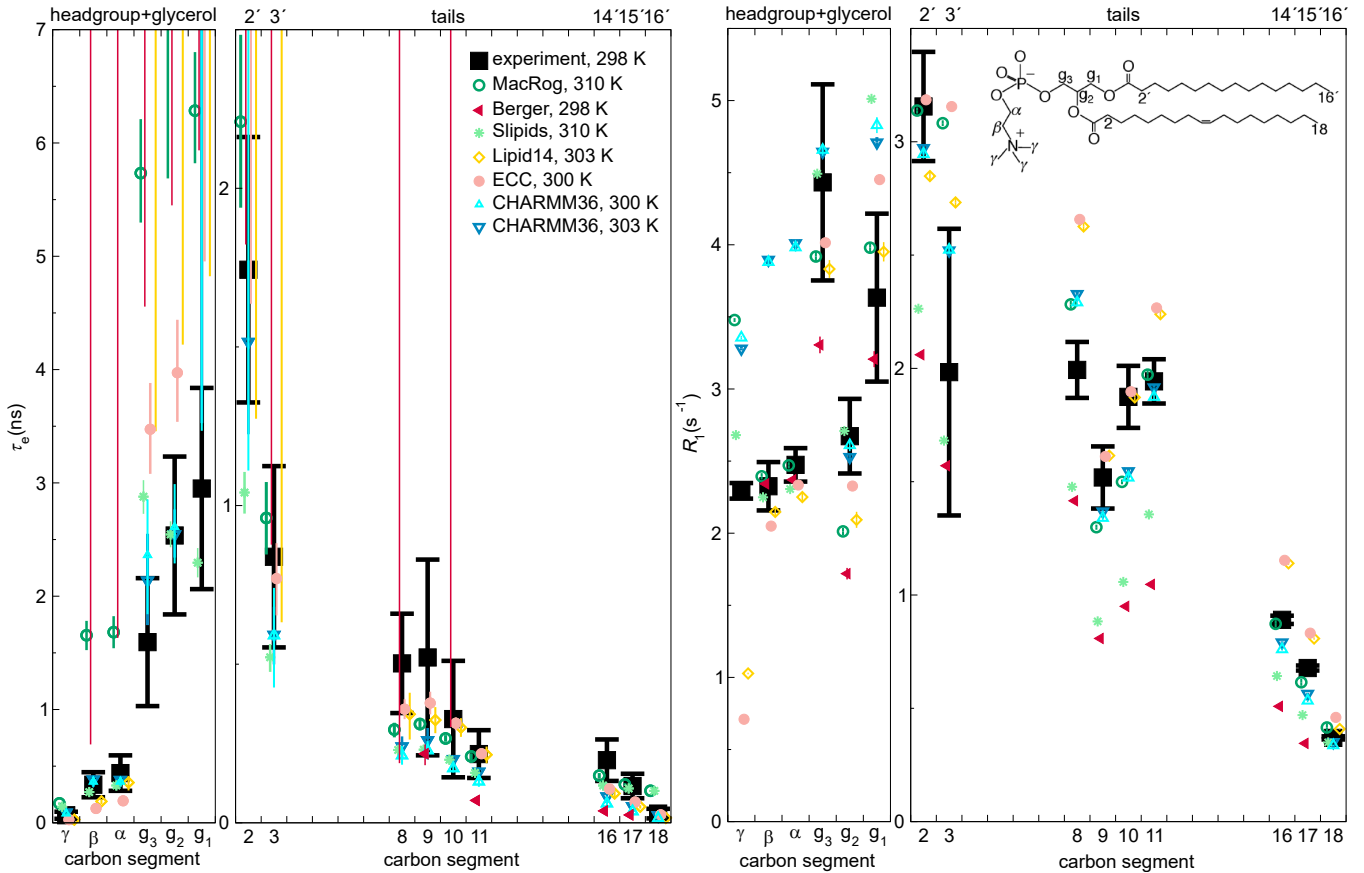


Figure 2: Effective correlation times (τ_e , left panels) and R_1 rates (right panels) in experiments (black) and MD simulations (colored) of POPC bilayers in L_α phase under full hydration. Inset on the right shows the POPC structure and carbon segment labeling. Each plotted value contains contributions 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.

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

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

ously R_1 and τ_e correct, has its largest weights at $\tau < 200$ ps. Indeed, considerable weights at short time scales (< 10 ps in α for Lipid14, ECC, Berger, CHARMM36) and at long time scales (> 10 ns in both β and α for MacRog and Berger) do not manifest at all in the R_1 rates. However, the latter contribute heavily on τ_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.

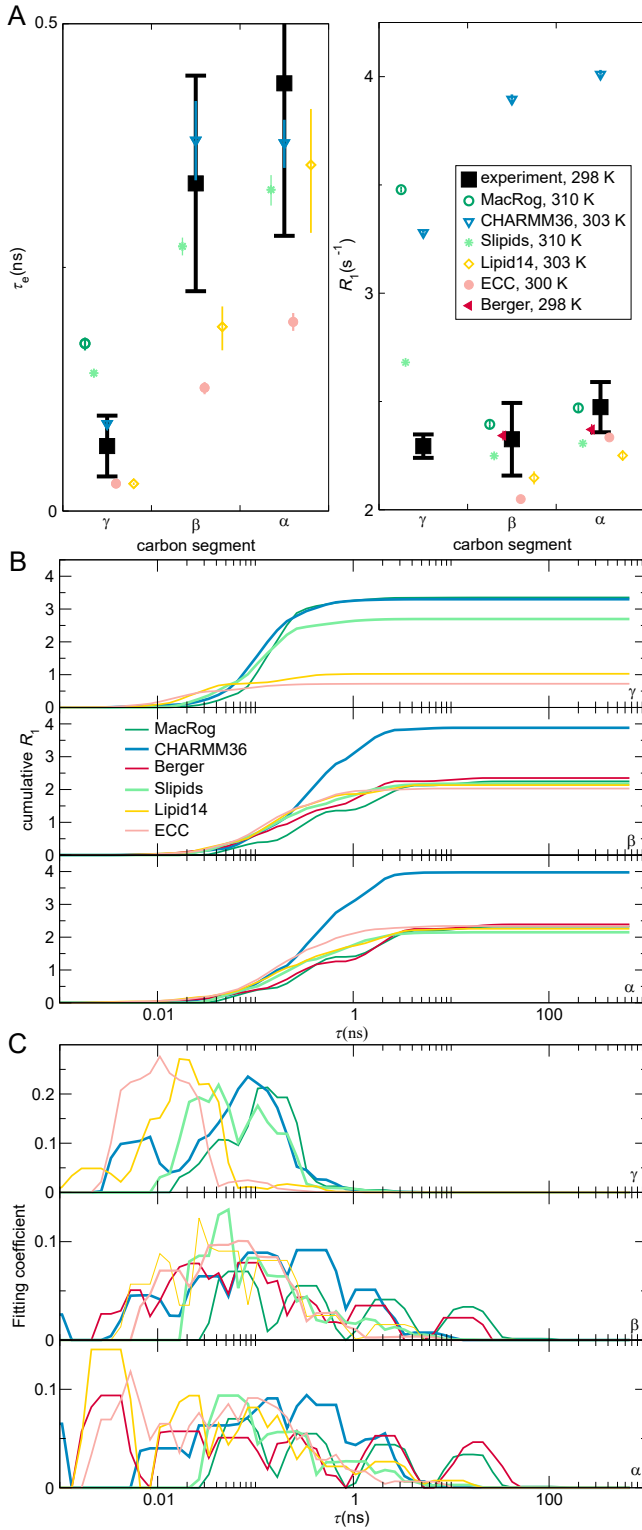


Figure 3: (A) Zoom on the headgroup τ_e (left panel) and R_1 (right). (B) 'Cumulative' R_1 (see Methods for definition) of the γ (top panel), β (middle), and α (bottom) segments. (C) Prefactor weights α_i from Eq. (11) of γ (top), β (middle), and α (bottom). In B and C, a sliding average over 3 neighboring data points is shown.

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.[?] Lipid headgroups are proposed to reorganize to shield cholesterol from interaction with water.[?] However, significant conformational changes in headgroup are not observed in NMR experiments upon addition of even 50% of cholesterol, while acyl chains exhibit substantial ordering, suggesting that acyl chain and headgroup regions behave essentially independently.^{18,46} 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, the dynamics of headgroup dynamics should be affected by cholesterol.

The experimental effective correlation times τ_e (Fig. 4A, top panels) slow down markedly when cholesterol is added. Tail segments slow down too, with most detectable effect close to the glycerol 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 (~ 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 .

6.Markus: look at the new figure, CHARMM36 error estimates pose difficulties here, change C36 data? 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 erroneously predict slow down also for the headgroup β and α carbons, for which experiments detect no change. **7.see new error bars for Macrog** Note that, while CHARMM36 correctly shows no change 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 findings⁷⁰ (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; **8.Samuli wants to remove this, why? such arrested rotations could manifest on τ_e and R_1 .** Interestingly, the tail ΔR_1 seem to be qualitatively reproduced by all three all-atom force fields, whereas Berger fails to capture the change 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 idea⁷ that acyl chains and headgroup can respond almost independently on even large membrane distractions. In line with general picture from order parameters,⁷ MD simulations capture the changes in acyl chain region rather well, but changes in headgroup and glycerol backbone can be overestimated.

Effect of dehydration.

The impact of dehydration on the structure and dynamics of lipid bilayers is of considerable interest as membrane fusion is always preceded by removal of water between the opposing membranes and therefore may affect the rate of fusion. 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 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 accuracy, 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 re-

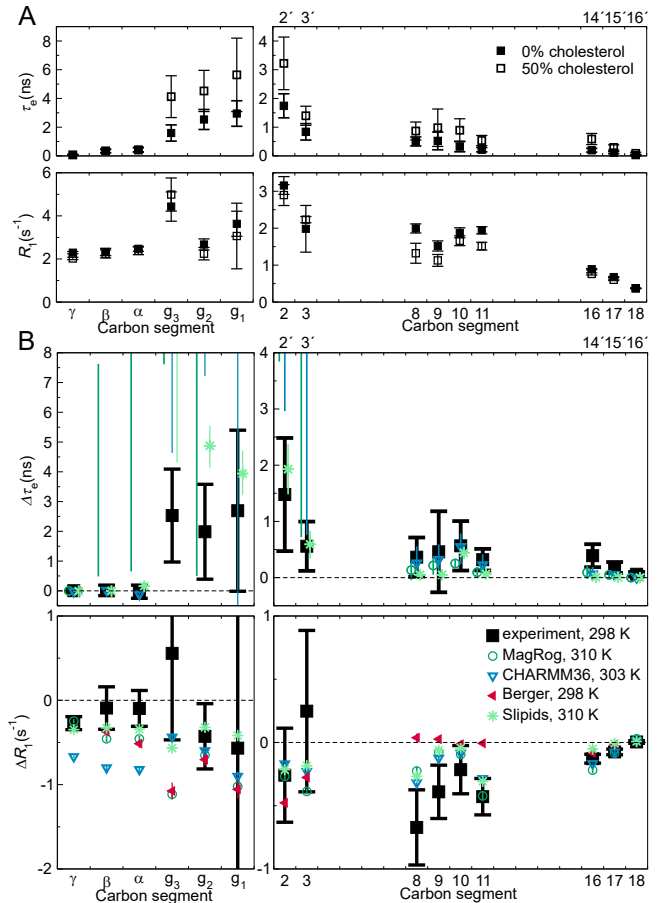


Figure 4: Effect of bilayer cholesterol content. (A) The experimental effective correlation times τ_e (top panels) and R_1 rates (bottom) in a pure POPC bilayer and in a bilayer containing 50% cholesterol. The data were measured at 298 K and full hydration. (B) The change in τ_e ($\Delta\tau_e$, top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

gions.

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 (Ta-

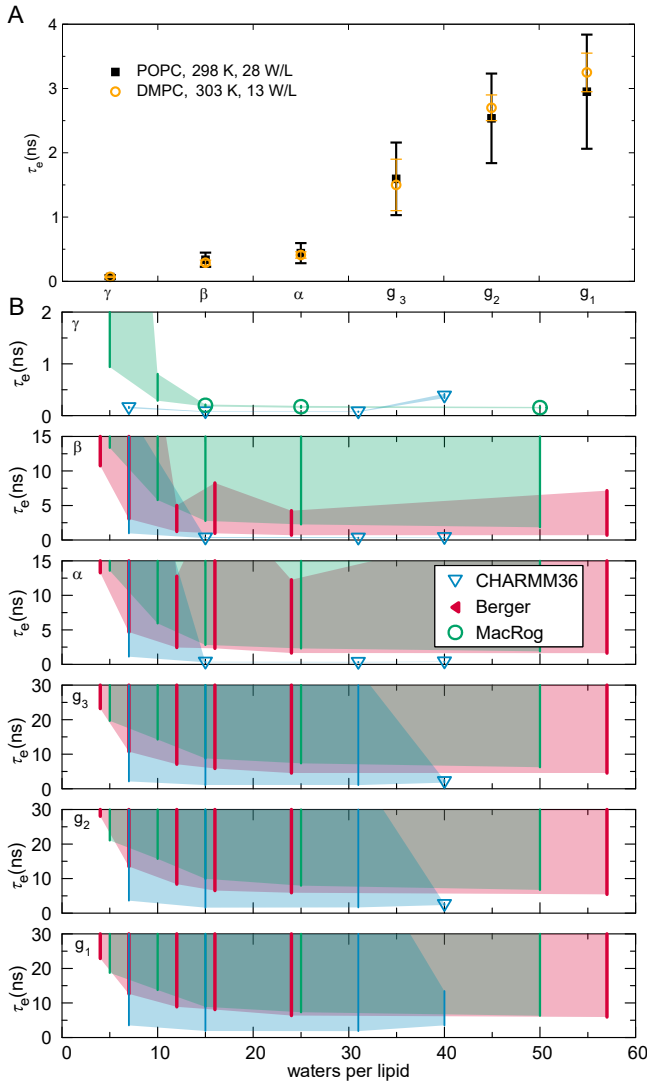


Figure 5: Effect of drying on effective correlation times in headgroup and glycerol backbone. (A) Experimental τ_e for DMPC (from Ref. 22) at low hydration do not significantly differ from the τ_e for POPC at full hydration. (B) Calculated τ_e for POPC at decreasing hydration in three MD models. Note that three Berger data points (4, 12, 16, and 24 w/l) are from DLPC bilayers. 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.

9. How to refer to full hydration POPC data?

ble 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 τ_e indicates slow down upon dehydration. This is manifested in the increase in the magnitude of the error estimate (cf. the Berger data for β and α) as well as in the increase of the lower limit of the error. For CHARMM36 the lower error estimates almost constant all the way until 7 w/l, whereas for Berger and MacRog they indicate 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 slow-down of the headgroup conformational dynamics upon dehydration that was indicated by $^2\text{H-NMR}$ R_1 measurements of DOPC bilayers: $R_1 \sim \exp(-n_{w/l}/4)$.⁷¹ The slowdown was attributed to the reduction in the effective volume available for the headgroup⁷¹ owing to its tilt towards the membrane upon dehydration; the tilt is observed via changes of the lipid headgroup order parameters,⁷² and is qualitatively reproduced by all the simulation models.¹⁸

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,²² and POPC at 20 and 5 w/l²³ 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 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.

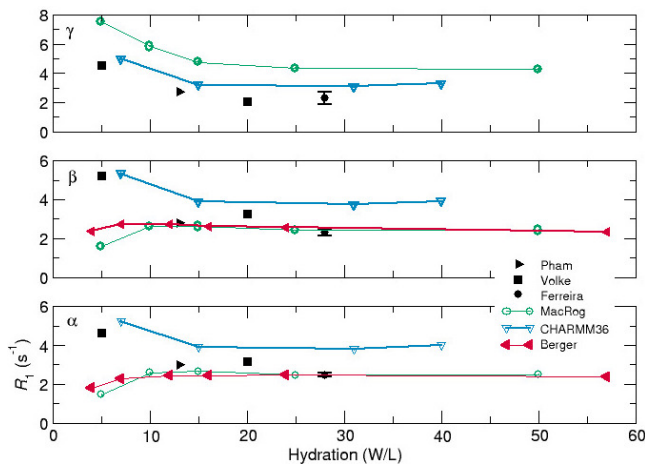


Figure 6: Effect of drying on ^{13}C -NMR R_1 rates of the headgroup segments (at 125 MHz) in experiments and simulations.

10.HA: merge this with the Fig. 5

Effect of cation binding.

11.This section will (possibly) be on ECC simulations and how teff changes with (divalent?) ion binding

Finally, we comment on the response of the MD model dynamics to increasing amounts of salt. Experimentally, the modulation of α and β carbon order parameters upon increasing ion concentration have been used to quantify ion binding to lipid bilayers (the molecular electrometer^{19,73}). The order parameters are constant for POPC bilayers under NaCl addition in experiments, indicating negligible ion binding, but change in the presence increasing amounts of divalent ions.

The molecular electrometer has been used to show that most molecular dynamics force fields overestimate the binding of monovalent ions to PC bilayers:¹⁹ 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. Here, we use existing data from a force field, the ECC model, that is known to have mostly realistic 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.

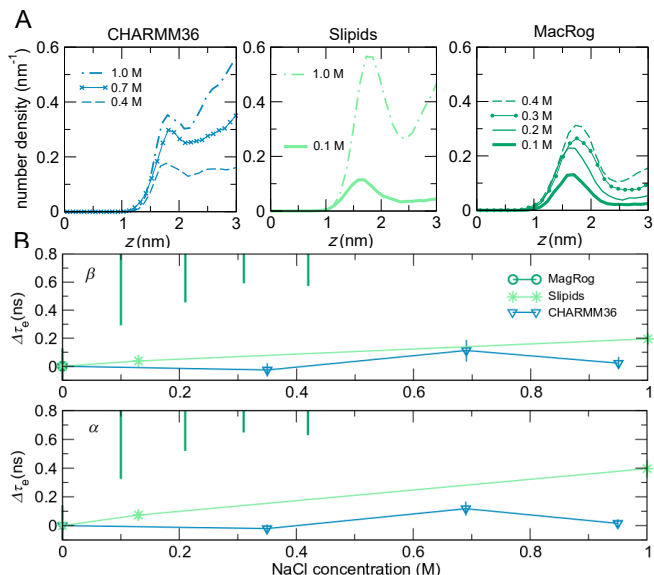


Figure 7: A picture of alpha and beta teff vs bound CA ions in the ECC model

12.analysis running

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, CHARM36 and Slipids have an overall impressive τ_e . This is particularly true for CHARM36, as it is also known to well reproduce the experimental conformational ensemble.¹⁸ That said, we find that CHARM36 struggles with the balance of dynamics 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).

13.Make the point that the 500-ns simulations indicated by Vogel⁶⁹ 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. **14.**

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,¹⁸ 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) Botan, A.; Favela-Rosales, F.; Fuchs, P. F. J.; Javanainen, M.; Kanduč, 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.
 - (19) Catte, A.; Giryck, M.; Javanainen, M.; Loison, C.; Melcr, J.; Miettinen, M. S.; Monticelli, L.; Määttä, J.; Oganessian, 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.
 - (20) Antila, H.; Buslaev, P.; Favela-Rosales, F.; Ferreira, T. M.; Gushchin, I.; Javanainen, M.; Kav, B.; Madsen, J. J.; Melcr, J.; Miettinen, M. S. et al. Headgroup Structure and Cation Binding in Phosphatidylserine Lipid Bilayers. *The Journal of Physical Chemistry B* **2019**, *123*, 9066–9079, PMID: 31574222.
 - (21) 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: ^1H – ^{13}C solid-state NMR and MD simulations. *The Journal of Chemical Physics* **2015**, *142*, 044905.
 - (22) 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.
 - (23) 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}EO_4 Model Membranes. *Biophys. J.* **1995**, *68*, 1960–1965.
 - (24) 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.
- (25) Eldho, N. V.; Feller, S. E.; Tristram-Nagle, S.; Polozov, I. V.; Gawrisch, K. Polyunsaturated Docosahexaenoic vs Docosapentaenoic Acid Differences in Lipid Matrix Properties from the Loss of One Double Bond. *Journal of the American Chemical Society* **2003**, *125*, 6409–6421, PMID: 12785780.
- (26) Wohllert, 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.
- (27) 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.
- (28) 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.
- (29) 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.
- (30) 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.
- (31) 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.
- (32) Bacle, A.; Fuchs, P. F. Berger pure POPC MD simulation (300 K - 300ns - 1 bar). 2018; {<https://doi.org/10.5281/zenodo.1402417>}.
- (33) 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.
- (34) 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.
- (35) 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.
- (36) Milan Rodriguez, P.; Fuchs, P. F. MacRog pure POPC MD simulation (300 K - 500ns - 1 bar). **2020**,
- (37) 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.
- (38) 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.

- (39) 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.
- (40) Javanainen, M. POPC with 0, 10, 20, and 30 mol-Slipids force field. 2016; <http://dx.doi.org/10.5281/zenodo.3243328>.
- (41) 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.
- (42) Melcr, J. Simulations of POPC lipid bilayer in water solution at various NaCl, KCl and CaCl₂ concentrations using ECC-POPC force field. **2019**,
- (43) 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.
- (44) 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>}.
- (45) 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.
- (46) 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 ¹H–¹³C NMR and MD simulation studies. *Phys. Chem. Chem. Phys.* **2013**, *15*, 1976–1989.
- (47) Ollila, O. H. S. MD simulation trajectory and related files for POPC/cholesterol (50 molmodified Höltje, Gromacs 4.5). **2014**,
- (48) 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.
- (49) 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.
- (50) Javanainen, M. POPC with 40 and 50 mol-force field. 2016; <https://doi.org/10.5281/zenodo.3238157>, DOI: 10.5281/zenodo.3238157.
- (51) Javanainen, M.; Kulig, W. POPC/Cholesterol @ 310K. 0, 10, 40, 50 and 60 mol-cholesterol. Model by Maciejewski and Rog. **2015**,
- (52) 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.
- (53) 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**,
- (54) 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.
- (55) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=24 w/l. 2015; DOI: 10.5281/zenodo.16289.
- (56) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=16 w/l. 2015; DOI: 10.5281/zenodo.16292.
- (57) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=12 w/l. 2015; DOI: 10.5281/zenodo.16293.

- (58) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=4 w/l. 2015; DOI: 10.5281/zenodo.16295.
- (59) 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.
- (60) 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.
- (61) 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.
- (62) 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.
- (63) 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.
- (64) 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.
- (65) 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, et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference. 2016; pp 98 – 105.
- (66) 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.
- (67) Schlenkrich, M.; Brickmann, J.; MacKerell, A. D.; Karplus, M. *Biological Membranes*; Springer, 1996; pp 31–81.
- (68) 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.
- (69) 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.
- (70) Leeb, F.; Maibaum, L. Spatially Resolving the Condensing Effect of Cholesterol in Lipid Bilayers. *Biophysical Journal* **2018**, *115*, 2179 – 2188.
- (71) Ulrich, A.; Watts, A. Molecular response of the lipid headgroup to bilayer hydration monitored by 2H-NMR. *Biophys. J.* **1994**, *66*, 1441 – 1449.
- (72) 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.
- (73) Seelig, J.; MacDonald, P. M.; Scherer, P. G. Phospholipid head groups as sensors of electric charge in membranes. *Biochemistry* **1987**, *26*, 7535–7541, PMID: 3322401.

Graphical TOC Entry

TOC here if needed