bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

Hanne S. Antila,[*,†] Tiago M. Ferreira,[‡] 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*
¶*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 monitor time-resolved motions of biomacromolecules, although it often remains unknown how closely the conformational dynamics correspond to those occurring in real life. Here, we used a large set of open-access MD trajectories of phosphatidylcholine (PC) lipid bilayers to benchmark the conformational dynamics in several contemporary MD models (force fields) against nuclear magnetic resonance (NMR) data available in the literature: effective correlation times and spin-lattice relaxation rates.

We found none of the tested MD models to fully reproduce the conformational dynamics. That said, the dynamics in CHARMM36 and Slipids are more realistic than in the Amber Lipid14, OPLS-based MacRog, and GROMOS-based Berger force fields, whose sampling of the glycerol backbone conformations is too slow. The performance of CHARMM36 persists when cholesterol is added to the bilayer, and when the hydration level is reduced. However, for conformational dynamics of the PC headgroup, both with and without cholesterol, Slipids provides the most realistic description, because CHARMM36 overestimates the relative weight of ∼1-ns processes in the headgroup dynamics.

We stress that not a single new simulation was run for the present work. This demonstrates the worth of open-access MD trajectory databanks for the indispensable step of any serious MD study: Benchmarking the available force fields. We believe this proof of principle will inspire other novel applications of MD trajectory databanks, and thus aid in developing biomolecular MD simulations into a true computational microscope—not only for lipid membranes, but for all biomacromolecular systems.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 scientific research. Constantly growing and improving in fidelity due to collaborative effort,[5–8] the now hundreds of databanks[9] fuel the data-driven development of biomolecular structure determination,[10] refinement,[11] prediction,[12] and design[13] approaches, as well as development of drugs,[14,15] materials,[16,17] and more.[18,19] 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,[20–22] and the discussion on how and by whom such databanks for dynamic structures would be set up is currently active.[23–26] While there are currently no general MD databanks in operation, individual databanks are accepting contributions on nucleic acid,[27] protein/DNA/RNA,[28] cyclodextrin,[29] G-protein-coupled receptor,[30] and lipid bilayer[31] simulations.

Since 2013, the NMRlipids Project (`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.[32] While its main focus has been on conformational ensembles of different lipid headgroups and on ion binding to lipid membranes,[32–34] the NMRlipids Project has also built a databank[31] (`zenodo.org/communities/nmrlipids`) containing hundreds of atomistic MD trajectories of lipid bilayers and indexed at `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.[32,34–37] 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.[38]

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 rate 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,[39] as well as to understand the dynamics of biological processes where molecular deformations play a rate-limiting role, such as membrane fusion.[40] 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

Lipid conformational dynamics in NMR data.

We analyzed the veracity of phosphatidylcholine (PC) lipid dynamics in MD based on

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

two quantities that are readily available from published[39,41–43] ^{13}C-NMR experiments and directly quantifiable from atomistic MD simulations: The effective C–H bond correlation times τ_e, and the spin-lattice relaxation rates R_1.

Effective C–H bond correlation times τ_e.

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

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

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

$$g(\tau) = g_f(\tau)g_s(\tau), \tag{2}$$

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 magic angle spinning conditions (Fig. 1). Ferreira et al.[41] have experimentally demonstrated that for all phospholipid carbons the motional correlation times contributing to g_f are well below μs, and to g_s well above 100 μs. This separation of time scales gives rise to the plateau $g(1\,\mu\text{s} \lesssim \tau \lesssim 100\,\mu\text{s}) = S_{CH}^2$ illustrated in Fig. 1. S_{CH} is the C-H bond order parameter

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

where $\theta(t)$ is the angle between the C–H bond and the bilayer normal. S_{CH} can be independently measured using dipolar coupling in ^{13}C or quadrupolar coupling in ^{2}H-NMR experiments. Knowing the set of S_{CH} for all the C–H bonds in a lipid is highly useful in order to evaluate its conformational ensemble.[38]

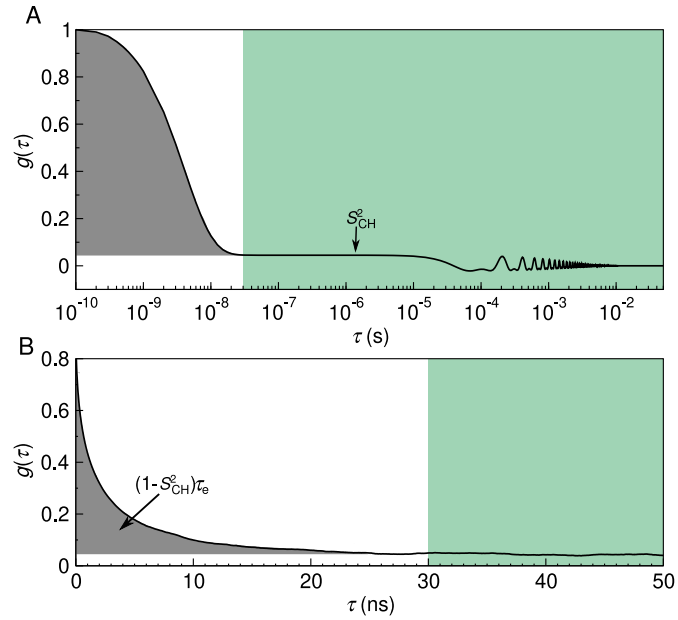


Figure 1: C–H bond autocorrelation function $g(\tau)$. (A) Idealised illustration of the fast (white background) and the slow (green) mode of the correlation function in solid-state NMR experiments. The fast mode decays to a plateau on which $g(\tau) = S_{CH}^2$, while the slow mode gives the final descent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

As S_{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

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

which is related to the gray-shaded area below the correlation function in Fig. 1. The τ_e detect essentially an average over all the time scales relevant for the lipid conformational dynamics. Their relation to process speeds is intuitive: Increase of long-lived correlations increases τ_e.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Spin-lattice relaxation rates R_1.

The C–H bond dynamics relate to R_1, the spin-lattice relaxation rate, through

$$R_1 = \frac{d_{\mathrm{CH}}^2 N_{\mathrm{H}}}{20} \left[j(\omega_{\mathrm{H}} - \omega_{\mathrm{C}}) + 3j(\omega_{\mathrm{C}}) + 6j(\omega_{\mathrm{H}} + \omega_{\mathrm{C}}) \right], \tag{5}$$

where ω_{H} is the ^{1}H and ω_{C} the ^{13}C-NMR Larmor frequency, and N_{H} the number of hydrogens covalently bonded to the carbon. The rigid dipolar coupling constant $d_{\mathrm{CH}} \approx -2\pi \times 22\,\mathrm{kHz}$ for the methylene bond. The spectral density $j(\omega)$ is given by the Fourier transformation

$$j(\omega) = 2 \int_0^\infty \cos(\omega\tau) g(\tau)\, \mathrm{d}\tau \tag{6}$$

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 ω_{H} and ω_{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 ∼0.1–10 ns. (In our experimental data[39,41–43] $\omega_{\mathrm{C}} = 125\,\mathrm{MHz}$ and $\omega_{\mathrm{H}} = 500\,\mathrm{MHz}$, which gives $(2\pi \times 125\ \mathrm{MHz})^{-1} = 1.3\,\mathrm{ns}$ and $(2\pi \times 625\ \mathrm{MHz})^{-1} = 0.25\,\mathrm{ns}$.) A change in given R_1, therefore, indicates a change in the relative amount of processes occurring in a window around the sensitive timescale; inferring also the direction to which the processes changed (speedup/slowdown) requires measuring R_1 at various field strengths.

Data acquisition and analysis.

All the experimental quantities used in this work were collected from the literature sources[39,41–43] cited at the respective figures.

The simulation trajectories were collected from the general-purpose open-access repository Zenodo (`zenodo.org`), with the majority of the data originating from the NMRlipids Project[32,33] (`nmrlipids.blogspot.fi`). The trajectories were chosen by hand based on how well the simulation conditions matched the available experimental data (lipid type, temperature, cholesterol content, hydration), and how precisely one could extract the quantities of interest from the trajectory (length of simulation, system size). Table 1 lists the chosen trajectories of pure POPC (1-palmitoyl-2-oleoyl-glycero-3-phosphocholine) bilayers at/near room temperature and at full hydration; Table 2 lists the trajectories with cholesterol; and Table 3 those with varying hydration. Full computational details for each simulation are available at the cited Zenodo entry.

Table 1: Analyzed open-access MD trajectories of pure POPC lipid bilayers at full hydration. Note that the temperature varied across these openly available simulation data, but in no case was T lower than in the experiment. Thus, as dynamics slows down when temperature drops, any overestimation of τ_{e} by MD (as typically seen in Fig. 2) would get worse if the simulations were done at the experimental 298 K.

force field	N_{l}[a]	N_{w}[b]	T[c](K)	t_{anal}[d](ns)	files[e]
Berger-POPC-07[44]	256	10240	300	300	[45]
CHARMM36[46]	256	8704	300	300	[47]
MacRog[48]	128	5120	300	500	[49]
Lipid14[50]	72	2234	303	50	[51]
Slipids[52]	200	9000	310	500	[53]
ECC[54]	128	6400	300	300	[55]

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

The trajectories were analyzed using in-house scripts. These are available on GitHub (`github.com/hsantila/Corrtimes/tree/master/teff_analysis`), along with a Python notebook outlining an example analysis run. To enable automated analysis of several force fields with differing atom naming conventions, we used the mapping scheme developed within the NMRlipids Project to automatically recognise the atoms and bonds of interest for each trajectory.

After downloading the necessary files from Zenodo, we processed the trajectory with Gro-

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2: Analyzed open-access MD trajectories of cholesterol-containing POPC bilayers at full hydration.

force field POPC/cholesterol	c_{chol}[a]	N_{chol}[b]	N_{l}[c]	N_{w}[d]	T[e](K)	t_{anal}[f](ns)	files[g]
Berger-POPC-07[44]	0%	0	128	7290	298	50	[56]
/Höltje-CHOL-13[57,58]	50%	64	64	10314	298	50	[59]
CHARMM36[46]	0%	0	200	9000	310	500	[60]
/CHARMM36[61]	50%	200	200	18000	310	500	[62]
MacRog[48]	0%	0	128	6400	310	500	[63]
/MacRog[48]	50%	64	64	6400	310	500	[63]
Slipids[52]	0%	0	200	9000	310	500	[53]
/Slipids[64]	50%	200	200	18000	310	500	[53]

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

Table 3: Analyzed open-access MD trajectories of PC lipid bilayers under varying hydration level.

force field	lipid	$n_{\mathrm{w/l}}$[a]	N_{l}[b]	N_{w}[c]	T[d](K)	t_{anal}[e](ns)	files[f]
Berger-POPC-07[44]	POPC	40	256	10240	300	300	[45]
	POPC	7	128	896	298	60	[65]
Berger-DLPC-13[66]	DLPC[g]	24	72	1728	300	80	[67]
	DLPC[g]	16	72	1152	300	80	[68]
	DLPC[g]	12	72	864	300	80	[69]
	DLPC[g]	4	72	288	300	80	[70]
CHARMM36[46]	POPC	40	128	5120	303	140	[71]
	POPC	34	128	5120	300	500	[49]
	POPC	31	72	2232	303	20	[72]
	POPC	15	72	1080	303	20	[73]
	POPC	7	72	504	303	20	[74]
MacRog[48]	POPC	50	288	14400	310	40	[75]
	POPC	25	288	7200	310	50	[75]
	POPC	15	288	4320	310	50	[75]
	POPC	10	288	2880	310	50	[75]
	POPC	5	288	1440	310	50	[75]

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

macs `gmx trjconv` to make the molecules whole. For the united atom Berger model, hydrogens were added using the Gromacs 4.0.2 tool `g_protonate`. We then calculated the S_{CH} (Eq. (3)) with the `OrderParameter.py` script that uses the MDanalysis[76,77] Python library. The C–H bond correlation functions $g(\tau)$ (Eq. (1)) were calculated with Gromacs 5.1.4[78] `gmx rotacf` (note that on MD timescales $g_{\mathrm{s}} = 1$ so that $g = g_{\mathrm{f}}$) after which the S_{CH} were used to normalize the g_{f} to obtain the reduced and normalized correlation function

$$g'_{\mathrm{f}}(\tau) = \frac{g_{\mathrm{f}}(\tau) - S_{\mathrm{CH}}^2}{1 - S_{\mathrm{CH}}^2}, \tag{7}$$

that is, the integrand in Eq. (4).

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

To quantify the error on τ_{e}, we first estimate the error on $g'_{\mathrm{f}}(\tau)$, where we account for two sources of uncertainty: $g_{\mathrm{f}}(\tau)$ and S_{CH}^2. Performing linear error propagation on Eq. (7) gives

$$\begin{aligned}\Delta g'_{\mathrm{f}}(\tau) = &\left|\frac{1}{1 - S_{\mathrm{CH}}^2}\right| \Delta g_{\mathrm{f}}(\tau) \\ &+ \\ &\left|\frac{2\,(g_{\mathrm{f}}(\tau) - 1)\,S_{\mathrm{CH}}}{(1 - S_{\mathrm{CH}}^2)^2}\right| \Delta S_{\mathrm{CH}}.\end{aligned} \tag{8}$$

Here the ΔS_{CH} was determined as the standard error of the mean of the S_{CH} over the N_{l} individual lipids in the system.[32] Similarly, we quantified the error on $g_{\mathrm{f}}(\tau)$ by first determining the correlation function $g_{\mathrm{f}}^m(\tau)$ for each individual lipid m over the whole trajectory, and then obtaining the error estimate $\Delta g_{\mathrm{f}}(\tau)$ as the standard error of the mean over the N_{l} lipids. Importantly, this gives an uncertainty estimate for $g_{\mathrm{f}}(\tau)$ at each time point τ.

To obtain the lower bound on τ_{e}, we integrate the function $g'_{\mathrm{f}}(\tau) - \Delta g'_{\mathrm{f}}(\tau)$ over time from $\tau = 0$

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

until $\tau = t_\mathrm{l}$. Here

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

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

To obtain the upper error estimate on τ_e, we first integrate the function $g'_\mathrm{f}(\tau) + \Delta g'_\mathrm{f}(\tau)$ over time from $\tau = 0$ until $t_\mathrm{u} = \min\{t_0, t_\mathrm{anal}/2\}$. 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, that is, we assume that all the decay of g_f from the time point t_u onwards comes solely from this hypothetical slowest process that decays with a time constant of 10^{-6} s. The additional contribution to the upper bound for τ_e then reads

$$(g'_\mathrm{f}(t_\mathrm{u}) + \Delta g'_\mathrm{f}(t_\mathrm{u})) \times \underbrace{\int_{t_\mathrm{u}}^{10^{-6}\,\mathrm{s}} \exp\left(-\frac{\tau - t_\mathrm{u}}{10^{-6}\,\mathrm{s}}\right) \mathrm{d}\tau}_{=\left(1-\exp\left(\frac{t_\mathrm{u}-10^{-6}\,\mathrm{s}}{10^{-6}\,\mathrm{s}}\right)\right)10^{-6}\,\mathrm{s}}. \quad (10)$$

The R_1 rates were calculated using Eq. (5). The spectral density $j(\omega)$ was obtained from the normalized correlation function g'_f by fitting it with a sum of 61 exponentials

$$g'_\mathrm{f}(\tau) \approx \sum_{i=1}^{61} \alpha_i e^{-\tau/\tau_i}, \quad (11)$$

with logarithmically spaced time-scales τ_i ranging from 1 ps to 1 μs, and then calculating the spectral density of this fit based on the Fourier transformation[41]

$$j(\omega) = 2(1 - S_\mathrm{CH}) \sum_{i=1}^{61} \alpha_i \frac{\tau_i}{1 + \omega\tau_i}. \quad (12)$$

The R_1 rate of a given C–H pair was first calculated separately for each lipid m (using Eq. (5) with $N_\mathrm{H} = 1$, and $j^m(\omega)$ obtained for the normalized correlation function g'^m_f). The resulting N_l measurements per C–H 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 C2 carbon in the acyl chains and the γ carbons), the R_1 from simulations was obtained as an average over carbons with overlapping peaks. The segment-wise error estimates were obtained by standard error propagation, starting from the uncertainties of the R_1 rates of the C–H 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 those terms 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' $R_1(\tau \to \infty)$ 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

$$\tau_\mathrm{e} = \int_0^\infty g'_\mathrm{f}(\tau)\,\mathrm{d}\tau \approx \sum_i \alpha_i \tau_i. \quad (13)$$

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 function, and thus we chose to quantify τ_e using the area under g'_f, and estimate its uncertainty as detailed above.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3 Results and Discussion

Using open-access MD simulation trajectories, we benchmark phospholipid conformational dynamics in six MD force fields. We start with pure POPC bilayers in their liquid crystalline fully hydrated state (see Table 1 for simulation details and Fig. 2 for the data), and then proceed to check the changes in dynamics when cholesterol is added to the bilayer (Table 2 and Fig. 4) and when the hydration level is reduced (Table 3 and Fig. 5). Our yardsticks are the effective correlations times τ_e (Eq. (4)) and the R_1 rates (Eq. (5)) measured at 125 MHz ^{13}C (500 MHz ^{1}H) Larmor frequency; an MD model with correct rotational dynamics in a window around ∼1 ns will match the experimental R_1 rates, whereas the τ_e reflect all the sub-μs time scales (Fig. 1).

Pure POPC at full hydration: Slipids and CHARMM36 reproduce τ_e excellently.

The top panels of Fig. 2 compare the effective correlation times τ_e obtained for fully hydrated POPC bilayers in experiments (black) and in six different MD force fields (color). We see that—as implied by the discussion leading to Eq. (10)—sub-μs MD simulations typically lead to asymmetric error bars on τ_e; if these open-access trajectories were extended, the τ_e values would more likely increase than decrease. 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, most MD simulations tend to produce too slow dynamics in the glycerol region (Fig. 2). This is consistent with previous results for the Berger model,[41] and with the insufficient conformational sampling of glycerol backbone torsions observed in 500-ns-long CHARMMc32b2[79,80] simulations of a PC lipid.[81]

The best overall τ_e-performance is seen in Slipids and in particular CHARMM36 (Fig. 2). This is in line with CHARMM36 reproducing the most realistic conformational ensembles for the headgroup and glycerol backbone among the MD simulation force fields benchmarked here.[32,34] Indeed, it is important to keep in mind that the conformational ensembles greatly differ between force fields and are not exactly correct in any of them.[32,34] Consequently, the calculated τ_e times and R_1 rates depict the dynamics of sampling a somewhat different and incorrect phase space for each model. To this end, we try to avoid overly detailed discussion on the models and rather concentrate on common and qualitative trends. That said, there are a few carbon segments in the data for which the experimental order parameters, R_1, and τ_e are all (almost) reproduced by simulations, suggesting that both the conformational ensemble and the dynamics are correctly captured by MD in these cases. For example, Slipids performs well at the β and α, and CHARMM36 at the g_3, g_2, and C2 segments. These are, however, exceptions.

An excellent τ_e may be accompanied by a poor R_1, or *vice versa.*

The lower panels of Fig. 2 compare the experimental and simulated R_1 rates under the same conditions that were used for the τ_e above. Notably, there are several instances where the R_1 comparison distinctly differs from what was seen for τ_e.

There are cases where a matching R_1 is accompanied by a larger-than-experimental τ_e. MacRog for the β, α, and g_1 segments provides a prominent example of this. Such a combination suggests that MD has the correct relative weight of 1-ns-scale dynamics, but has too slow long-time dynamics.

There are also cases where τ_e matches experiments, but R_1 does not, such as the β and α segments in CHARMM36. Therein a cancellation of errors occurs in τ_e: The overestimation of the relative weight of 1-ns-scale dynamics is compensated by wrong dynamics at the other time scales. As CHARMM36 overall performs rather well for C–H bond order parameters, R_1, and τ_e, we proceed to study this shortcoming on the headgroup R_1 rates in some more detail.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

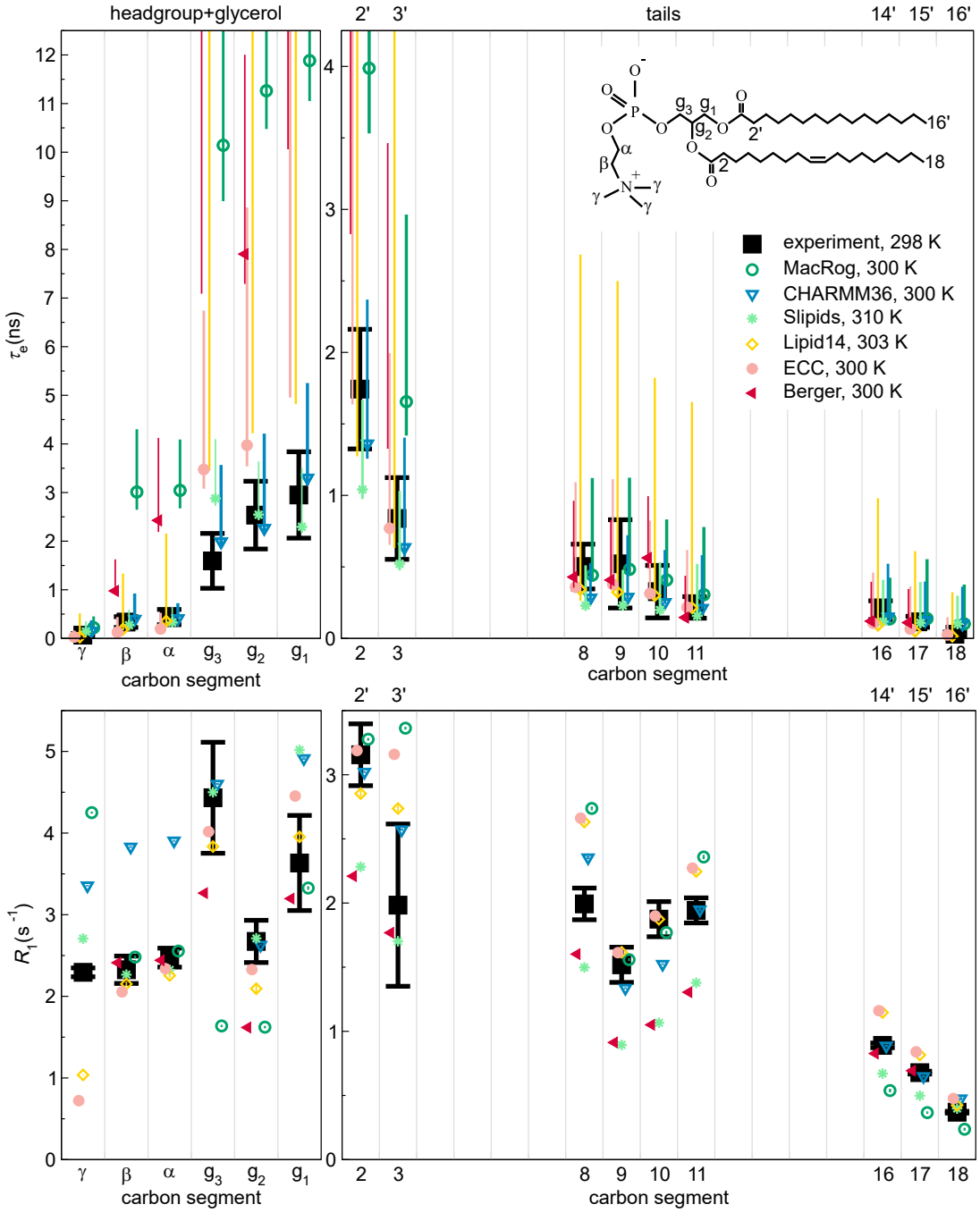


Figure 2: Effective correlation times (τ_e, top) and R_1 rates (bottom) in experiments[39] (black) and MD simulations (colored) of POPC bilayers in L_α phase under full hydration. Inset shows the POPC chemical structure and carbon segment labeling. Each plotted value contains contributions from all the 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 for which experimental data was available. 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 segments γ, C18, and C16' are left out, as 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, while information on the experiments is available at Ref. 39.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conformational dynamics of PC headgroup segments in MD.

Figure 3A zooms in on the headgroup (γ, β, α) segments, whose τ_e were not clearly visible on the scale of Fig. 2. We see that for γ, no force field provides both τ_e and R_1, but Slipids comes closest. For β and α, Slipids captures both measurables near perfectly. In other words, among the benchmarked force fields Slipids gives the most realistic description of the conformational dynamics in the headgroup region. CHARMM36, e.g., overestimates (R_1) the relative weight of timescales around ∼1 ns.

To investigate closer how the differences between force fields arise, Fig. 3B shows the 'cumulative' $R_1(\tau)$, where the ranges of steepest increase indicate time scales that most strongly contribute to R_1 rates.

For the γ segment, Fig. 3B shows that for models that overestimate the R_1 rate (MacRog, CHARMM36, and Slipids, see Fig. 3A) the major contribution to R_1 arises at $\tau > 50$ ps, whereas for models that underestimate R_1 (Lipid14 and ECC) the major contribution comes from $\tau < 50$ ps. This also manifests in the distribution of fitting weights (α_i in Eq. (11)) in Fig. 3C: The later non-zero weights occur, the larger is the resulting R_1 of γ.

For the β and α segments, Fig. 3B shows that the main contribution to R_1 rates arises from processes between 100 ps and 1 ns. CHARMM36 has the largest relative weights of all models in this window (Fig. 3C), which explains its overestimation of R_1 of β and α. All the other models have R_1 rates close to experiments, but only Slipids simultaneously gives also the τ_e correctly. Notably, Slipids has its largest weights at $\tau < 100$ ps. Indeed, the considerable weights at short (< 10 ps) time scales in all models except MacRog and at long (> 10 ns) time scales in MacRog and Berger hardly manifest in R_1. However, the latter contribute heavily to τ_e, which is thus considerably overestimated by MacRog and Berger (Fig. 2).

It would be highly interesting to identify the origins of the observed artificial timescales, particularly for the 0.1–1 ns window over-presented in CHARMM36, and propose how to correct those in the simulation models. However, the connection between the fitted correlation times

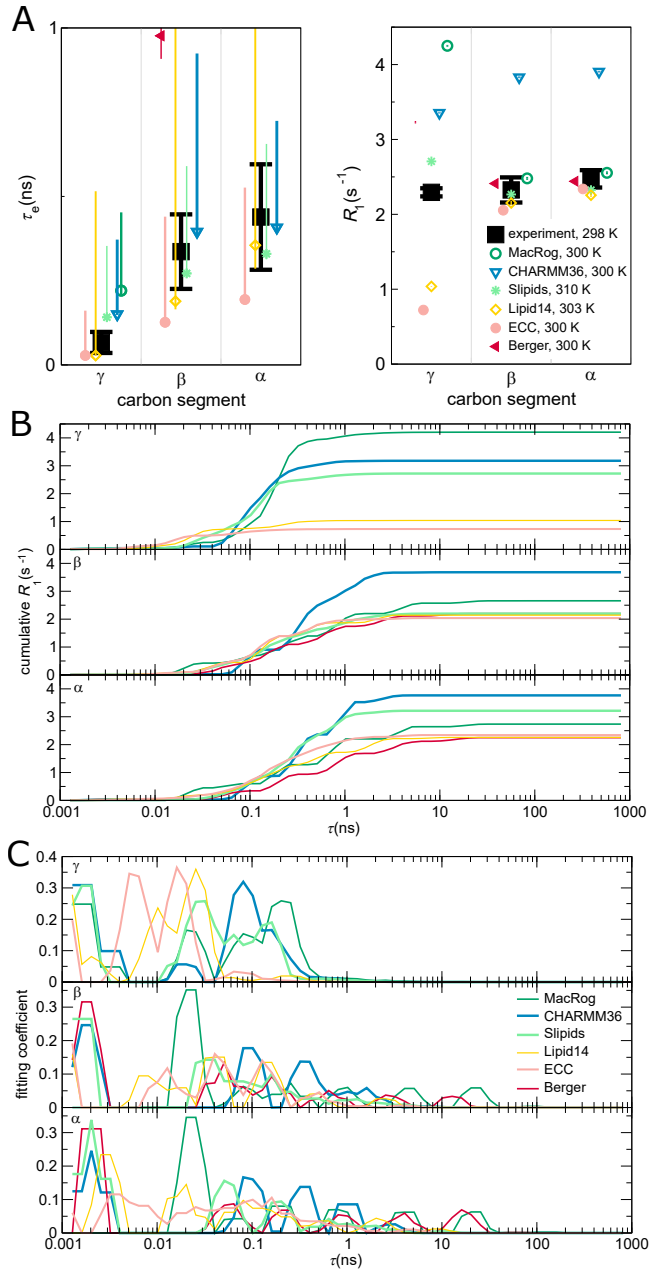


Figure 3: Contributions to the dynamics of the headgroup segments. (A) Zoom on the headgroup τ_e (left panel) and R_1 (right). (B) 'Cumulative' $R_1(\tau)$ of the γ (top panel), β (middle), and α (bottom) segments. $R_1(\tau)$ is obtained, as detailed in Methods, by including in the sum of Eq. (12) only terms with $\tau_i < \tau$. Consequently, at $\tau \to \infty$ the $R_1(\tau)$ approaches the actual R_1. (C) Prefactor weighs α_i from Eq. (11) of γ (top), β (middle), and α (bottom). Note that panels B and C show a sliding average over 3 neighboring data points.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

An essential component in cell membranes, cholesterol has various biological functions. It is well known to order the acyl chains in lipid bilayers, but its effect on the headgroup is more controversial.[58,82] For example, it has been proposed that lipid headgroups reorganize to shield cholesterol from water.[82] However, while acyl chains do substantially order, NMR experiments show no significant conformational changes in the headgroup upon addition of even 50% of cholesterol—which suggests that the tail and head regions behave essentially independently.[32,58] In principle, the headgroups could shield cholesterol from water even without changing their conformational ensemble: By reorienting only laterally on top of the cholesterol. In this case, one would expect the rotational dynamics of headgroup segments to change when cholesterol is added.

Top panels of Fig. 4A depict the experimental effective correlation times τ_e in pure POPC bilayers and in bilayers containing 50% cholesterol. The τ_e at the glycerol backbone slow down markedly when cholesterol is added. Tail segment dynamics slows down too, most notably close to the glycerol backbone. In stark contrast, τ_e of the headgroup segments (γ, β, α) remain unaffected. Furthermore, cholesterol induces no measurable change in the headgroup β and α segment dynamics at short ($\sim$1 ns) time scales, as demonstrated by the experimental R_1 rates (Fig. 4A, bottom panels). That said, there is a small but measurable impact on R_1 at γ. In summary, these experimental findings support the idea[39] that the acyl chains and the headgroup can respond almost independently to changes in conditions and composition.

All four benchmarked force fields (Fig. 4B) qualitatively reproduce the experimental increase in τ_e: Slipids and CHARMM36 give rather decent magnitude estimates, while MacRog grossly overestimates the slowdown of glycerol, C2, and C3 segments. Notably, MacRog appears to predict slowdown also for the headgroup (β and α), for which experiments detect no change. Note that while CHARMM36 correctly shows no change in τ_e of the γ, β, and α segments, it does predict an erroneous ΔR_1 for all three, indicating some inaccuracies in the headgroup rotational dynamics. Such inaccuracies might be reflected in the recent findings[83] (obtained using CHARMM36) that the headgroups of PCs neighboring (within 6.6 Å) a lone cholesterol spend more time on top of the said cholesterol than elsewhere. Interestingly, the tail ΔR_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. All these findings are in line with the general picture obtained from C–H bond order parameters:[38] MD simulations capture the changes in acyl chain region rather well, but changes in and near the glycerol backbone region can be overestimated. Of the benchmarked force fields, CHARMM36 appears most realistic in reproducing the effects of cholesterol on the glycerol backbone—and Slipids on the PC headgroup—conformational dynamics.

Effect of drying.

Understanding the impact of dehydration on the structure and dynamics of lipid bilayers is of considerable biological interest. Dehydrated states are found, e.g., in skin tissue. Most prominently, the process of membrane fusion is always preceeded by removal of water between the approaching surfaces, and thus the dehydration-imposed changes can considerably affect fusion characteristics, such as its rate.

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 rather insensitive

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

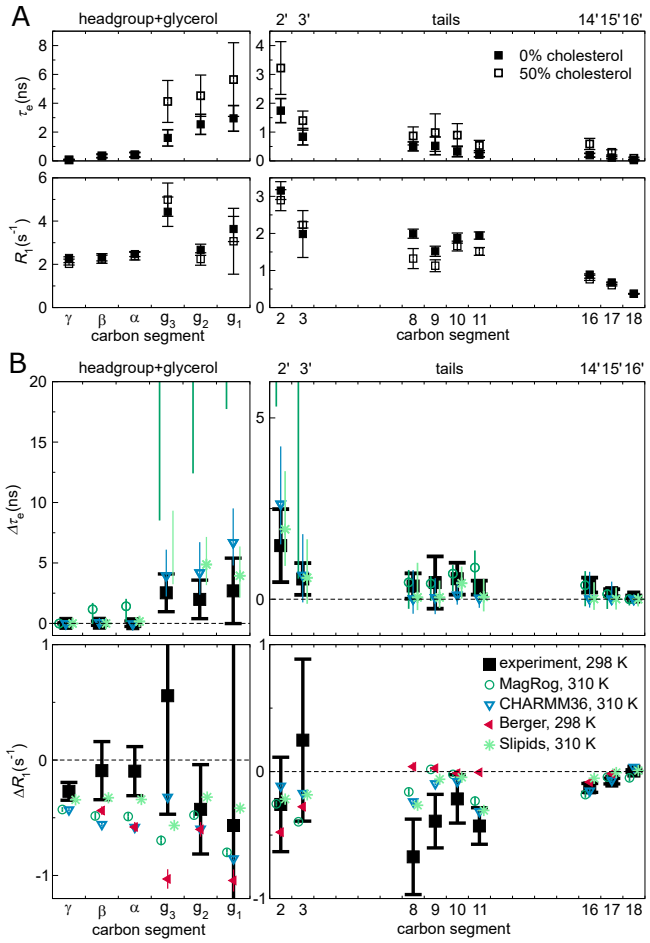


Figure 4: Effect of cholesterol on POPC conformational dynamics. (A) Experimental effective correlation times τ_e (top panels) and R_1 rates (bottom) in 100/0 and 50/50 POPC/cholesterol bilayers at full hydration, see Ref. 39 for further details. (B) The change in τ_e ($\Delta\tau_e$, top panels) and R_1 (ΔR_1, bottom), in NMR (black) and MD (color), when bilayer composition changes from pure POPC to 50% cholesterol. 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. The Berger $\Delta\tau_e$ is not shown, because the available open-access trajectories were too short to determine meaningful error estimates. Table 2 provides further simulation details; for segment labeling, see Fig. 2.

to the chemical identities of the tails. This is analogous to what was seen experimentally when adding cholesterol (Fig. 4A): Structural changes in the tail and glycerol regions do not (need to) affect the headgroup dynamics. Secondly, a mild dehydration does not alter the τ_e in the headgroup and glycerol regions.

Figure 5B shows the effects of dehydration in three MD models. Combination of the unrealistically slow dynamics, especially in the glycerol backbone (Fig. 2), and the relatively short lengths of the available open-access trajectories (Table 3) led to large uncertainty estimates; thus we only point out qualitative trends here. For all headgroup and glycerol segments, the simulated τ_e indicate slowdown 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 stay almost constant all the way until 7 w/l, whereas for Berger and MacRog they hint that a retardation of dynamics starts already between 15 and 10 w/l.

These simulational findings suggest that experiments reducing hydration levels below 10 w/l would also show an increase in τ_e. This prediction is in line with the exponential slowdown of the headgroup conformational dynamics upon dehydration that was indicated by ^{2}H-NMR R_1 measurements of DOPC bilayers: $R_1 \sim \exp(-n_{w/l}/4)$.[84] The slowdown was attributed to the reduced effective volume available for the headgroup[84] as it tilts towards the membrane upon dehydration; such tilt is observed via changes of the lipid headgroup order parameters,[85] and is qualitatively reproduced by all the simulation models.[32]

Figure 5C shows a collection of experimental ^{13}C-NMR R_1 rates for the headgroup segments at different water contents; in addition to the full hydration POPC data from Fig. 2, DMPC at 13 w/l,[42] and POPC at 20 and 5 w/l[43] are shown. Experimentally, an increasing trend with decreasing hydration is observed for all three 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 slowdown characteristics discussed here are of significance not only for computational studies of intermembrane interactions, such as fusion, but also when simulating a bilayer

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

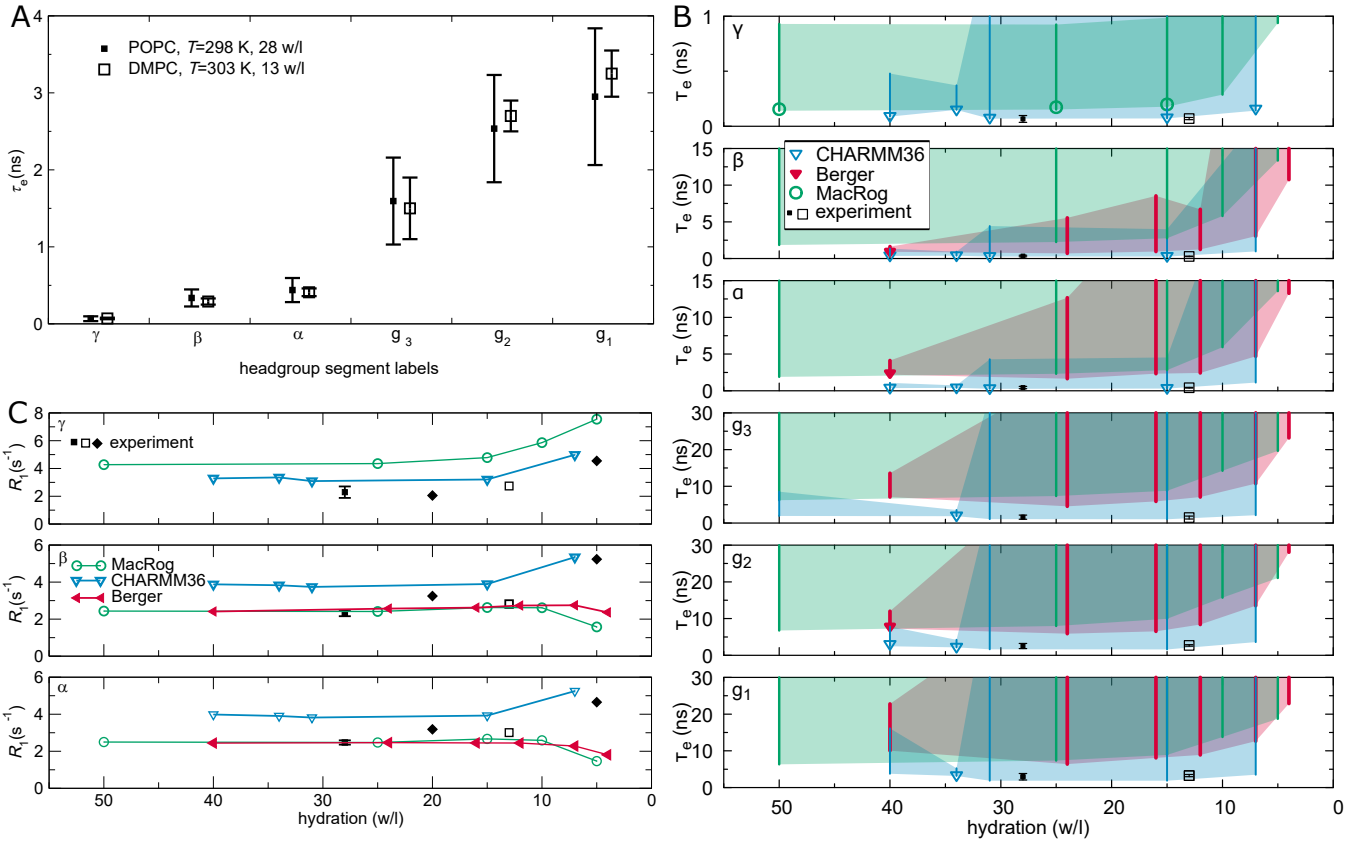


Figure 5: Effect of drying on PC headgroup and glycerol backbone conformational dynamics. (A) Experimental effective correlation times τ_e for DMPC at low hydration (from Ref. 42) do not significantly differ from the τ_e for POPC at full hydration (from Ref. 39). (B) Calculated τ_e for POPC at decreasing hydration in three MD models. Symbols indicate the mean of segment hydrogens if τ_e could be determined for all of them; otherwise, only the error bar (extending from the mean of the lower to the mean of the upper uncertainty estimates) is drawn. The area limited by the error bars shaded for visualization. Note that four Berger data points (4, 12, 16, and 24 w/l) are from DLPC. (C) ^{13}C-NMR R_1 rates (at $\omega_C = 125$ MHz) of the PC headgroup segments in experiments and simulations: Experiments indicate an increasing trend upon dehydration. Experimental POPC data at 28 w/l is from Ref. 39 (■), POPC at 20 and 5 w/l from Ref. 43 (♦), and DMPC at 13 w/l from Ref. 42 (□). See Table 3 for simulation details.

(stack) under low hydration: Slower dynamics require longer simulation times for equilibration, for reliably quantifying the properties of the bilayers, and for observing rare events.

4 Conclusions

We have here demonstrated that open access databanks of MD trajectories enable the creation new scientific information without running a single new simulation. More specifically, we have benchmarked (against published NMR data[39,41–43]) the conformational dynamics of a wide range of phosphatidylcholine MD models using existing open-access trajectories from the Zenodo repository, in particular those belonging to the NMRlipids Databank (`zenodo.org/communities/nmrlipids`).

We found that every MD model captures the ^{13}C-NMR effective correlation time (τ_e) profile of POPC qualitatively, but that most are prone to too slow dynamics of the glycerol backbone C–H bonds (Fig. 2). While no force field perfectly reproduces all the experimental data, CHARMM36 and Slipids have overall impressive τ_e. This is a particularly exciting finding concerning CHARMM36, as it is also known to reproduce quite well the experimental con-

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

formational ensemble.[32] That said, we do find that CHARMM36 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). In fact Slipids, which also reproduces the experimental headgroup order parameters,[32] appears to outperform CHARMM36 when it comes to headgroup conformational dynamics (Fig. 3).

Further, we found that when cholesterol is mixed into a POPC bilayer, MD qualitatively captures the slowdown of conformational dynamics in the tail and glycerol regions (Fig. 4). However, the benchmarked force fields overestimate the changes in the ∼1-ns dynamics of the headgroup—except Slipids, which captures well the effects of cholesterol on PC headgroup conformational dynamics.

Finally, we found that upon reducing the water content below 10 waters per lipid, MD exhibits slowdown of headgroup and backbone dynamics in qualitative agreement with experimental data. That said, only CHARMM36 (but not Berger or MacRog) qualitatively captures the experimentally detected increase of R_1 rates upon dehydration (Fig. 5).

While work is still needed in capturing even the correct phospholipid conformations,[32] realistic dynamics will be an essential part of developing MD into a true computational microscope. Here we gathered a set of published experimental ^{13}C-NMR data on phosphatidylcholine dynamics, and charted the typical features of the existing MD models against it, thus laying the foundation for further improvement of the force fields. Importantly, our work demonstrates the potential of open-access MD trajectories in achieving such benchmarks at a reduced computational and labor cost. A pool of well indexed and documented open-access data provides an ideal platform for automation, which in turn will facilitate faster progress in pinpointing the typical failures of the existing force fields, in identifying key differences in models describing chemical variations of the same molecule type (such as different lipid headgroups), and in developing better models through data-driven approaches.

Acknowledgement

H. S. A. gratefully acknowledges the support from Osk. Huttunen Foundation, Finnish Academy of Science and Letters (Foundations' Post Doc Pool), Instrumentarium Science Foundation, and the Alexander von Humboldt Foundation. O. H. S. O. acknowledges the Academy of Finland (315596, 319902) for financial support.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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*, 320–327.

(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.; Verdiguel, 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.; Ostropytskyy, 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.; Tiemann, 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(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 Databank 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.

(33) Catte, A.; Girych, M.; Javanainen, M.; Loison, C.; Melcr, J.; Miettinen, M. S.; Monticelli, L.; Määttä, J.; Oganesyan, 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.; Iwaï, H.; Ollila, O. H. S. Heterogeneous dynamics in partially disordered proteins. *Phys. Chem. Chem. Phys.* **2020**, *22*, 21185–21196.

(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:

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1H–13C 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.

(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/$C_{12}E0_4$ Model Membranes. *Biophys. J.* **1995**, *68*, 1960–1965.

(44) 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.

(45) Bacle, A.; Fuchs, P. F. Berger pure POPC MD simulation (300 K - 300ns - 1 bar). 2018; `https://doi.org/10.5281/zenodo.1402417`.

(46) 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.

(47) Papadopoulos, C.; Fuchs, P. F. CHARMM36 pure POPC MD simulation (300 K - 300ns - 1 bar). 2018; `http://dx.doi.org/10.5281/zenodo.1306800`.

(48) 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.

(49) Milan Rodriguesz, P.; Fuchs, P. F. MacRog pure POPC MD simulation (300 K - 500ns - 1 bar). 2020; `http://dx.doi.org/10.5281/zenodo.3741793`.

(50) 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.

(51) Ollila, O. H. S.; Retegan, M. MD simulation trajectory and related files for POPC bilayer (Lipid14, Gromacs 4.5). 2014; `https://dx.doi.org/10.5281/zenodo.12767`.

(52) 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.

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

(54) 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.

(55) Melcr, J. Simulations of POPC lipid bilayer in water solution at various NaCl, KCl and CaCl2 concentrations using ECC-POPC force field. 2019; `http://dx.doi.org/10.5281/zenodo.3335503`.

(56) 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`.

(57) 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cholesterol. *Biochim. Biophys. Acta* **2001**, *1511*, 156 – 167.

(58) 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.

(59) Ollila, O. H. S. MD simulation trajectory and related files for POPC/cholesterol (50 molmodified Höltje, Gromacs 4.5). 2014; `http://dx.doi.org/10.5281/zenodo.13285`.

(60) Javanainen, M. POPC with 0, 10, 20, and 30 mol-Charmm36 force field. 2016; `https://doi.org/10.5281/zenodo.3237420`.

(61) 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.

(62) Javanainen, M. POPC with 40 and 50 mol-force field. 2016; `https://doi.org/10.5281/zenodo.3238157`.

(63) Javanainen, M.; Kulig, W. POPC/Cholesterol @ 310K. 0, 10, 40, 50 and 60 mol-cholesterol. Model by Maciejewski and Rog. 2015; {`http://dx.doi.org/10.5281/zenodo.3252052`}.

(64) 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.

(65) 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; `http://dx.doi.org/10.5281/zenodo.13814`.

(66) 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.

(67) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=24 w/l. 2015; `http://dx.doi.org/10.5281/zenodo.16289`.

(68) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=16 w/l. 2015; `http://dx.doi.org/10.5281/zenodo.16292`.

(69) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=12 w/l. 2015; `http://dx.doi.org/10.5281/zenodo.16293`.

(70) Kanduc, M. MD trajectory for DLPC bilayer (Berger, Gromacs 4.5.4), nw=4 w/l. 2015; `http://dx.doi.org/10.5281/zenodo.16295`.

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

(72) 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`.

(73) 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`.

(74) 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`.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(76) 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.

(77) 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.

(78) 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.

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

(80) 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.

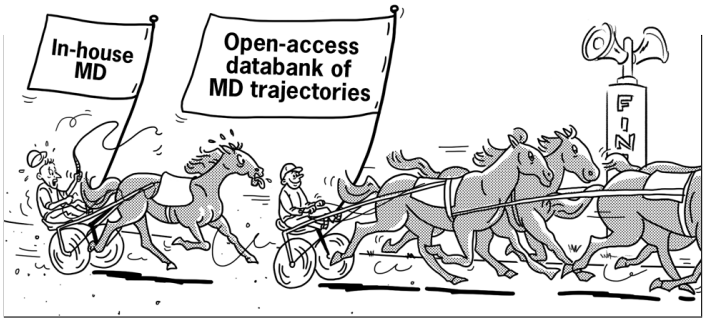
(81) Vogel, A.; Feller, S. E. Headgroup Conformations of Phospholipids from Molecular Dynamics Simulation: Sampling Challenges and Comparison to Experiment. *The Journal of Membrane Biology* **2012**, *245*, 23–28.

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

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

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

(85) 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.09.374850; this version posted November 9, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Graphical TOC Entry