Boosting conformational sampling in lipid bilayer simulations using hamiltonian

Boosting conformational sampling in lipid bilayer simulations using Hamiltonian Replica Exchange

Chiara Cardelli[1], *Alessandro Barducci*[2], *Piero Procacci*[3*]

[1] *Computational Physics Department*
University of Wien, Sensengasse 8/9 1090 Wien, Austria

[2] *Laboratoire de Biophysique Statistique*
École Polytechnique Fédérale, CH-1015 Lausanne, Switzerland

[3] *Chemistry Department*
University of Florence, Via Lastruccia 3 50019 Sesto Fiorentino, Italy

Abstract. In this report we have tested a parallel implementation for the simulation of lipid bilayers at atomistic level, based on a generalized ensemble (GE) protocol where only the torsional degrees of freedom of the alkyl chains of the lipid are scaled. To this aim, we have used our in-house code ORAC where parallelism is implemented exclusively via the Hamiltonian Replica Exchange algorithm (H-REM) with the atomic forces being integrated serially for each simultaneously evolving GE trajectories. The results in terms of configurational sampling enhancement have been compared with a conventional simulation produced with a widespread molecular dynamics code with parallelism based on a domain decomposition approach for parallel computation of the forces. Results show that the proposed thermodynamic-based multiple trajectories parallel protocol for membrane simulations is competitive with the conventional single trajectory domain decomposition approach as far as area and volume fluctuations are concerned while the gain is only moderate for transport/mixing properties, decisively pointing to a mixed strategy as the optimal parallelization approach in lipid bilayer.

1 Introduction

Molecular dynamics simulation is an important computational tool for the study of biomolecular systems, such as biological membranes, that have lipid bilayers as main constituents. The recent development of massively parallel environments exploiting high speed communication links such as InfiniBand has nowadays made possible simulations in time range of hundreds of nanosecond of lipid bilayers of the extension in the tens nanometers. Although the indisputable and tremendous gain in the performance with respect to early applications, computational scientists still face a severe length-scale and a time-scale problem in membrane simulation. The actual limits in length and time scale (20-30 nm and $\simeq 1$ μs

*Corresponding author. E-mail: procacci@unifi.it.

respectively) severely restrain the possibility of studying key properties of bilayers like the bending rigidity via determination of the undulation spectrum and/or cooperative transport phenomena. Both these properties are intimately connected with important biological situations including endocytosis, lipid raft formation and stability, membrane fusions an membrane trafficking.[1, 2] As stated, flat lipid bilayers under periodic boundary conditions provide a simple and effective model system for biological membranes. Nonetheless, the simulations of a hydrated bilayer implies a number of atoms to the least in the order of tens of thousands, resulting in a high wall-time even resorting to efficient parallel algorithms such the as the dynamic domain decomposition approach.[3] Typically, on the CRESCO(1-2) platform a moderately sized system, such as a hydrated lipid bilayer of 36 molecules of palmitoyl oleoyl phosphatidylcholine (POPC) per leaflet (about 17000 atoms), can run with a maximum speed of 20-25 nanosecond per day using the popular GROMACS MD program exploiting at most 160 processors with an efficiency of less than 50%. This is so since after a certain processor number threshold, the inter-domain communication overhead dominates over the time spent in the parallelized computation of the forces within each domain. By trading model accuracy and reliability in exchange for computational speed, recent approaches for membrane simulation are based on so-called *coarse-grained* models, where larger molecular units are considered as single particles.[4]

In this report we investigate on the effectiveness of using advanced Hamiltonian Replica exchange schemes with selective scaling of specific degrees of freedom of the system[5, 6, 7] as a mean for boosting configurational sampling in simulations of model membranes at the atomistic level. In this respect, Mori *et al.*[8] have recently proposed a new GE algorithm for membrane systems, based on exchanges between few replicas spanning the surface tension space, from zero of the target replica to higher tensions, obtaining a moderate gain in the convergence time of structural parameters. On the other hand, recent advances in membrane science have highlighted the key importance of lipid flexibility and entanglement in shaping the transport and undulation phenomena in biological membranes.[2]. These molecular properties, in turn, have time scale dynamics that are essentially dictated by the free energy barriers separating, e.g., *gauche* and *trans* states for the dihedral conformation of the torsion around the sp^3 bonds of the alkyl chains in the hydrophobic interior of the bilayer. By selectively scaling, along the replica progression, the energy terms implied in these barriers (i.e. aliphatic torsion angles and 1-4 coulomb and Lennard-Jones interactions), the jump rate for *gauche trans* interconversion can be exponentially increased in the hot replicas, thereby enhancing diffusion and area/volume modulation throughout the GE. This approach should hopefully allow to collect, in few ns or tens of ns time span, a manifold of configurations statistically out of the reach of conventional (single trajectory) simulations.

The present report is organized as follows. In the Section Methods we succinctly describe the system and simulation techniques used in our contributions as well as the basics of the H-REM approach as opposed to the conventional single trajectory technique. In the section Results, we compare configurational properties such as volume and area fluctuations, diffusion and bilayer structure. In Conclusion Section a discussion and conclusive remarks are presented.

2 Method

POPC simulation setup: The simulations of the lipid bilayer, whether conventional of in the GE, comprised 72 POPC lipid units (36 for leaflet) with approximately 30 water molecules per lipid. The total number of atoms in the system was 16374. The starting PDB configuration was an equilibrated charmm36 configuration, downloaded from ref. [9]. The force field employed is a minor modification[10] of the most recent update of CHARMM parameters for lipids by Jämbeck et al., called *Slipids*, recently developed for fully saturated phospholipids.[11]

All the simulations - both conventional and GE - were performed in the isothermal-isobaric ensemble, NPT, at an external pressure of 0.1 MPa. The pressure was held constant by a Parrinello-Rahman barostat [12], with a 70 cm^{-1} oscillator frequency and compressibility set to $5.3 \times 10^{-4} MPa^{-1}$ and semi-isotropic stress. The temperature was held at 303 K by means of a Nosé-Hoover thermostat.[13, 14] Electrostatic interactions were treated using Particle Mesh Ewald[15] with a b-spline order parameter of 4 and a grid spacing of 1.2 Å. The TIP3P water model [16] has been used. The switch-off of Lennard-Jones interactions has been set at 13 Å, with no long-range correction.

H-REM simulation setup: The GE simulation was performed using the ORAC program[6] running 10 independent sets of 24 GE replica walkers, for a total of 240 parallel processes. The Hamiltonian scaling protocol is identical for each of the 24-walkers sets and involves only the torsion angles around the sp_3 bonds of the aliphatic chains in the bilayer interior. In Fig. 1 we show the full torsional energy around a C-C bond for the target state (scaling factor 1.0) and for the highest replica (scaling factor 0.3, corresponding to a "torsional temperature" of 1000 K).

The scaling factors for the intermediate states along the replica progression were obtained as the arithmetic mean between a simple linear scaling and the standard scheme given in Ref [5]. In Fig. 1 (right), we report the dihedral angle distributions for the aliphatic torsion angles of the lipid chains in the GE and, in the inset, the jump rate between *gauche* and *trans state* as a function of the scaling state. These properties were determined by performing a 1 ns long GE simulation with no exchanges using the above defined scaling protocol and by averaging on all torsion angles of the alkyl chains of the lipids. As expected, the free energy barrier between *gauche* and *trans* states steadily decreases along the replica progression with jump rate being consequentially boosted exponentially the higher the scaling factors, thereby enhancing the conformational sampling of the bilayer.

For the production run, MPI communication groups are defined only within each of the 24-walkers batteries, that hence run independent GE simulations of 5.5 ns length, accumulating statistics for a total of 55.0 ns on the target state. Each of the 240 exchanging serial simulations are numerically integrated using an efficient multiple time step setup[17, 18], running at a speed of 0.4-0.5 ns/ day on the CRESCO1-2 platforms. The whole GE simulation on 240 processors required about a week of wall clock time on CRESCO2, amounting to a total of $\simeq$ 1700 days of CPU.

Conventional Simulation set up: The conventional simulation of POPC bilayer was performed using the GROMACS program[3] (version 4.0.7) using the same physical setup

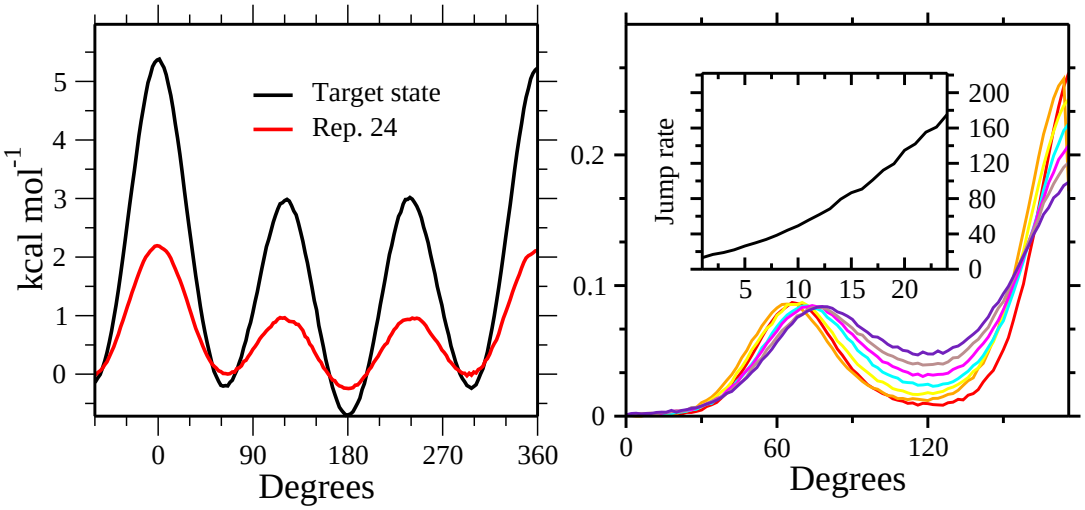

Figure 1: Left: torsional energy round the internal aliphatic C-C bond (Jämbeck type a3, see Ref. [11]); black trait, target replica; red trait, hottest replica. Right: probability distributions of the dihedral angle averaged on all aliphatic torsion angles in a non exchanging 1 ns GE simulation. Color coded scheme for distributions: from red (target state) to violet (hottest torsional state). In the inset, the jump rate per ns is reported as a function of the scaled state.

previously described. Equations of motion were integrated with a time step of 1 fs. As stated in the introduction, the moderate size of the system (about 17000 atoms) allows to run with a maximum speed of 20-25 ns per day on CRESCO2 using at most 160 processors with an efficiency of less than 50%.

Table 1: Performances of the GROMACS code on the POPC system (16374 atoms). Measures were done on the ENEA-grid CRESCO2 cluster.

Nprocs.	Speed (ns/day)	efficiency
1	0.27	1
32	8.37	0.96
64	14.85	0.85
96	19.296	0.74
160	20.900	0.48

In Table 1 we show the speed up ratio obtained on the POPC system with GROMACS. Starting form the same initial structure, we run two independent 96 processors conventional simulation of the POPC system each lasting 100 ns for about 9 days. The total amount of CPU employed was $\simeq$ 1700 days, i.e. comparable to that used in the H-REM 10-batteries simulations.

3 Results

In the following we shall assess the effect of torsional tempering in GE H-REM simulation on two key properties of biological membranes, i.e area fluctuation and lipid diffusion. The former is important for the determination of the bending rigidity of the membrane in the continuum (low wavelength) limit, while the latter determines the mixing rate and equilibration of in-homogeneous systems. In Fig. 2 we compare the time record of the

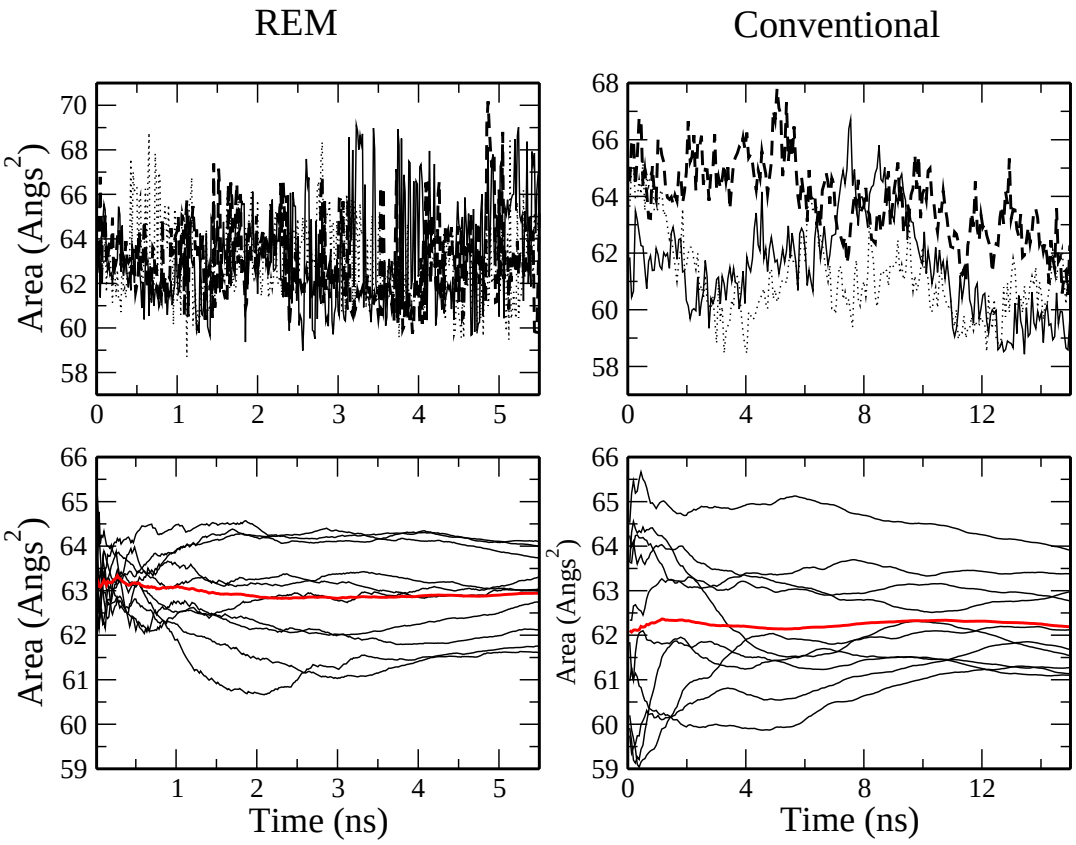

Figure 2: Top: Time record of the area per lipids in REM and conventional simulations. The time record of only three 5.5 ns H-REM batteries and three 15 ps conventional sub-simulations are shown for clarity. Bottom: Area per lipid running averages evaluated in REM for each of the 10 batteries and for the conventional simulation in 15 ns time span taken from two independent 100 ns simulations. The red curve is the cumulative running average over ten 5.5 batteries (REM) and on 200 ns of conventional simulation.

area per lipid for the H-REM and conventional simulation. In REM, the average area over all GE configurations in 5.5 ns appears to have reached a stationary value. Given the variance of the running averages recorded in each of the ten independent batteries, a more accurate results can be straightforwardly attained by simply increasing the number of the independent H-REM simulations with no impact on the wall-clock time. In the H-REM simulation, the fluctuations of the area undergo a clear boost from the GE exchanges,

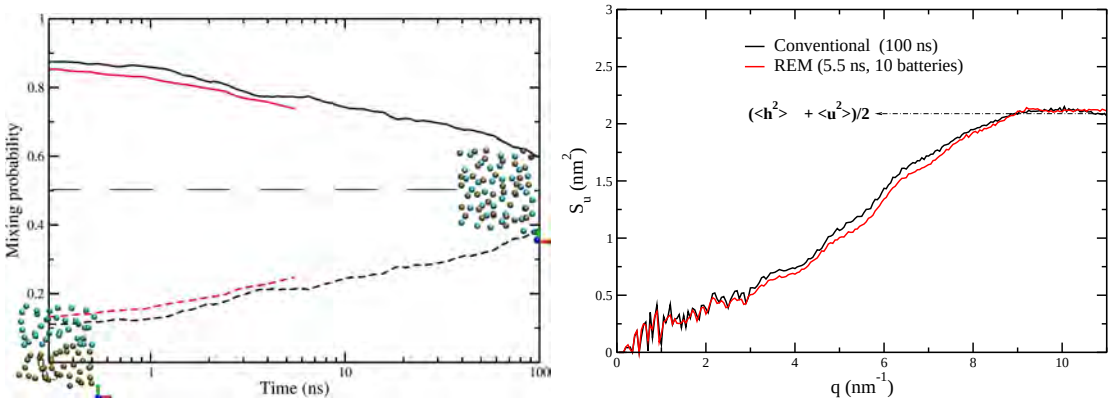

Figure 3: Left panel: Time record of the mixing probability. For the REM data (red curve) the mixing function was averaged over all GE states. Right Panel: Fourier transform of the membrane undulation function[19] $u(x,y) = \frac{1}{2}[z_1(x,y) + z_2(x,y)]$.

exhibiting variations up to 8 Å^2 in a time as small as the fraction of the picosecond. This is so since the average area per lipid of the hot GE states is sensibly larger than that of the target state increasing up to 66.7 Å^2 in the last GE state. These large area states are transmitted through the exchanges in the target replica with the correct Boltzmann weight. The conventional simulation shows in general much smaller fluctuation with respect to REM in the short time scale. Larger fluctuations can only be observed in the super-nanosecond timescale, partly inhibiting the acquisition of a stationary value of the mean and standard deviation of the area in a 100 ns time span.

In Fig. 3 (left) we show the undulation spectrum[19] of the POPC system for H-REM and conventional simulation. The two approaches yield essentially identical results, exhibiting similar trends and the same limiting value at long wavelength. Due to the moderate size of the sample, the q^4 short wavelength behavior[19, 2] could not be observed. On the right of Fig. 3 we show the mixing function for the H-REM and the conventional simulations, obtained by labeling at the start of the run with two different colors two identical sets of phosphorous atoms based on their t=0 y-coordinate. This labeling defines the unmixed zero time state, showed in the bottom left part of the plot. We then evaluate the probability of a phosphorous atom to have the nearest neighbor bearing a different label (or colors) as the simulation proceeds. In the starting state, such probability simply reflects the impact of the surface boundary separating the labeled particles. Clearly, for a perfectly mixed state such probability should tend to 0.5. In the H-REM simulation, because of the exchanges, the time ordered trajectories yield comparable probability evolution and hence one can define a time dependent probability of mixing averaged over the whole GE ensemble. The H-REM protocol produces indeed an apparent moderate acceleration of the mixing speed with respect to the conventional simulation. However, the standard MD at the end of the 100 ns shows a much more pronounced mixing with respect to the final H-REM state at 5.5 ns. Such partly disappointing outcome is due to the fact that the diffusion of the polar heads on the membrane surface is only moderately increased in H-REM via the indirect effect of enhanced torsional *gauche-trans* switching across the GE.

4 Conclusions

In Table 2 we finally collect the results obtained with H-REM (10 independent 5.5 24-replicas GE batteries) and the two independent 100 ns simulation conducted conventionally with domain decomposition. The two methods, that uses the same CPU allocation and and involve the same wall-clock time, appears to yield in essence the same physical picture of the POPC system, as long as averages lipid area and volume are concerned. Regarding fluctuations, H-REM exhibits a substantial enhancement of the time dependent area variations as well as of the amplitude of the area fluctuation producing a significantly less rigid bilayer compared to the GROMACS outcome. On the other hand H-REM promotes only a moderate increase of the polar head diffusion. These results on the overall suggest to

	$A_L(\text{Å}^2)$	$V_L(\text{Å}^3)$	K_A $(10^{-3}Nm^{-1})$	$\frac{1}{2}(\langle h^2\rangle + \langle u^2\rangle)$	G/T ratio
HREM simulation	62.9±1.2	1262.0±0.8	164±40	2.10±0.05	0.261
Standard MD simulation	62.2±1.3	1247.7±1.3	388±65	2.10±0.05	0.260
Experiments	64.3[a]±1.3	1256[b]	180-330[c]	2.1[a]± 0.1	-

Table 2: Area per lipid (A_L), volume per lipid (V_L), isothermal area compressibility modulus (K_A), S_q long wavelength limit (see Fig. 3) and *Gauche/Trans* ratio from POPC HREM simulation, standard MD simulation and experiments: **a.** from ref. [20] **b.** from ref. [21] **c.** from ref. [22].

adopt a mixed strategy in H-REM for an optimal allocation of CPU on a parallel platform: "vertical" sampling though independent GE batteries can be reduced by allowing parallel force computation within each GE trajectory, achieving "horizontal" configurational sampling, via increase of the GE simulation time span. For example, based on the data reported in Table 1, on 240 processors, a single 24-replica simulation of the POPC system can in principle be extended up to about 50 ns, using a 10 processor communication group per trajectory for parallel force computation, with little or no degradation of the parallel efficiency. Work in this direction is currently in progress.

References

[1] Nagle J. F. and Tristram-Nagle S. Structure of lipid bilayers. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes*, 1469(3):159–195, 2000.

[2] Watson Max C., Brandt Erik G., Welch Paul M., and Brown Frank L. H. Determining biomembrane bending rigidities from simulations of modest size. *Phys. Rev. Lett.*, 109:028102, Jul 2012.

[3] Hess Berk, Kutzner Carsten, van der Spoel David, and Lindahl Erik. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.

[4] Marrink S. J., Risselada H. J., Yefimov S., Tieleman D. P., and de Vries A. H. The martini force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B*, 111(27):7812–7824, 2007.

[5] Fukunishi H., Watanabe O., and Takada S. On the hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction. *The Journal of chemical physics*, 116(20):9058–9067, 2002.

[6] Marsili S., Signorini G. F., Chelli R., Marchi M., and Procacci P. Orac: A molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level. *Journal of computational chemistry*, 31(5):1106–1116, 2010.

[7] Mitsutake A., Sugita Y., and Okamoto Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Peptide Science*, 60(2):96–123, 2001.

[8] Mori T., Jung J., and Sugita Y. Surface-tension replica-exchange molecular dynamics method for enhanced sampling of biological membrane systems. *Journal of Chemical Theory and Computation*, 9(12):5629–5640, 2013.

[9] Klauda J. B. Laboratory of molecular & thermodynamic modeling. http://terpconnect.umd.edu/ jbklauda/research/download.html.

[10] In the present force field variant, the standard AMBER fudge factors were applied to *all* 14 interactions and only X-H bonds were constrained.

[11] Jämbeck J. P. M. and Lyubartsev A. P. Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *The Journal of Physical Chemistry B*, 116(10):3164–3179, 2012.

[12] Parrinello M. and Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.

[13] Nosé S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511–519, 1984.

[14] Hoover W. G. Canonical dynamics: equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985.

[15] Essmann U., Perera L., Berkowitz M. L., Darden T., Lee H., and Pedersen L. G. A smooth particle mesh ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.

[16] Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W., and Klein M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[17] Procacci P., Darden T. A., Paci E., and Marchi M. Orac: A molecular dynamics program to simulate complex molecular systems with realistic electrostatic interactions. *Journal of computational chemistry*, 18(15):1848–1862, 1997.

[18] Procacci Piero, Darden Tom, and Marchi Massimo. A very fast molecular dynamics method to simulate biomolecular systems with realistic electrostatic interactions. *The Journal of Physical Chemistry*, 100(24):10464–10468, 1996.

[19] Brandt Erik G., Braun Anthony R., Sachs Jonathan N., Nagle John F., and Olle Edholm. Interpretation of fluctuation spectra in lipid bilayer simulations. *Biophys J.*, 100:2104–2111, 2011.

[20] Kučerka N., Nieh M.-P., and Katsaras J. Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1808(11):2761–2771, 2011.

[21] Kučerka N., Tristram-Nagle S., and Nagle J. F. Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains. *The Journal of membrane biology*, 208(3):193–202, 2006.

[22] Binder H. and Gawrisch K. Effect of unsaturated lipid chains on dimensions, molecular order and hydration of membranes. *The Journal of Physical Chemistry B*, 105(49):12378–12390, 2001.