A machine learning study of the two states model for lipid bilayer phase transitions - Electronic Supplementary Informations -

Vivien WALTER
Department of Chemistry
King's College London
Britannia House, 7 Trinity Street, London, United Kingdom

Céline RUSCHER
Stewart Blusson Quantum Matter Institute
University of British Columbia
Vancouver BC V6T 1Z1, Canada

Olivier BENZERARA, Carlos M MARQUES, and Fabrice THALMANN*
Institut Charles Sadron
CNRS and University of Strasbourg
23 rue du Loess, F-67034 Strasbourg, France
(Dated: June 30, 2020)

* fabrice.thalmann@ics-cnrs.unistra.fr

I. SIMULATIONS

A. Simulation conditions

All simulations were performed using GROMACS 2016.4 [1, 2] along with the CHARMM-36 all-atom force-field [3] (June 2015 version). A lipid bilayer made of 106 lipid molecules per leaflet, each containing 130 explicit atoms, was created using CHARMM-GUI [4–7]. It was hydrated with two 8 nm thick water layers on each side (connected through periodic boundary conditions), using the TIP3P water model, for a total of 29826 solvent molecules. The force field parameters for DPPC molecules were provided directly by CHARMM-GUI [8, 9].

The above system was first subject to energy relaxation using steepest descent energy minimization, followed by a 10 ps NVT thermalization stage at 288 K. Then, the bilayer was subject to a 1 ns NPT equilibration run coupled to a semi-isotropic barostat (1 bar in all directions). The system was then further equilibrated at the desired temperature with the same semi-isotropic barostat during a second NPT equilibration step of 10 ns. Molecular dynamics production runs of 50 ns were finally generated at the same temperature and with the same semi-isotropic barostat. The analysis were performed on the last 25 ns of simulations. All time steps were set to 2 fs.

All the molecular dynamics simulations used the leap-frog integration algorithm [10]. Temperature and pressure were kept constant using respectively a Nosé-Hoover thermostat [11, 12] (correlation time $\tau_T = 0.4$ ps) and a Parrinello-Rahman semi-isotropic barostat [13, 14] (correlation time $\tau_P = 2.0$ ps, compressibility 4.5×10^{-5} bar^{-1}).

Lipid and water molecules were separately coupled to the thermostat. Following GROMACS recommendations for the CHARMM-36 all-atom force field, a Verlet cut-off scheme on grid cells was used with a distance of 1.2 nm, and non-bonded interactions cut-offs (Van der Waals and Coulombic) were also set to 1.2 nm. Fast smooth Particle-Mesh Ewald electrostatics was selected for handling the Coulombic interactions, with a grid spacing of 4 nm. A standard cut-off scheme with a force-switch smooth modifier at 1.0 nm was applied to the Van der Waals interactions. We did not account for long range energy and pressure corrections, and constrained all the hydrogen bonds of the system using the LINCS algorithm.

B. Nature of the gel phase

Whenever a large system (212 lipids or more) was simulated at low temperature, either starting from a molecular builder configuration (Charmm-gui), or resulting from the annealing of a high temperature configuration, a structure showing a longitudinal corrugation in the x and y direction was obtained (Fig. S1 right). As many other authors before, we think that this structure could actually be reminiscent of the experimental $P_{\beta'}$ DPPC ripple phase [15–17], see also [18]. In this work we refer to this corrugated phase as a **disordered gel phase** to distinguish it from the flat, tilted chains, $L_{\beta'}$ structure.

On the other hand, simulations of small systems made of 64 lipids each show much less corrugation, and looks closer to a flat tilted $L_{\beta'}$ [19] gel phase. The same holds if the 64 lipids system originates from a slow cooling of the high temperature phase. Figure S2 summarizes the various pathways leading to either a disordered, or a flat tilted structure.

The stability of the disordered gel phase was challenged by putting the system in contact with an anisotropic barostat (3 independent axis, same pressure) for 50 ns at 288 K in order to remove any box induced residual stress. The ripple phase was not perturbed except but a 6% change in the box lateral size. For all practical purposes, the bilayer behaves mechanically as a solid (fluid bilayers display large box size fluctuations when subject to an anisotropic barostat).

In addition, the ripple phase was put under tension, both under anisotropic and semi-isotropic barostat conditions, imposing a 10 mN/m stretching condition during 50 ns at 288 K. The longitudinal instability persisted and no tilted $L_{\beta'}$ emerged. The stresses and box sizes obtained at low temperature (288 K) are given in Table I.

The outcome of these "stress-tests" was that the disordered gel phase shows robustness and metastability (*e.g.* apparent stability without evidence of thermodynamic stability). Meanwhile, it is possible to duplicate a 64 lipids tilted flat configuration, and simulate it at low temperature (Fig. S2). The resulting 256 lipids solid phase also showed (meta)stability within accessible simulation times. However, once this system was heated and melted, the flat tilted configuration could not be recovered under quenching, or cooling. Only the disordered gel structure seems to be spontaneously favored upon system cooling, and reversible temperature cycling.

We therefore considered that the disordered "ripple-like" gel phase was indeed our low temperature reference phase, and performed the training and analysis on it.

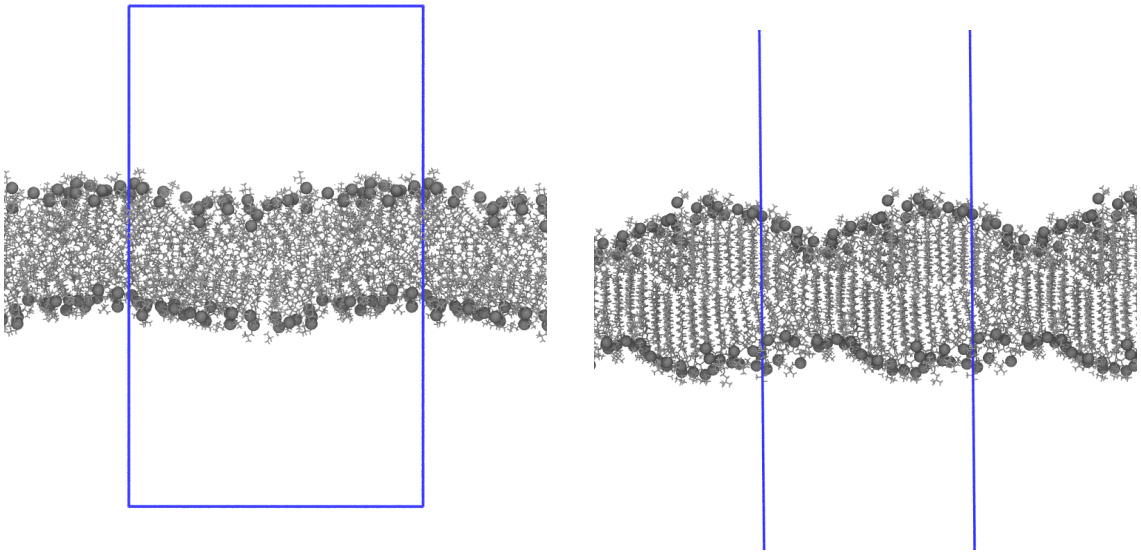

FIG. S1. Snapshots of bilayer configurations in the fluid (left) and disordered gel (right) phase, with periodic boundary conditions (box).

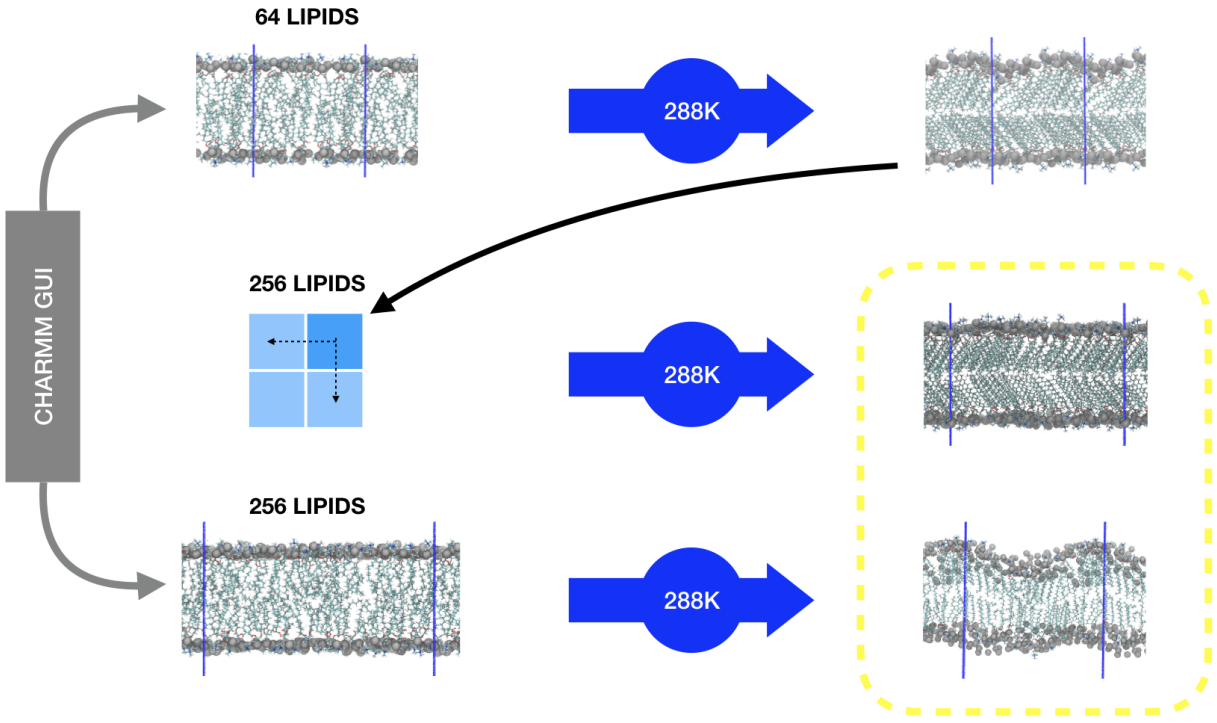

FIG. S2. Synopsis of the conditions allowing for the emergence of a disordered gel (large system, bottom row) or tilted gel (small system 64 lipids, top row and replication of the small system, middle row). The tilted lipid structures does not show up spontaneously if the number of lipids is larger than 64, and can only be found by replicating a small system.

II. SYSTEM ANALYSIS

A. Determination of structural parameters

Values of the average area per lipid A_l and order parameter $S_{\rm mol}$ of the bilayer were respectively obtained using the GROMACS built-in commands `gmx energy` and `gmx order`.

For measurements of individual lipid properties (area, order parameter, elongation), atom positions were collected from trajectories using the Python 3 `MDAnalysis` module [20, 21]. The individual areas per lipid were obtained from Voronoi tessellations of the two-dimensional projections of the lipid center of masses, computed using the `Voro++` library [22]. The individual volumes per lipid were derived from three-dimensional tessellations of the lipid centers of mass, again using the `Voro++` library. Note that the bilayer geometry requires a specific tessellation procedure: this was done by introducing *ghosts* lipids in the water regions. Without these ghost lipids, the tessellation cells cannot be correctly defined and are unbounded across the membrane-water interface, thus overestimating significantly the individual volume per lipid. Ghosts lipids are mirror images of bilayer lipids across the local lipid-water interface (*cf.* Fig. S3). After the tessellation was made, ghost lipids described in the previous section and their corresponding cells were removed the lists, and only the volumes of physical lipids were collected and analyzed.

The molecular order parameter $S_{\rm mol}$ of individual lipids was calculated by measuring, for every $N_C - 2 = 14$ non-terminal carbon atoms $k = 2 \dots 15$ of the 2 tails of the lipids, the angle formed between the z-axis of the system directed along $\vec{u}_z$ and the vector $\overrightarrow{C_{(j,k-1)}C_{(j,k+1)}}$ defined by the carbon atoms surrounding atom k within the same tail $j = 1, 2$. The order parameter $S_{{\rm mol},(j,k)}$ of the atom k is obtained from the 2^{nd} Legendre polynomial P_2 using $\cos(\theta_{(j,k)}) = \vec{u}_z \cdot \overrightarrow{C_{(j,k-1)}C_{(j,k+1)}}$, and averaging over j and k:

$$S_{\rm mol} = \frac{1}{2(N_C-2)} \sum_{j=1}^{2} \sum_{k=2}^{N_C-1} \frac{1}{2} \left(3\cos(\theta_{(j,k)})^2 - 1 \right) \tag{1}$$

B. Next-nearest neighbors statistics

After completion of the 3d Voronoi tessellation using `Voro++`, a list of next-nearest neighbors was established for each lipid center of mass. We also collected the areas of the polygonal surfaces separating each pair of neighboring Voronoi cells. The ghost lipids and their corresponding faces were removed from the lists. The neighbor lists were further curated by removing all the neighbor pairs for which the corresponding face area accounted for less than 1% of the total surface area of each Voronoi cell in contact. The number of next-nearest neighbors were finally counted to build the coordination statistics (n_g, n_f), where each lipid molecule has n_g gel and n_f fluid neighbors.

III. COMPARISON BETWEEN MACHINE LEARNING DECISIONS AND STRUCTURAL CHARACTERIZATIONS OF THE LIPID CONFIGURATIONS

Machine Learning predictions were compared to two typical lipid structural properties: the carbon carbon (CC) order parameter $S_{\rm mol}$ along the chains and the area per lipid A in the 2d Voronoi tessellation of the lipid projected centers of mass. The corresponding results are presented in Fig. S4(a) and (b).

The order parameter curves (Fig. S4(a)) clearly discriminate among the low temperature (288 K) and high temperature (358 K) fluid phases, in agreement with published results on these systems [23].

Simulation	Stress xx	Stress yy	Box x	Box y
(1)	-6	8	7.2	7.2
(2)	0.8	1	7.0	7.4
(3)	-21	-18	7.5	7.5
(4)	-20	-20	7.4	7.6

TABLE I. Mechanical resistance of the low temperature phase (288 K). Simulation conditions: (1) tensionless semi-isotropic barostat, (2) tensionless anisotropic barostat, (3) under tension semi-isotropic barostat, (4) under tension anisotropic barostat. Stress xx: virial stress in the x direction in bars. Stress yy: virial stress in the y direction. Box x: lateral x box size in nm. Box y: lateral y box size.

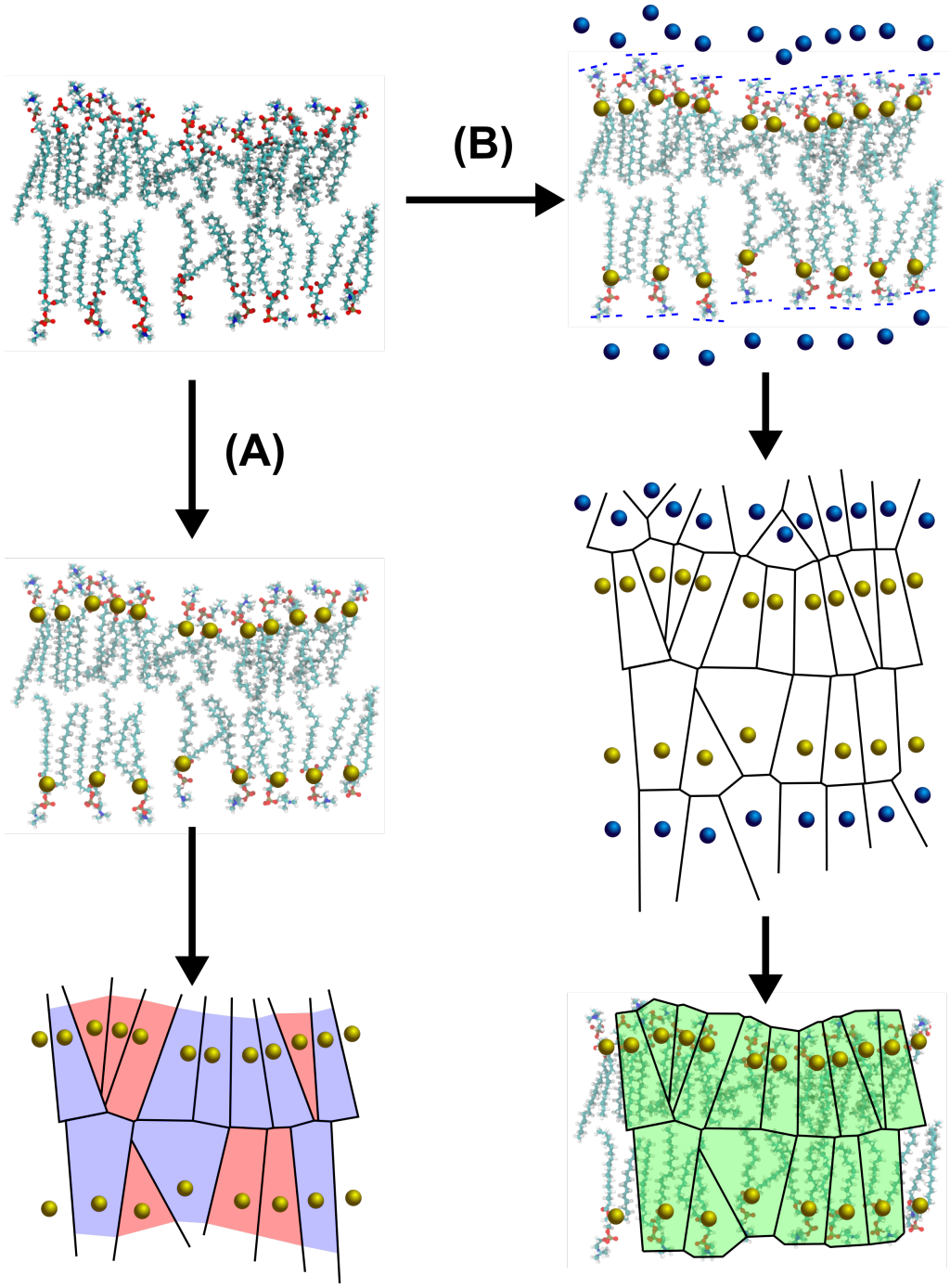

FIG. S3. Comparison of 3-dimensional Voronoi tessellations of a lipid bilayer configuration (A) without and (B) with ghost lipids. Without ghost lipids, most cells are unbounded, with infinite volume, due to the absence of particle on the opposite side of the water-membrane interface. As a practical solution of this problem, ghost lipids are added to the data set, mirroring the lipid center of mass positions. The resulting cells for lipids inside the bilayer display a realistic volume and shape, accounting for the water interface in a natural way.

The order parameter of the atoms in the lipid tails at low (288 K) and high (358 K) temperature is characteristic from membranes in the gel and fluid phases respectively [19]. The phase transition can be clearly seen in Fig. S4(b) as a significant variation in the evolution of the area per lipid A_l around 321 K.

Experimental structural values are available at 323 K [24–27]. Nagle *et al.* obtained for DPPC an area per lipid equals to 64 $\pm$ 1 Å^2 significantly close to the value *circa* 60 Å^2 we obtained in our simulations. Using the average DPPC bilayer thickness reported by Nagle *et al.*, we could estimate an experimental volume per lipid of 1220 $\pm$ 50 Å^3,

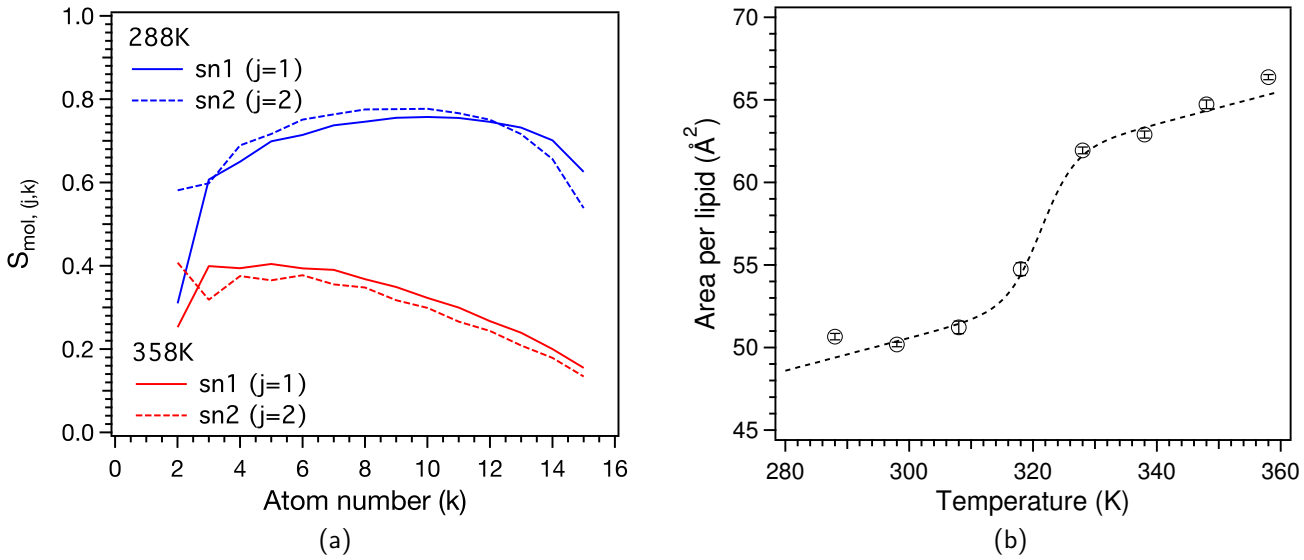

FIG. S4. Confirmation of the thermodynamic phase of the bilayer using two common structural parameters: (a) the order parameter $S_{\mathrm{mol},(j,k)}$ with k the atom number, $j = 1$ (tail sn1) or $j = 2$ (tail sn2), and (b) the area per lipid A_l. (Left) the average order parameter is shown as a function of the carbon atom index along the chain (from glycerol to terminal end), for each *sn1* and *sn2* chain. (Right) the phase transition can be clearly seen in the evolution of the area per lipid as a function of temperature. A sigmoid fit points to a transition temperature T_m equal to 321 K in our system.

which agrees fairly with our Voronoi value of 1300 Å^3

IV. NAIVE CLASSIFICATIONS

The distributions of the areas per lipid and molecular elongations at low and high temperatures are shown in Fig. S5. Using a naive classification scheme based on a single threshold value for either of the two previous scalar parameters would at best result in a prediction accuracy of respectively 69% and 67%.

Fig. S6 compares the histogram of molecular order parameters S_{mol} as a function of the temperature of the lipid bilayer from which the configurations are extracted (288 K or 358 K), and as a function of the result of the Machine Learning classification procedure. The difference between the distribution at 288 K and the distribution in the *a posteriori* gel state ensemble indicates that a small fraction of lipids in the fluid state are already present at 288 K.

[1] H Berendsen, D van der Spoel, and R van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 1995.
[2] M J Abraham, T Murtola, R Schulz, S Páll, J C Smith, B Hess, and E Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 2015.
[3] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell Jr. Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain khi1 and khi2 dihedral angles. *Journal of Chemical Theory and Computation*, 8(9):3257–3273, 2012.
[4] S. Jo, T. Kim, V. G. Iyer, and W. Im. Charmm-gui: A web-based graphical user interface for charmm. *Journal of Computational Chemistry*, 29:1859–1865, 2008.
[5] E. L. Wu, X. Cheng, S. Jo, H. Rui, H. K. Song, E. M. Davila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R. M. Venable, J. B. Klauda, and W. Im. Charmm-gui membrane builder toward realistic biological membrane simulations. *Journal of Chemical Theory and Computation*, 35:1997–2004, 2014.
[6] S. Jo, J. B. Lim, J. B. Klauda, and W. Im. Charmm-gui membrane builder for mixed bilayers and its application to yeast membranes. *Biophysical Journal*, 97:50–58, 2009.
[7] S. Jo, T. Kim, and W. Im. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS ONE*, 2(9):880, 2007.
[8] B. R. Brooks, C. L. Brooks III, A. D. MacKerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels,

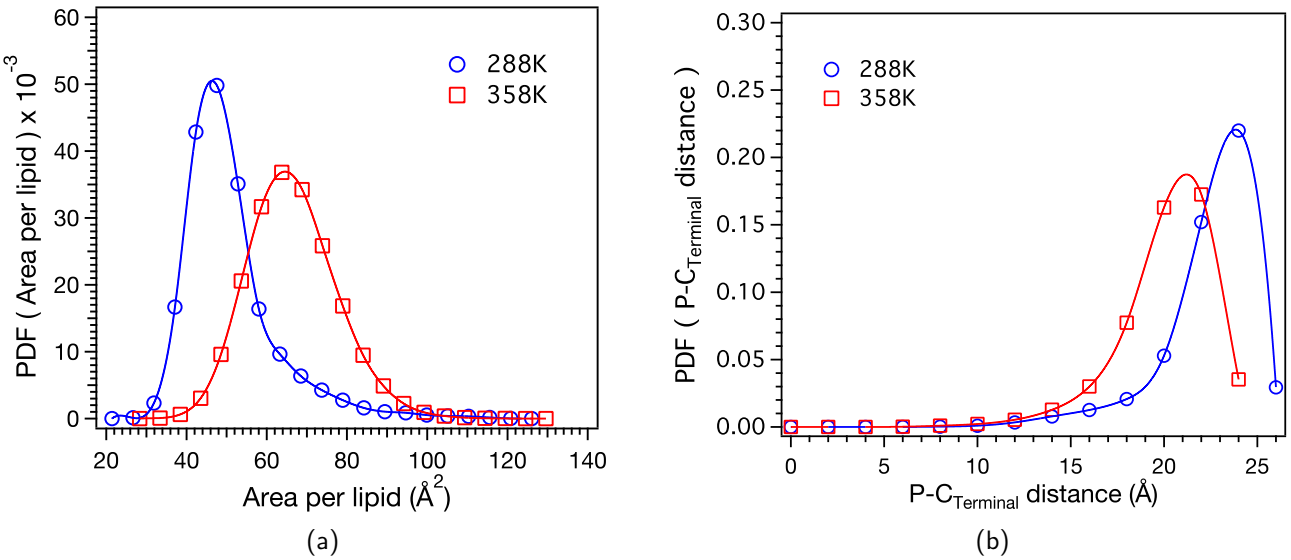

FIG. S5. Distributions of the area per lipid (Left) and the average elongation between phosphorus and *sn1* terminal carbon atoms (Right) from lipids conformations at 288 and 358 K.

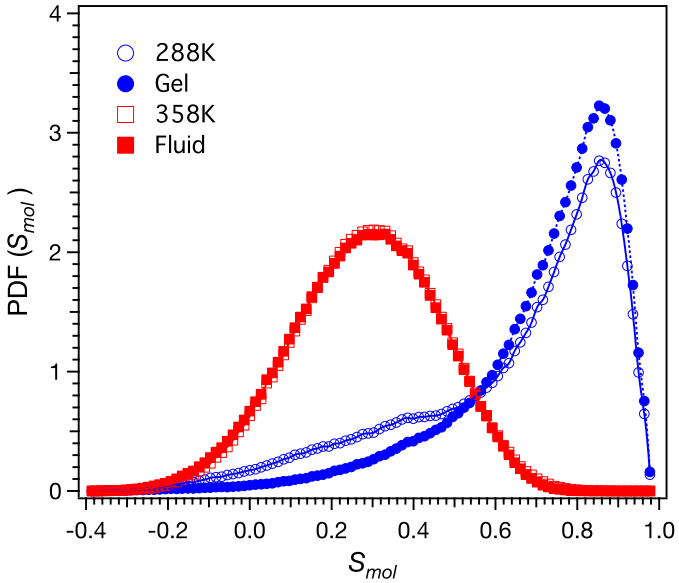

FIG. S6. Histograms of the molecular order parameter S_{mol} of lipids sorted by temperature (blue circles) and state ML classification (red squares).

S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. Charmm: The biomolecular simulation program. *Journal of Computational Chemistry*, 30:1545–1614, 2009.

[9] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks III, A. D. MacKerell Jr, J. B. Klauda, and W. Im. Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field. *Journal of Chemical Theory and Computation*, 12(1):405–413, 2016.

[10] R. W. Hockney, S. P. Goel, and J. W. Eastwood. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics*, 14(2):148–158, 1974.

[11] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[12] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31:1695, 1985.
[13] S. Nose and M. L. Klein. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, 50:1055–1076, 1983.
[14] M. Parinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52:7182, 1998.
[15] Alex H. de Vries, Serge Yefimov, Alan E. Mark, and Siewert J. Marrink. Molecular structure of the lecithin ripple phase. *Proceedings of the National Academy of Sciences*, 102(15):5392–5396, 2005.
[16] Olaf Lenz and Friederike Schmid. Structure of symmetric and asymmetric "ripple" phases in lipid bilayers. *Phys. Rev. Lett.*, 98(5):058104, 2007.
[17] Kiyotaka Akabori and John F. Nagle. Structure of the DMPC lipid bilayer ripple phase. *Soft Matter*, 11(5):918–926, 2015.
[18] Ananya Debnath, Foram M. Thakkar, Prabal K. Maiti, V. Kumaran, and K. G. Ayappa. Laterally structured ripple and square phases with one and two dimensional thickness modulations in a model bilayer system. *Soft Matter*, 10(38):7630–7637, 2014.
[19] Gregor Cevc and Derek Marsh. *Phospholipid Bilayers. Physical Principles and Models.* John Wiley & Sons, New-York, 1987.
[20] R J Gowers, M Linke, J Barnoud, T J E Reddy, M N Melo, S L Seyler, D L Dotson, J Domanski, S Buchoux, I M Kenney, and O Beckstein. Mdanalysis: A python package for the rapid analysis of molecular dynamics simulations. *Proceedings of the 15th Python in Science Conference*, pages 98–105, 2016.
[21] N MichaudAgrawal, E J Denning, T B Woolf, and O Beckstein. Mdanalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 32(10):2319–2327, 2011.
[22] C R Rycroft. Voro++: A three-dimensional voronoi cell library in c++. *Chaos*, 19:041111, 2009.
[23] Jeffery B. Klauda, Richard M. Venable, J. Alfredo Freites, Joseph W. OConnor, Douglas J. Tobias, Carlos Mondragon-Ramirez, Igor Vorobyov, Alexander D. MacKerell, and Richard W. Pastor. Update of the charmm all-atom additive force field for lipids: Validation on six lipid types. *Journal of Physical Chemistry B*, 114(23):7830–7843, 2010.
[24] J F Nagle, R Zhang, S Tristram-Nagle, W Sun, H I Petrache, and R M Suter. X-ray structure determination of fully hydrated l alpha phase dipalmitoylphosphatidylcholine bilayers. *Biophysical Journal*, 70(3):1419–1431, 1996.
[25] J F Nagle and S Tristram-Nagle. Structure of lipid bilayers. *Biochimica and Biophysica Acta*, 1469(3):159195, 2000.
[26] N Kuçerka, S Tristram-Nagle, and J F Nagle. Closer look at structure of fully hydrated fluid phase dppc bilayers. *Biophysical Journal*, 90(11):L83–L85, 2006.
[27] N Kuçerka, J F Nagle, J N Sachs, S E Feller, J Pencer, A Jackson, and J Katsaras. Lipid bilayer structure determined by the simultaneous analysis of neutron and x-ray scattering data. *Biophysical Journal*, 95(5):23562367, 2008.