Supporting information for: Quantifying lateral inhomogeneity of cholesterol-containing membranes

Celsa Díaz-Tejada, Igor Ariz-Extreme, Neha Awasthi, and Jochen S. Hub*

Georg-August-University Göttingen, Institute for Microbiology and Genetics, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

E-mail: jhub@gwdg.de

Phone: +49-551-3914189

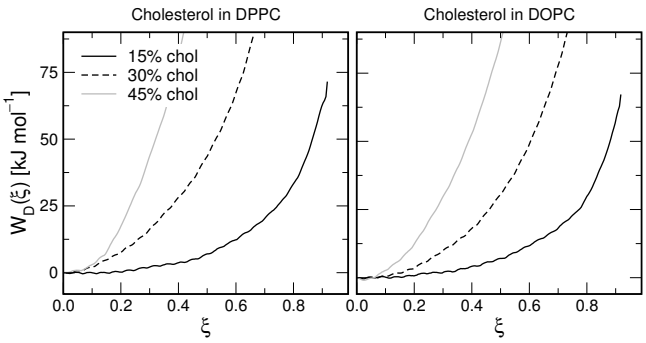

Figure S1: PMFs for the formation of a cholesterol-depleted domain, versus the reaction coordiante ξ by Tolpekina *et al.*[1] ξ is defined in the supporting information text. Left: in DPPC; right: in DOPC. Cholesterol concentration highlighted by line style (see legend).

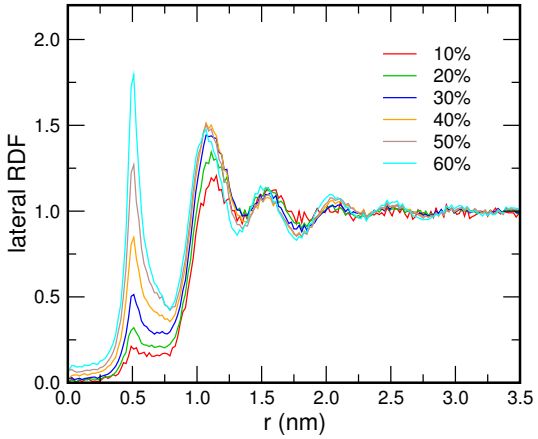

Figure S2: Lateral radial distribution of functions (RDFs) of cholesterol in DOPC at cholesterol concentrations between 10% and 60% (for color coding see legend). The abscissa r denotes the center-of-mass distance between two cholesterol molecules projected onto the membrane plane. RDFs were computed using only pairs of cholesterol molecules in the same lipid monolayer.

Methodological details

Coarse-grained simulations

All simulations were conducted with the Gromacs simulation software.[2] Molecular interactions were taken from the Martini 2.0 force field.[3] Cholesterol parameters were taken from a more recent Martini update,[4] allowing for an integration time step of 30 fs. An exception is the simulations used to compute the control PMFs that did not contain any additional

cholesterol (Fig. 2C, black curves). For these simulations only, cholesterol parameters prior to the update by Ingolfsson *et al.* and a 18 fs time step were applied. All simulation parameters were chosen according to the default Martini parameters.[3] Accordingly, Coulomb interactions were cut-off at a 1.2 nm, and Lennard-Jones interactions where shifted to zero between 0.9 and 1.2 nm using the Gromacs shift function.[2] The temperature of all simulations was controlled at 323 K using a stochastic dynamics integrator,[5] and the pressure was kept at 1 bar using a semi-isotropic weak coupling scheme.[6]

PMFs across cholesterol-enriched slabs

The cholesterol-enriched slabs were enforced by a flat-bottomed quadratic potential $V_{\mathrm{fb}}(y)$ acting on the centers of mass (COMs) of cholesterol, implemented as $V_{\mathrm{fb}}(y) = k_{\mathrm{fb}}(|y| - y_{\mathrm{fb}})^2/2\, H(|y| - y_{\mathrm{fb}})$. Here, y is the coordinate in the membrane plane perpendicular to the cholesterol-enriched slab (Fig. 2B). $k_{\mathrm{fb}} = 1000\,\mathrm{kJ\,mol^{-1}nm^{-2}}$ is the force constant, $2y_{\mathrm{fb}}$ the thickness of the slab, which was chosen as half of the box width in y-direction. H denotes the Heaviside step function. The simulation systems contained ∼300 phospholipids, ∼2200 coarse-grained (CG) water beads, and between zero and 112 cholesterol molecules, corresponding to a cholesterol mole fraction of up to 50 mol% in the cholesterol-enriched slab. Random frames were taken from equilibrium simulations for starting configurations for umbrella sampling simulations.[7] The coordinate y perpendicular to the cholesterol slab was taken as reaction coordinate, which was split into ∼110 1Å-wide equidistant umbrella windows. One cholesterol molecule was inserted at the center of the umbrella window, and restrained with an umbrella potential (force constant $500\,\mathrm{kJ\,mol^{-1}nm^{-2}}$). To remove atomic overlaps, interactions between the inserted cholesterol and the rest of the system were gradually switched on along an alchemical reaction coordinate. Subsequently, the energy of the structure was minimized, and each umbrella window was simulated for 2μs. The PMFs were computed using the weighted histogram analysis method (WHAM), omitting the first 50 ns for equilibration.[8]

Error estimates for PMFs

Statistical errors were computed using the Bayesian bootstrap of complete histograms, as implemented into the software g_wham with options "-bs-method b-hist".[9] This method does *not* try to estimate the autocorrelation times *within* individual umbrella sampling trajectories, and it does not try to split individual trajectories into time bins, with the hope that these time bins would be statistically independent. Because MD trajectories frequently suffer from long and unknown autocorrelations, such procedures may underestimate the statistical error.[9]

Instead, bootstrapping of complete histograms, as employed here, considers only complete umbrella histograms as statistically independent data, thereby leading to larger and presumably more realistic error estimates. However, the method requires (a) highly overlapping histograms, such that each point along the reaction coordinate is covered by multiple histograms; and (b) that these available umbrella histograms represent the accessible phase space. Systems of DPPC or POPC at high cholesterol content entered a gel-like phase, leading to very long autocorrelations, such that the accessible phase space may not be fully sampled. Consequently, we cannot exclude that statistical errors reported here for high cholesterol content are underestimated (e.g., Fig. 2B, red and magenta lines at $x_c = 50\%$). At lower cholesterol content, and in systems containing DOPC and DUPC, no indications of poor sampling were observed, suggesting that the error estimates are correct.

Excess chemical potentials

The excess chemical potential (or free energies for insertion) of one cholesterol molecule into lipid membranes μ_{ex} was computed using thermodynamic integration (TI). Lipid membrane patches containing one out of four types of phospholipid (DPPC, POPC, DOPC, or DUPC) plus a cholesterol content between 0 and 60 mol% were built with the Insane tool.[10] The systems contained between 160 and 286 cholesterol-lipid molecules, and the membranes were fully hydrated with 1727 to 3094 coarse-grained water beads. All systems were simulated

until the box dimensions and the potential energy were fully converged. One additional cholesterol molecule was inserted in each system at a random position into one leaflet, and the system was again equilibrated. Subsequently, TI was conducted along an alchemical reaction coordinate λ, which scales the interactions between the additional cholesterol molecule and all other molecules. Accordingly, $\lambda = 0$ corresponds to a non-interacting cholesterol molecule, and $\lambda = 1$ to the fully interacting state. For membranes containing DPPC, POPC, and DOPC, the following 43 λ-values were used: 0, 0.05, 0.1, 0.125, 0.14, 0.145, 0.15, 0.155, 0.16, 0.165, 0.17, 0.175, 0.18, 0.185, 0.19, 0.195, 0.2, 0.205, 0.21, 0.215, 0.22, 0.225, 0.23, 0.235, 0.24, 0.245, 0.25, 0.275, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1. For membranes containing DUPC, the following 25 λ-values were used: 0, 0.05, 0.1, 0.125, 0.15, 0.175, 0.2, 0.225, 0.25, 0.275, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1. Each λ-window was simulated for 1 μs, where the first 100 ns were discarded for equilibration. Lennard-Jones potentials were turned on using soft-core potentials. No separate TI-calculations for turning on the partial charges were required because the Martini cholesterol model does not contain partial charges. μ_{ex} was computed via the integral $\mu_{\text{ex}} = \int_0^1 \langle \partial H/\partial\lambda \rangle \mathrm{d}\lambda$. The statistical error of $\langle \partial H/\partial\lambda \rangle$ was computed by binning analysis.[11] Subsequently, the statistical error of μ_{ex} was given by straight-forward error propagation. To validate the procedure, μ_{ex} was also computed using the Bennet's acceptance ratio (BAR) method, which yielded nearly identical results.[12]

PMFs for domain formation

PMFs for the formation of a circular cholesterol-depleted domain were computed using the reaction coordinate introduced by Tolpekina *et al.*[1] Accordingly, we conducted umbrella sampling along the coordinate $\xi = (\Sigma - \Sigma_0)/(N - \Sigma_0)$, where $\Sigma = \sum_{i=1}^{N} \tanh(r_i/\zeta)$. Here, N is the number of cholesterol molecules, r_i denotes the COM distance in the membrane plane of cholesterol i from the center of the cholesterol-depleted domain. Σ_0 is the equilibrium value of Σ, which can be computed analytically assuming a random distribution of cholesterol in

the membrane. ζ specifies the approximate radius of the domain when ξ is close to one, and it was here chosen as $\zeta = 1.5\,\mathrm{nm}$. Hence, the reaction coordinate ξ is normalized such that $\xi = 0$ corresponds to a homogeneous cholesterol distribution, and ξ close to one corresponds to the fully formed domain. The PMFs as a function of ξ are shown in Fig. S1. Because ξ is little intuitive, we translated the PMFs into functions of the approximate radius R of the domain, where R was defined as the radius where cholesterol density reached 50% of the value far away from the domain center (Fig. 4B/C).

The simulations systems contained ∼880 phospholipid-cholesterol molecules, and ∼13.000 CG water beads. The total area of the equilibrated square membranes was between 198 and $263\,\mathrm{nm}^2$. After building and equilibrating the membrane, the domain was generated by pulling the system along ξ from the homogeneous system to a fully formed domain. Subsequently, we conducted umbrella sampling between $\xi = 0$ and $\xi = 0.9$ using 21 equally-spaced umbrella windows. A force constant of 12500 was applied along ξ, which yielded sufficient overlap between adjacent histograms. Each window was simulated for 1 μs, omitting the first 40 ns for equilibration. The PMFs $W(\xi)$ were again computed using WHAM, and the errors estimated by bootstrap analysis (Fig. S1), suggesting standard errors that are smaller than $1\,\mathrm{kJ\,mol^{-1}}$.[8,9]

The loss of lateral entropy $\Delta S_{\mathrm{lat}}(R) = S_{\mathrm{lat}}(R) - S_{\mathrm{lat}}(R = 0)$ upon domain formation was quantified via the Shannon entropy $S_{\mathrm{lat}}(R) = -k_B \int_{\mathrm{box}} p_R(x, y) \ln p_R(x, y)\,\mathrm{d}x\mathrm{d}y$, where $p_R(x, y)$ denotes the lateral density of cholesterol at domain radius R, taken from the respective umbrella sampling simulations. The lateral entropy in the absence of a domain, $S_{\mathrm{lat}}(R = 0)$, was computed analytically by assuming a homogeneous cholesterol density $p_{R=0}(x, y) = N/A_b$, where A_b is the membrane area.

The contribution $\Delta W_\mu(R)$ to the PMFs for domain formation was computed as follows. Upon formation of the circular cholesterol-depleted domain, $N_R = \rho\pi R^2$ cholesterol molecules must be transferred from the area inside to the area outside of the domain. Here, $\rho = N/A_b$ is the average cholesterol 2D density. Because the domain is much smaller than

the total membrane area, we approximate the cholesterol concentration x_c outside of the domain as constant. Then, we have

$$\Delta W_\mu(R) = \int_0^{N_R} \left[\Delta\mu(x_c^D(N_R')) - \Delta\mu(x_c)\right] \, \mathrm{d}N_R', \tag{1}$$

where $x_c^D(N_R') = N_R' x_c/(\rho\pi R^2)$ is the cholesterol concentration inside the domain if N_R' cholesterol molecules are left. Approximating the curves in Fig. 2D by $\Delta\mu(x_c) = m_1 x_c + m_2 x_c^2$, yields for the integral $\Delta W_\mu(R) = \pi R^2 \rho x_c(\frac{1}{2}m_1 + \frac{2}{3}m_2 x_c)$.

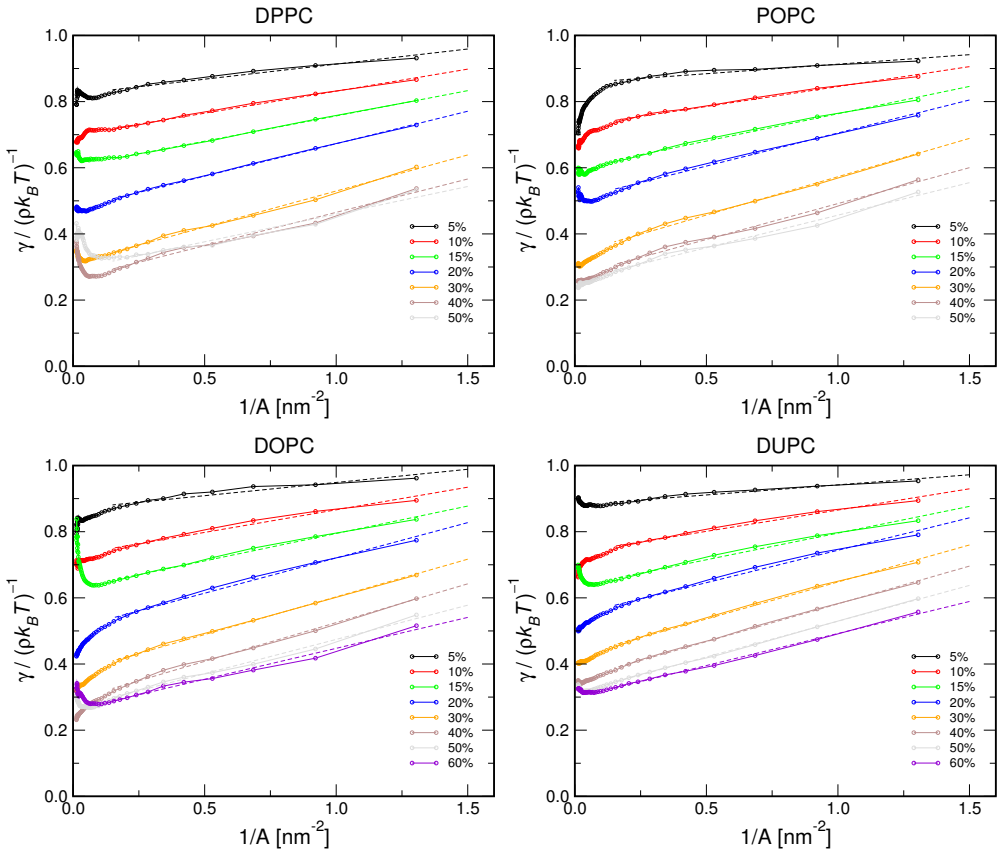

Figure S3: The quantity $\gamma(A)$, which is proportional to the variance of cholesterol number, versus the inverse probe area $1/A$ (see eq. 2), for systems containing DPPC (top left), POPC (top right), DOPC (bottom left), and DUPC (bottom right). The color indicates cholesterol mole fraction x_c (see legends). Dashed lines: linear fit at moderate to high $1/A$. The normalized lateral compressibility $\chi_c/(\rho k_B T)^{-1}$ of cholesterol was taken as the y-intercept of the linear fits.

Compressibility calculation

The lateral compressibility was computed from equilibrium simulations that contained between 5200 and 6500 lipid-cholesterol molecules and ∼75000 CG water beads. The total area of the equilibrated square membranes was between 1165 and 1655 nm^2. The simulation systems contained one type of phospholipid plus a cholesterol content between 5 and 60 mol%. Each system was simulated for 1 μs. Phospholipid, cholesterol and water beads were coupled separately to a heat bath of 323 K using the weak coupling scheme ($\tau = 1\,\mathrm{ps}$).[6] All other parameters were chosen as described above.

The compressibility was computed from the equilibrium MD simulations as follows. First, the quantity

$$\gamma(A) = \frac{\langle N_a^2\rangle - \langle N_a\rangle^2}{\langle N_a\rangle^2}\frac{A}{k_B T}, \tag{2}$$

was computed using various probe areas A, where N_a is the number of cholesterol molecules within the probe area. The probe area was taken as the area in the x-y-plane, that is, effects from membrane undulation on the N_a were neglected. The quantity $\gamma(A)$ for the DOPC systems is plotted versus $1/A$ in Fig. S3 as solid lines. At large probe areas (small $1/A$), $\gamma(A)$ is affected by (i) finite size effects since we simulated at constant cholesterol number and not in the grand canonical ensemble; and (ii) by poor sampling since independent estimates of $\gamma(A)$ require the diffusion of many cholesterol molecules over long distances. At moderate to low probe areas (moderate to large $1/A$), $\gamma(A)$ versus $1/A$ exhibits an approximately linear regime. The increase in $\gamma(A)$ in that regime was previously attributed to "missing correlations" near the boundary of the probe area.[13–15] The compressibility can be computed by extrapolating $\gamma(A)$ to $A \to \infty$ using only the data poins at moderate to small A.

Hence, a line was fitted to $\gamma(A)$ in that linear regime (Fig. S3, dashed lines), between $1/A = 0.15\,\mathrm{nm}^{-2}$ and $1.5\,\mathrm{nm}^{-2}$, and $\chi_\mathrm{c}/(\rho k_B T)^{-1}$ was taken as the y-intercept of the fitted line. We note that, in contrast to Rovere *et al.*, we here linearly extrapolated along $1/A$ and not along $1/\sqrt{A}$.[13–15] Visual inspection of the curves suggested that linear fits accurately

model the $\gamma(A)$-vs-$1/\sqrt{A}$ curves only at small cholesterol content. At higher cholesterol content, however, any attempts to linearly extrapolate along $1/\sqrt{A}$ gave unsatisfactory fits. We therefore decided to extrapolate linearly along $1/A$, noting that the scaling of $\gamma(A)$ with A deserves further attention.

The error in χ_c was taken from the uncertainty of the fit. Noteworthy, we computed $\chi_c/(\rho k_B T)^{-1}$ using either all cholesterol molecules in the double layer, or restricted to cholesterol molecules within a monolayer. As expected, the compressibility of the double-layer was approximately half of that of a monolayer. Hence, χ_c shown in the figures was computed from the double-layer, but normalized to a monolayer (by multiplying by two). That procedure avoided some uncertainty from cholesterol molecules that are at times located horizontally at the center of the membranes, in particular in the highly unsaturated DUPC membranes.[16]

References

(1) Tolpekina, T.; Den Otter, W.; Briels, W. Nucleation Free Energy of Pore Formation in an Amphiphilic Bilayer Studied by Molecular Dynamics Simulations. *J. Chem. Phys.* **2004**, *121*, 12060–12066.

(2) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D. et al. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(3) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

(4) Ingólfsson, H. I.; Melo, M. N.; van Eerden, F. J.; Arnarez, C.; Lopez, C. A.; Wasse-

naar, T. A.; Periole, X.; De Vries, A. H.; Tieleman, D. P.; Marrink, S. J. Lipid Organization of the Plasma Membrane. *J. Am. Chem. Soc.* **2014**, *136*, 14554–14559.

(5) van Gunsteren, W. F.; Berendsen, H. J. C. A Leap-Frog Algorithm for Stochastic Dynamics. *Mol. Sim.* **1988**, *1*, 173–185.

(6) Berendsen, H. J. C.; Postma, J. P. M.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(7) Torrie, G. M.; Valleau, J. P. Monte Carlo Free Energy Estimates Using Non-Boltzmann Sampling: Application to the Sub-Critical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974**, *28*, 578–581.

(8) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comp. Chem.* **1992**, *13*, 1011–1021.

(9) Hub, J. S.; de Groot, B. L.; van der Spoel, D. G_wham–A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713–3720.

(10) Wassenaar, T. A.; Ingólfsson, H. I.; Böckmann, R. A.; Tieleman, D. P.; Marrink, S. J. Computational Lipidomics with Insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 2144–2155.

(11) Hess, B. Determining the Shear Viscosity of Model Liquids from Molecular Simulations. *J. Chem. Phys.* **2002**, *116*, 209–217.

(12) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245–268.

(13) Rovere, M.; Hermann, D.; Binder, K. Block Density Distribution Function Analysis of Two-Dimensional Lennard-Jones Fluids. *Europhys. Lett.* **1988**, *6*, 585.

(14) Rovere, M.; Heermann, D.; Binder, K. The Gas-Liquid Transition of the Two-Dimensional Lennard-Jones Fluid. *J. Phys.: Condens. Matter* **1990**, *2*, 7009.

(15) Rovere, M.; Nielaba, P.; Binder, K. Simulation Studies of Gas-Liquid Transitions in Two Dimensions Via a Subsystem-Block-Density Distribution Analysis. *Z. Phys.* **1993**, *90*, 215–228.

(16) Marrink, S. J.; de Vries, A. H.; Harroun, T. A.; Katsaras, J.; Wassall, S. R. Cholesterol Shows Preference for the Interior of Polyunsaturated Lipid Membranes. *J. Am. Chem. Soc.* **2008**, *130*, 10–11.