

Cryo-EM images of phase-separated lipid bilayer vesicles analyzed with a machine-learning approach

Karan D. Sharma,¹ Milka Doktorova,² M. Neal Waxham,³ and Frederick A. Heberle^{1,*}

¹Department of Chemistry, University of Tennessee, Knoxville, Tennessee; ²Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia; and ³Department of Neurobiology and Anatomy, University of Texas Health Science Center, Houston, Texas

ABSTRACT Lateral lipid heterogeneity (i.e., raft formation) in biomembranes plays a functional role in living cells. Three-component mixtures of low- and high-melting lipids plus cholesterol offer a simplified experimental model for raft domains in which a liquid-disordered (Ld) phase coexists with a liquid-ordered (Lo) phase. Using such models, we recently showed that cryogenic electron microscopy (cryo-EM) can detect phase separation in lipid vesicles based on differences in bilayer thickness. However, the considerable noise within cryo-EM data poses a significant challenge for accurately determining the membrane phase state at high spatial resolution. To this end, we have developed an image-processing pipeline that utilizes machine learning (ML) to predict the bilayer phase in projection images of lipid vesicles. Importantly, the ML method exploits differences in both the thickness and molecular density of Lo compared to Ld, which leads to improved phase identification. To assess accuracy, we used artificial images of phase-separated lipid vesicles generated from all-atom molecular dynamics simulations of Lo and Ld phases. Synthetic ground-truth data sets mimicking a series of compositions along a tieline of Ld + Lo coexistence were created and then analyzed with various ML models. For all tieline compositions, we find that the ML approach can correctly identify the bilayer phase with >90% accuracy, thus providing a means to isolate the intensity profiles of coexisting Ld and Lo phases, as well as accurately determine domain-size distributions, number of domains, and phase-area fractions. The method described here provides a framework for characterizing nanoscopic lateral heterogeneities in membranes and paves the way for a more detailed understanding of raft properties in biological contexts.

SIGNIFICANCE Lipid rafts are important for cell function, but in most cases they cannot be detected with conventional optical microscopy because of their extremely small size. Cryogenic electron microscopy (cryo-EM), because of its much greater spatial resolution, is capable of imaging domains as small as 5–10 nm. In this report, we show how machine-learning techniques can be used to automatically and accurately identify raft-like domains in simulated cryo-EM images, a powerful approach that could ultimately lead to a better understanding of raft properties.

INTRODUCTION

The membranes that surround and partition living cells can play both structural and functional roles. The latter is exemplified by the lipid raft hypothesis, which proposes the existence of plasma membrane (PM) domains enriched in sphingolipids and cholesterol that are thought to direct the spatial organization of membrane proteins and thus mediate processes such as cell signaling (1). Multiple lines of evidence have converged on a picture of raft domains as small (<20 nm) and highly dynamic in resting conditions but with

the ability to coalesce into larger structures in response to certain stimuli (2). The structural and elastic properties of the raft and non-raft environments can differ substantially, with the latter being thinner, easier to bend, less viscous, and more compressible owing to a greater proportion of unsaturated lipids (3–5). These local membrane properties influence the free energy of protein conformations (6) and can result in distinct phase preferences for different proteins (7). In the raft model of PM organization, the size, connectivity, and composition of both raft and non-raft domains are all variables that can affect the local concentrations of membrane proteins and, thus, the frequency of encounters between interaction partners.

One of the great challenges in studying lipid rafts is their small size, which precludes direct observation by

Submitted December 22, 2023, and accepted for publication April 26, 2024.

*Correspondence: fheberle@utk.edu

Editor: Tamar Schlick.

<https://doi.org/10.1016/j.bpj.2024.04.029>

© 2024 Biophysical Society.

conventional optical microscopy. Instead, researchers have relied heavily on biochemical and spectroscopic data to infer the presence or absence of multiple membrane environments in cell membranes. While the lack of visual evidence initially provided fuel for controversy, those concerns have been largely put to rest by the sheer quantity of indirect data that consistently points toward heterogeneous membranes as a rule rather than an exception (8). Yet even as the “seeing is believing” debate recedes into the background, a fact often lost in the discussion is that the utility of image data extends far beyond visual confirmation of rafts. To take one example, images can reveal spatial domain patterns that might help distinguish between competing theories of raft formation (9). Image data also open a window to heterogeneity in domain structures that are often obscured in ensemble-averaged spectroscopic or scattering measurements, thus providing a more complete picture of rafts and their role in specific biological processes.

The great wealth of information contained in suboptical images is exemplified by the emergence of single-particle cryogenic electron microscopy (cryo-EM) as a pre-eminent tool in structural biology. A crucial breakthrough that has accelerated its rise is the use of machine learning (ML) at multiple stages of the experimental workflow (10) including image denoising (11), particle picking (12), image segmentation (13), and 3D reconstruction (14,15). Far from being limited to water-soluble proteins, cryo-EM is increasingly being used to determine structures of membrane proteins in native or near-native environments (16–20). As the focus of these studies has been fixed squarely on the protein, their membrane hosts have received less attention, yet bilayers also appear in images at a resolution high enough to observe density variation in the separate leaflets. This raises the possibility that, when visualized with cryo-EM, relatively thicker raft domains might be distinguishable from the thinner “sea” of mostly unsaturated lipids that surrounds them.

To this end, we recently developed a method for obtaining local intensity profiles along the projected circumference of lipid vesicles in cryo-EM images (21), which we then used to measure the local bilayer thickness at 5-nm lateral resolution in liposomes and vesicles derived from cell plasma membranes (21) as well as HIV pseudoviral membranes (22). When thicknesses from a large population of vesicles were plotted as a histogram, the appearance of the distribution (i.e., unimodal or multimodal) was consistent with the number of coexisting phases determined in independent experiments. Concurrently with our study, the Keller group performed similar analyses and reached similar conclusions using tomographic reconstructions of cryopreserved liposomes (23). These outcomes were expected, since it is well established that ordered phases are 5–10 Å thicker than disordered phases (3,24,25). However, while those studies demonstrated critical proof of principle for using

cryo-EM to investigate lipid phase separation, its unique capabilities as an imaging technique have not yet been fully realized. Most importantly, the development of a robust, automated method for determining the membrane phase state with high spatial resolution would enable additional valuable characterization including domain-size distributions and an assessment of heterogeneity within the sample (26).

Although the local phase state of the membrane can in principle be determined from spatially resolved bilayer thickness measurements such as those in (21), in practice this is hampered by the substantial noise inherent to cryo-EM images, which results in broad and overlapping thickness distributions for Ld and Lo phases. One consequence is that a relatively large portion of the projected bilayer cannot be unambiguously classified as Ld or Lo by thickness alone. However, bilayer thickness is not the only characteristic that distinguishes ordered and disordered phases. Equally important from the perspective of electron scattering in image formation is the difference in molecular packing density, which manifests as a greater intensity contrast for ordered phases compared to disordered phases (27). Here, we show that ML models trained to recognize both thickness and intensity contrast can predict the local bilayer phase state with significantly greater fidelity than methods based on thickness measurements alone. Key to this conclusion is the use of *in silico* data that allow us to establish a ground truth for assessing the accuracy of phase determination for a variety of methods, including both unsupervised and supervised ML techniques. We also describe additional structural information that can be gleaned from segmented images such as area fractions and average intensity profiles of ordered and disordered phases, and measurements of domain size and number.

MATERIALS AND METHODS

Analysis software

The analyses described in the following sections were implemented in Wolfram Mathematica 13.2 (Wolfram Research, Champaign, IL) unless stated otherwise.

Molecular dynamics simulations of Ld and Lo bilayers

All-atom molecular dynamics (MD) simulations of three-component mixtures containing 1,2-distearoyl-*sn*-glycero-3-phosphocholine (DSPC), 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC), and cholesterol (Chol) were performed with NAMD (28) using the CHARMM36 force-field parameters (29,30). Two compositions were simulated, corresponding to the endpoints of a tieline in the Ld + Lo region of the room-temperature phase diagram (Fig. S1). Each bilayer contained 100 lipids per leaflet, 50 waters per lipid, and no ions, and was constructed and equilibrated with the CHARMM-GUI protocols (31–33). The production runs for the two bilayers (917 ns for Ld and 1345 ns for Lo) were performed at constant temperature of 293 K and constant pressure of 1 atm using the same simulation

TABLE 1 Compositions and structural parameters of simulated Ld and Lo phases at 20°C

Mixture	χ_{DSPC}	χ_{DOPC}	χ_{CHOL}	A_L (Å ²)	D_B (Å)
Ld	0.09	0.79	0.12	59.9	40.0
Lo	0.60	0.11	0.29	40.0	52.8

parameters as in (34). Number and charge density profiles of each lipid and water atom in the systems were calculated from the last 480 ns of the centered bilayer trajectories with the Density Profile tool in VMD (35). The dipole potential profile for each system was calculated from the charge density profile following the approach in (36). The protocol is based on Poisson's equation and utilizes a double integral calculation, setting the potential to be the same on the opposite sides of the simulation box, which is a good approximation for symmetric bilayers (37). Table 1 lists the compositions, average area per lipid A_L , and bilayer thickness D_B of the simulated Lo and Ld phase bilayers.

Synthetic cryo-EM images

Synthetic cryo-EM images of lipid vesicles were generated following a previously described method that we summarize here (21). In phase-contrast imaging mode, the observed signal is related to variation in the electrostatic potential Φ within the lipid bilayer, which depends on lipid composition and packing density and is therefore different for Lo and Ld phases. We approximated $\Phi(w)$ (where w denotes position along the direction normal to the plane of the bilayer) as a sum of contributions from individual lipid and water atoms with an additional contribution from the membrane dipole potential Φ_d , i.e.,

$$\Phi(w) = \sum_i V_i \rho_i(w) + \Phi_d(w). \quad (\text{Equation 1})$$

In Eq. 1, the sum is over individual lipid atoms, $\rho_i(w)$ is the atom's time-averaged atomic number density profile (units of Å⁻³) calculated from the MD simulation trajectory, and V_i is the spatially integrated, shielded Coulomb potential for an isolated neutral atom ($V_i = 25, 130, 108, 97$, and 267 V Å^3 for H, C, N, O, and P, respectively). The phase shift experienced by an electron wave passing through the bilayer is given by

$$g(w) = \sigma_e \Phi(w), \quad (\text{Equation 2})$$

where σ_e accounts for the dependence of the electron phase on the projected potential and is equal to $0.65 \text{ mrad V}^{-1} \text{ Å}^{-1}$ for 300 keV electrons (38). Following Wang et al. (39), we refer to $g(w)$ as the electron "scattering profile" of the flat simulated bilayer; as Φ has units of V, the scattering profile has units of mrad Å⁻¹. A cryo-EM image of a liposome corresponds to a projection of the vesicle's spherical scattering profile, $\gamma(r, \theta, \phi)$, onto a plane, followed by convolution with a contrast transfer function (CTF). For a compositionally uniform and spherically symmetric liposome of radius R , γ has no angular dependence and can be approximated with the flat bilayer scattering profile $g(w)$ using the coordinate transformation $r = w + R$, such that $\gamma(r, \theta, \phi) = \gamma(r) = g(r - R)$. In this case, the projected density $\Gamma(r)$ is easily obtained by the Abel transform:

$$\Gamma(r) = 2 \int_r^\infty \frac{g(a - R)}{\sqrt{1 - (r/a)^2}} da. \quad (\text{Equation 3})$$

Equation 3 is not straightforward to evaluate for a phase-separated liposome where γ may have a complicated angular dependence. Instead, we used a Monte Carlo technique to approximate $\Gamma(r)$. In brief, a 3D representation of the vesicle was constructed as a set of discrete

points whose spatial density was proportional to $\Delta\gamma(r, \theta, \phi) = \gamma(r, \theta, \phi) - \gamma_s$, where γ_s is the electron phase shift caused by vitreous ice under given imaging conditions and is equal to 3.28 mrad Å^{-1} here. We used a geometrical model for the vesicle in which the Ld and Lo phases each comprised a single spherical cap centered on opposite poles of the vesicle. For this domain arrangement, the position of the domain boundary is specified by the polar angle $\theta_b = \cos^{-1}(1 - 2a_{Lo})$, where a_{Lo} is the area fraction of Lo phase. With the Lo and Ld domains centered at $\theta = 0$ and π , respectively, the azimuthally symmetric spherical scattering profile is

$$\gamma(r, \theta; R, \theta_b) = \begin{cases} g_{Lo}(r - R) & \text{for } 0 \leq \theta < \theta_b \\ g_{Ld}(r - R) & \text{for } \theta_b \leq \theta < \pi \end{cases} \quad (\text{Equation 4})$$

To generate images of phase-separated vesicles, the vesicle radius was drawn from a Schulz distribution centered at 25 nm with a width of 5 nm, consistent with experimentally observed vesicle-size distributions after extrusion through a 50-nm pore-size filter. Random points were then generated within the Lo and Ld domains delineated by θ_b and further subdivided radially into thin (0.2 Å width) concentric shells, with the shell radii r_j chosen to span the full thickness of the simulated flat Lo and Ld bilayers. Within the Lo and Ld shells, the point density was proportional to $\gamma(r_j, \theta)$ given by Eq. 4, with the constant of proportionality chosen to generate a total of $\approx 10^8$ points in the vesicle. The full set of points thus generated was then randomly rotated in 3D, projected onto the xz plane, and binned into 2.5-Å square pixels to create the image $\Gamma(r)$ where the intensity at each pixel was proportional to the density of projected points. At this step, the fraction of signal in each pixel originating from Lo and Ld domains was recorded and used to generate ground-truth maps of the bilayer phase state in the projection image. These maps were subsequently used to assess the accuracy of machine-learning analyses as described in "machine-learning classification of bilayer phase state" and "quantifying the accuracy of ML analysis."

To mimic experimental images, $\Gamma(r)$ must be convolved with a CTF $c(s)$ and associated phase perturbation factor $\chi(s)$:

$$c(s) = [\sin \chi(s) - Q \cos \chi(s)] \exp(-B|s|^2), \quad (\text{Equation 5})$$

$$\chi(s) \approx -\pi\lambda\Delta Z|s|^2 + \frac{\pi}{2}C_s\lambda^3|s|^4. \quad (\text{Equation 6})$$

In Eqs. 5 and 6, s is the spatial frequency, B is the amplitude decay factor, Q is the dimensionless amplitude contrast factor, λ is the electron wavelength, ΔZ is the defocus length (defined such that positive values indicate underfocus), and C_s is the spherical aberration coefficient. Fig. S2 plots the CTF for our simulated parameter set (Table 2) with and without the spherical aberration correction term in Eq. 6, demonstrating that the effects of spherical aberration are negligible for simulated imaging conditions, and thus justifying our choice to omit the second term in Eq. 6 in our calculations. We also note our use of a single electron wavelength, which carries the implicit assumption of zero chromatic aberration. Although chromatic aberration exists in any real electron microscope, the effects can be partially corrected by energy filtering. We previously showed that energy filtering

TABLE 2 Parameters used to generate simulated images (see materials and methods for details)

Electron		Contrast transfer function			
Energy	λ	ΔZ	Q	B	m
300 keV	1.97 pm	$2.5 \mu\text{m}$	0.075	300 Å^2	400

has minimal effects on bilayer intensity profiles that are used in the present analyses (27).

Accounting for the effects of the CTF, the reciprocal space image is calculated as

$$I(s) \propto 1 + mc(s)\mathcal{F}_s[\Gamma(r)](s), \quad (\text{Equation 7})$$

where $\mathcal{F}_s$ is the Fourier transform of the projected 2D phase-shift image $\Gamma(r)$ and m is a scale factor that adjusts the intensity contrast in the spatial domain (we define contrast as the difference between maximum and minimum values of bilayer intensity profiles described in “bilayer intensity profiles”). We chose a value of $m = 400$ that resulted in average contrasts of 85–110 a.u. depending on vesicle size and bilayer composition (see Fig. S3); these values are consistent with ranges observed in experimental images collected at 300 keV and 2.5 μm under-focus. The corresponding real-space image $I(r)$ is the inverse Fourier transform of $I(s)$,

$$I(r) = \mathcal{F}_r^{-1}[I(s)](r). \quad (\text{Equation 8})$$

Finally, zero-mean and frequency-independent (i.e., white) Gaussian noise with a standard deviation of 120 a.u. was added to simulated images, resulting in final image contrast-to-noise ratios of ≈ 0.7 – 0.9 that fall within the range of values observed in experimental images (≈ 0.7 – 1.05) collected under the simulated conditions. The complete sequence of steps for generating synthetic images described in this section is graphically summarized in Fig. S4, while Fig. S5 demonstrates how the profiles depend on vesicle size. For the analyses presented here, a total of 1000 vesicle images were generated for each of the two timeline endpoint samples to use as training data, and a total of 490 vesicle images were generated for each of the timeline compositions for validation data.

Bilayer intensity profiles

Within projection images, vesicles were identified and subdivided as previously described (21) to obtain spatially resolved intensity profiles (IPs) in the direction w normal to the bilayer, with $w = 0$ corresponding to the bilayer midplane. In brief, vesicle contours (i.e., the set of points approximating the midplane of the projected bilayer as defined by a relatively bright central peak) were first generated using a neural network-based algorithm (MEMNET) developed and kindly provided by Dr. Tristan Bepko (MIT, Cambridge, MA). While we retain the original MEMNET nomenclature, these algorithms (and others) are now part of the TARDIS software package (40). The MEMNET contour was resampled at arc length intervals of ≈ 5 nm, resulting in a polygonal representation of the 2D contour. For each polygon face, all pixels within a 5×20 -nm rectangular region of interest centered at the face were selected, and their intensities binned at 1-Å intervals in the long dimension (i.e., normal to the face) and subsequently averaged in the short dimension to produce a local IP.

Bilayer thickness measurement

A key measurement obtained from the IP is the spatially resolved bilayer thickness D_{TT} , details of which are provided in previous work (21). In brief, D_{TT} was calculated as the distance between the two minima in the IP. Two methods were used to locate the minima: 1) a “model-free” method, in which a local 5-point Gaussian smoothing was first performed, and the distance between the two absolute minimum intensity values on either side of $w = 0$ was calculated; and 2) a “model-fit” method that fits the profile as a sum of four Gaussians and a quadratic background, with the troughs corresponding to the two absolute minimum intensity values on either side of $w = 0$. The two methods generally agree to within 1 Å, and the D_{TT} values we report are the average of the two measurements.

Machine-learning classification of bilayer phase state

We tested various unsupervised and supervised learning methods for identifying the phase state of the bilayer, using the spatially resolved IPs as features for discriminating Lo and Ld phases. Each profile can be considered as a 201-dimensional vector of intensities for input to the ML algorithm (where noted, the dimensionality of the input vectors was first reduced using principal component analysis). For unsupervised learning, we used k -means and k -medoids clustering with the number of clusters (k) set to 2. For supervised learning, we used five methods as implemented in Mathematica v13.2: decision tree (DT), gradient boosted trees (GBT), random forest (RF), nearest neighbors (NN), and logistic regression (LR) (see section S1 in supporting methods for additional details of these models and their default parameter settings in Mathematica).

Models were trained on IPs from vesicles in the pure Lo and Ld data sets, with vesicle sizes drawn from a uniform distribution to minimize size-dependent bias. As shown in Fig. S3, IP contrast depends on both the bilayer phase state and the vesicle size. Therefore, to maximize the ability of ML models to predict the former, it is necessary to account for the effects of the latter. We investigated two ways of addressing this issue. 1) In the “IP pre-processing” procedure (described in detail in supporting methods section S2), intensity profiles were normalized prior to model training using an empirically determined scale factor that corrects for vesicle size dependence. 2) In the “size-training” procedure, vesicle size was included as a second feature (along with the un-normalized IP) at the stage of model training. In both cases, trained models were then used to predict the phase state of IPs from the validation data sets.

Quantifying the accuracy of ML analysis

We assessed the accuracy of each ML model by comparing the predicted phase state to the ground-truth phase state at the level of individual 5-nm segments. Because each projected bilayer segment may contain contributions from both phases, we define here the segment’s “ordered character” ($OC \in [0, 1]$) as the fraction of the segment’s intensity that originated from an Lo domain. For vesicles that contain a single domain of each phase (i.e., the geometric model used here), most segments are purely Ld ($OC = 0$) or Lo ($OC = 1$). Exceptions occur in the vicinity of the domain boundary, where intermediate OC values are typically observed. Importantly, all supervised ML methods yield an estimate of the probability that a given segment is Lo (i.e., a prediction of the segment’s OC value). We therefore used the residual (i.e., $OC - OC'$, where OC is the ground-truth ordered character and OC' is the prediction of the ML model) as one metric for accuracy.

Unsupervised ML methods result in a binary classification of the segment phase state (i.e., a given segment is either Ld or Lo). In some cases, it is also useful to simplify the phase classification from supervised ML methods as a strictly binary outcome. For this purpose, we used a threshold of $OC = 0.5$ (i.e., a segment is classified as Ld if $OC < 0.5$ and Lo otherwise). The accuracy of binary classification was assessed from confusion matrices as the number of correct predictions divided by the total number of predictions. One notable advantage of binary classification is the ability to estimate domain size by counting runs of identical Lo or Ld segments within a vesicle.

RESULTS AND DISCUSSION

Model system and generation of image data sets

The primary inputs for generating images of phase-separated vesicles are the electron-scattering profiles of Ld and Lo phases, which we calculated from MD simulations as described in materials and methods. We chose the

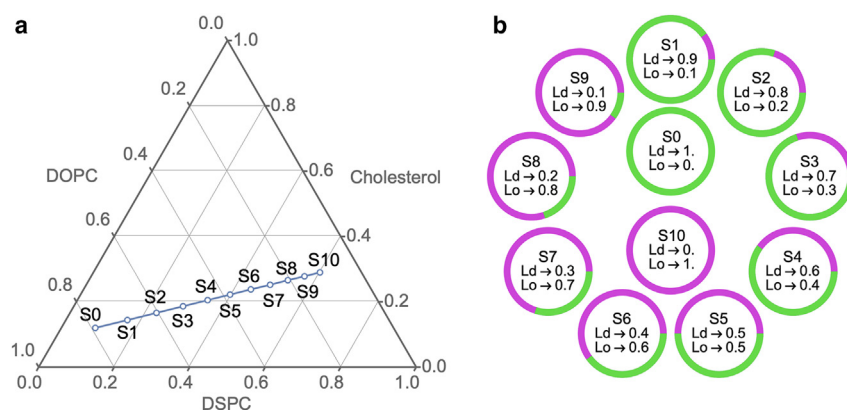


FIGURE 1 Lipid compositions of synthetic image data sets. (a) All compositions in this study lie on a thermodynamic tieline of the mixture DSPC/DOPC/Chol at 20°C (41). On this tieline, an Ld phase of fixed composition (S0) coexists with an Lo phase of fixed composition (S10), and only the relative amounts of Ld and Lo change. (b) The compositions S1–S9 represent a series of increasing Lo area percentage in increments of 10%. To see this figure in color, go online.

three-component mixture DSPC/DOPC/Chol as a model system because the room-temperature phase diagram has been determined experimentally at high compositional resolution (41,42). As shown in Fig. S1, the phase diagram contains a region of coexisting Ld and Lo phases that serve as a model for lipid rafts. We simulated the endpoints of a tieline that includes the composition DSPC/DOPC/Chol 39:39:22 mol%, in part because this composition has been extensively studied with confocal microscopy (43), fluorescence resonance energy transfer (44), X-ray scattering (24), and neutron scattering (3). The large line tension for this mixture (45) results in essentially complete coalescence of domains in giant unilamellar vesicles that can be easily visualized with fluorescence microscopy, such that each vesicle has a single domain of each phase.

Table 1 lists bilayer structural parameters of Ld and Lo phases obtained from the molecular simulations. Most important for cryo-EM are the substantial differences in bilayer thickness D_B and area per lipid A_L , the latter being inversely related to lipid packing such that the smaller A_L of Lo phase indicates greater lipid packing density compared to Ld phase. The large differences in bilayer thickness and lipid packing result in markedly different electron-scattering profiles for Ld and Lo bilayers, as shown in Fig. S4. We used these electron-scattering profiles to generate image data sets as a function of Lo area percentage ranging from 0% to 100% in increments of 10%, as shown in Fig. 1, with the workflow for image generation demonstrated schematically in Fig. S4. Throughout the text, we refer to these compositions as S0–S10. For each composition, the size of individual vesicles was randomly drawn from a distribution consistent with extrusion through a 50-nm pore-size filter and thus producing diameters ranging from ≈ 20 to 80 nm. Each vesicle object was given a different random 3D orientation prior to projection. Projection images were convolved with a CTF assuming an electron energy of 300 keV and using typical values for CTF parameters (Table 2). A defocus range of 2.0–3.0 μm was previously found to be optimal for determining bilayer thickness (27), which guided our choice to use a fixed defo-

cus value of 2.5 μm . Fig. 2 shows representative images from each simulated composition before and after the addition of white Gaussian noise. For most of the phase-separated vesicles (i.e., S1–S9), distinct domains of different thickness and intensity are visible in the noisy images.

Data sets of noisy images were then analyzed to obtain IPs around the vesicle circumference at 5-nm resolution as detailed in materials and methods. Fig. 3 compares the ground-truth profiles (obtained by convolving the Ld and Lo electron-scattering profiles with the CTF) to representative IPs obtained from synthetic images of pure Ld and Lo vesicles. The ground-truth IPs of Ld and Lo phases (Fig. 3 b) are characterized by significant differences in both the position and depths of the “troughs” on either side of the bilayer center. These features, which arise from differences in bilayer thickness and molecular packing density of Ld and Lo phases, are partially obscured by noise at 5-nm resolution but are precisely recovered when IPs from many segments are averaged (Fig. 3, c and d). As we now describe, the spatially resolved IPs constitute the raw data that we used to classify the bilayer phase state.

Phase classification based on bilayer thickness

As a baseline for comparing the performance of ML methods, we employed a simple classification scheme for phase state based on bilayer thickness measured from segment IPs as described in materials and methods. Fig. S6 a shows probability distributions of segment thicknesses D_{TT} for each data set. As expected, the distributions are approximately Gaussian for the single-phase Ld and Lo data sets S0 and S10, while data sets S1–S9 show non-Gaussian character owing to Ld + Lo coexistence within the vesicles. The mean D_{TT} was 29.7 Å for S0 and 35.8 Å for S10, closely matching the ground-truth values of 30.6 Å and 36.4 Å for Ld and Lo phases, respectively.

To classify the phase state of individual segments, we first fit the thickness distributions for the tieline endpoint data sets S0 and S10 to Gaussians (Fig. S6 a, upper left). Thickness distributions for phase-separated data sets S1–S9 were

Sharma et al.

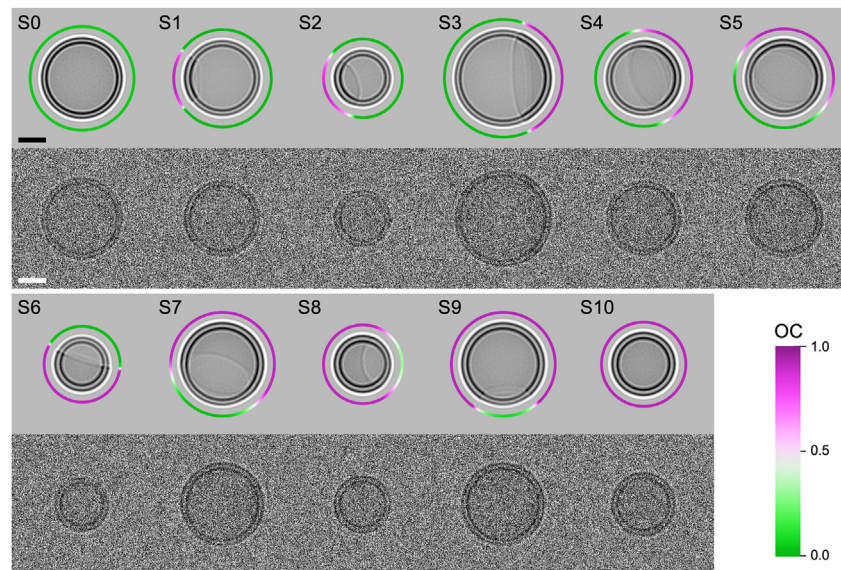


FIGURE 2 Representative vesicle images. A pair of images of a representative vesicle from each uniform (S0, S10) and phase-separated (S1–S9) data set is shown (see Fig. 1 for compositions). For each pair, the top and bottom images are before and after addition of Gaussian noise, respectively. The ring surrounding the vesicle in the noise-free images shows the angular dependence of the OC value in the projection, which quantifies the fraction of signal arising from the Lo domain. To see this figure in color, go online.

then modeled as a weighted sum of the endpoint Gaussians (i.e., with the mean and standard deviation values fixed), and the obtained weights were used to calculate the phase fractions for each data set, which resulted in excellent agreement with ground-truth phase fractions (Fig. S6 *b* and Table 3). For each data set, the best fit phase fraction f_{Lo} was then combined with the cumulative thickness distribution function (CDF) to determine a cutoff thickness D_{TT}^* satisfying the relationship $CDF(D_{TT}^*) = f_{Lo}$, which was then used to classify the phase state of each segment (i.e., a segment whose thickness was greater than or equal to

D_{TT}^* was classified as Lo). The accuracy of this thickness-based classification scheme, shown in Fig. S6 *c* and Table 4, is slightly greater than 80% for samples near the middle of the tieline where substantial fractions of both phases are present and increases for samples closer to the tieline endpoints.

Phase classification by unsupervised ML

We next used unsupervised learning (k -means and k -medoids) to group the segment IPs into two clusters

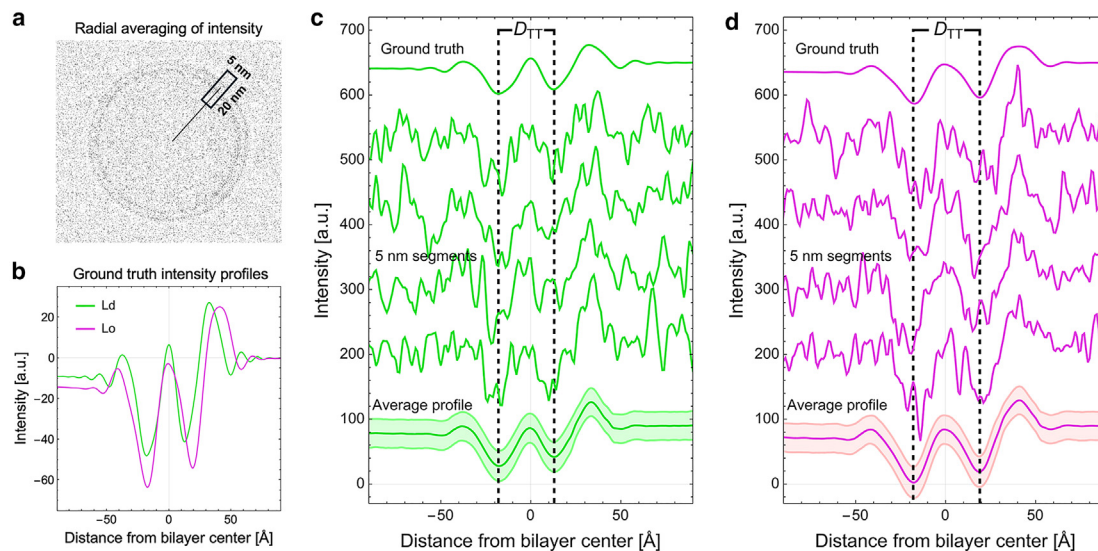


FIGURE 3 Segment intensity profiles. (a) Schematic illustration of the procedure for obtaining spatially resolved intensity profiles (IPs). IPs are generated by radial averaging of the intensity within a 5×20 nm sampling region centered at the projected vesicle contour. The sampling region is shown as a black rectangle superimposed on a projection image that has been faded for clarity. (b) Ground-truth IPs of Ld and Lo phases show significant differences in the position and depth of the troughs on either side of the bilayer center. (c) and (d) Comparison of Ld (c) and Lo (d) IPs: ground truth (top), four randomly chosen 5-nm segments (middle), and the average of all segments in all vesicles (bottom, with shaded area denoting ± 1 standard deviation). IPs in (c) and (d) are offset vertically for clarity. To see this figure in color, go online.

TABLE 3 Area percentage of Lo phase determined by different phase-classification methods

Sample	Ground truth	D_{TT}	Unsupervised ML		Supervised ML with size training				
			2-Means cluster	2-Medoids cluster	DT	GBT	RF	LR	NN
S0	0.0	0	2	4	4	1	2	1	1
S1	10.0	11	11	14	13	11	11	11	10
S2	20.0	20	20	24	22	21	21	21	19
S3	30.0	30	30	34	32	31	31	31	30
S4	40.0	40	39	43	40	40	40	40	38
S5	50.0	50	48	52	49	50	49	50	48
S6	60.0	59	57	60	58	60	58	60	57
S7	70.0	69	67	70	68	70	69	71	68
S8	80.0	78	76	79	76	79	78	80	77
S9	90.0	88	86	88	85	89	87	90	87
S10	100.0	98	96	97	94	99	97	99	97

(we note that for this analysis, segments from all 11 samples were pooled prior to clustering). As shown in Table 3, both clustering methods resulted in reasonable area percentage predictions for all compositions, deviating by <5% from ground-truth values. Both methods also achieved good accuracy (>90%) at the level of individual segment classification (Table 4). The accuracy was correlated with the position of the sample on the tieline, with the lowest accuracy consistently occurring near the middle of the tieline (i.e., sample S5). While this trend is similar to that found for thickness-based classification, unsupervised learning resulted in substantially improved overall prediction accuracy, especially near the middle of the tieline. To assess the impact of dimensionality reduction, we also performed clustering after first subjecting the IPs to a global principal component analysis. Fig. S7 reveals that segment accuracy exceeds 90% for all data sets even when only the first two principal components are used for clustering. We note that for data sets of the size used here, k -means and k -medoids clustering are computationally inexpensive and fast on a desktop computer even when the full 201-dimensional IP is used.

Phase classification by supervised ML

We next tested five supervised ML algorithms as described in materials and methods using models trained on segment IPs obtained from 1000 vesicles drawn from a uniform size distribution in each of the tieline endpoint samples. Table 3 compares the Lo area percentage predicted by these models to predictions from thickness-based and unsupervised methods, while Table 4 compares the overall accuracy of binary phase classification for the various data sets and methods at the level of individual segments. Like the thickness-based and unsupervised methods, accuracy is strongly correlated with the position of the sample on the tieline (Fig. 4 *a* and *b*). Comparing different methods, GBT and LR emerged as the most accurate overall, followed closely by NN and RF; DT was significantly less accurate. Fig. S8 shows an example of the confusion matrices that were used to calculate binary accuracy.

As described in materials and methods, we tested two procedures to correct for the influence of vesicle size on contrast, in which models were trained either on size-normalized segment IPs (“IP pre-processing”) or unnormalized IPs and vesicle radii (“size training”). For the

TABLE 4 Segment-level accuracy determined by different phase-classification methods

Sample	D_{TT}	Unsupervised ML		Supervised ML									
		2-Mean	2-Medoids	DT		GBT		RF		LR		NN	
				ST ^a	PP ^b	ST ^a	PP ^b	ST ^a	PP ^b	ST ^a	PP ^b	ST ^a	PP ^b
S0	100	98	96	96	96	99	99	98	99	99	99	99	99
S1	89	94	92	92	93	96	96	95	95	96	96	95	95
S2	84	92	91	91	92	95	95	94	94	95	95	94	94
S3	81	92	91	90	91	94	95	93	94	95	95	94	94
S4	82	91	91	90	91	94	94	93	94	94	94	93	94
S5	81	90	90	89	90	93	93	92	93	93	93	92	93
S6	82	90	91	90	90	94	94	92	94	94	94	92	94
S7	84	92	93	91	92	95	95	93	94	95	95	93	95
S8	86	92	92	91	92	95	95	93	94	95	95	93	95
S9	90	92	93	91	93	96	96	94	95	96	96	94	96
S10	98	96	97	94	97	99	99	97	99	99	99	97	99

^aSize training.

^bIP pre-processing.

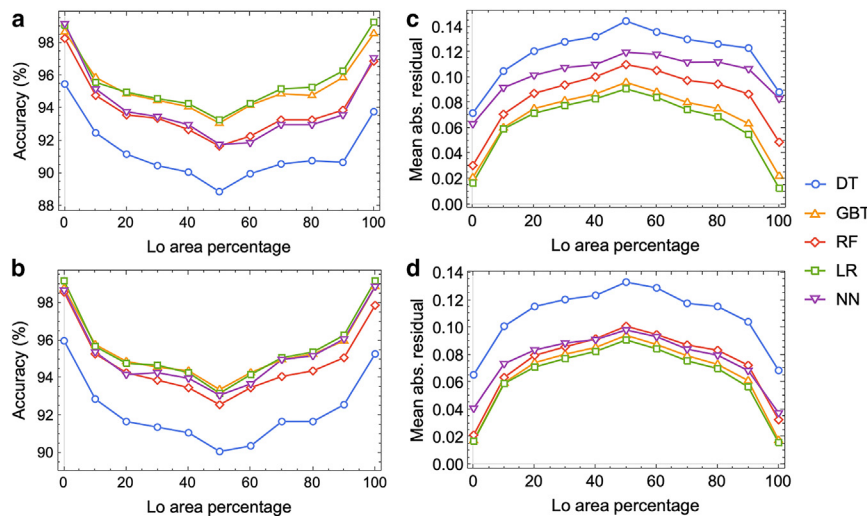


FIGURE 4 Accuracy of phase classification by supervised ML methods. (a and b) Percentage of correctly predicted segments vs. Lo area percentage using either size-training (a) or IP pre-processing (b) procedures. (c and d) Mean absolute residual of segment ordered character using either size-training (c) or IP pre-processing (d) procedures. Methods used: decision tree (DT), gradient boosted trees (GBT), random forest (RF), logistic regression (LR), and nearest neighbors (NN). To see this figure in color, go online.

best-performing algorithms (i.e., GBT and LR), the two procedures produced nearly identical results. Interestingly, IP pre-processing resulted in small but consistent gains in segment-level accuracy compared to size training for the NN, RF, and DT algorithms, as shown in Table 4 and Fig. 4.

Because supervised methods yield probability-based predictions (i.e., the probability that a given segment is Lo phase), we can further assess the ability of these models to recognize domain boundaries that are inherently weighted superpositions of Lo and Ld IPs. For each segment, we first calculated the fraction of total intensity originating from the Lo phase, which we termed the segment's OC . The residual $OC - OC'$, where OC is the ground truth value and OC' is the probability determined by a supervised model, is a natural alternative metric for quantifying accuracy, with smaller values indicating greater

accuracy. The mean absolute residual (i.e., averaged over all segments in a data set) is plotted vs. Lo area percentage in Fig. 4 c and d for the different supervised methods, revealing similar overall trends compared to the binary phase classification. The accuracy of boundary classification is discussed in the next section.

Visualizing phase-classification accuracy

It is instructive to visualize the accuracy of phase classification at the segment level within individual vesicles. Fig. 5 presents ring plots of representative vesicles in which ground-truth OC values of individual segments are found in the innermost ring and OC' values predicted by the various supervised models are arranged concentrically. The green-pink color scale facilitates the identification of misclassified

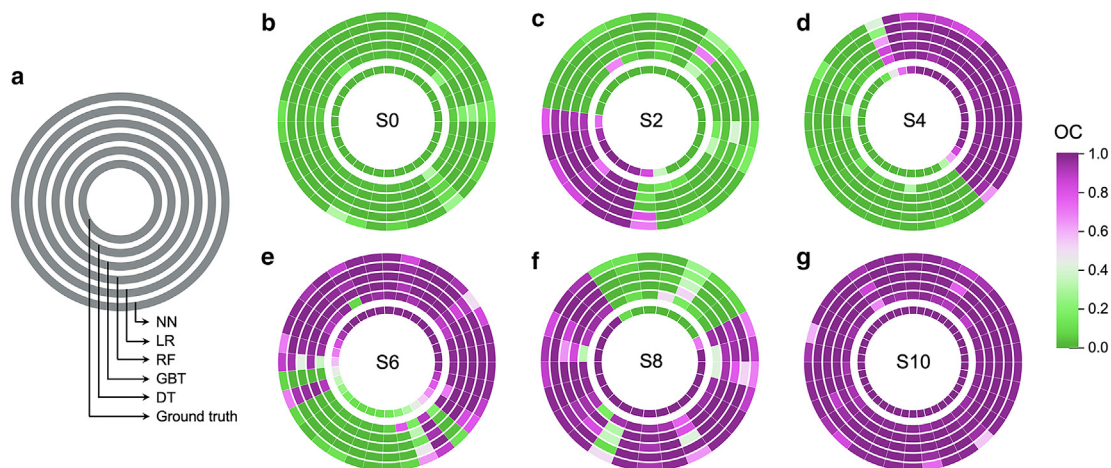


FIGURE 5 Visualizing the accuracy of phase classification for individual vesicles. (a) The inner ring gives the ground-truth OC for individual 5-nm segments of a vesicle, with model predictions OC' arranged concentrically as indicated. (b–g) Ring plots of representative vesicles for different data sets as indicated. Methods used are decision tree (DT), gradient boosted trees (GBT), random forest (RF), logistic regression (LR), and nearest neighbors (NN), in each case with the size-training procedure. To see this figure in color, go online.

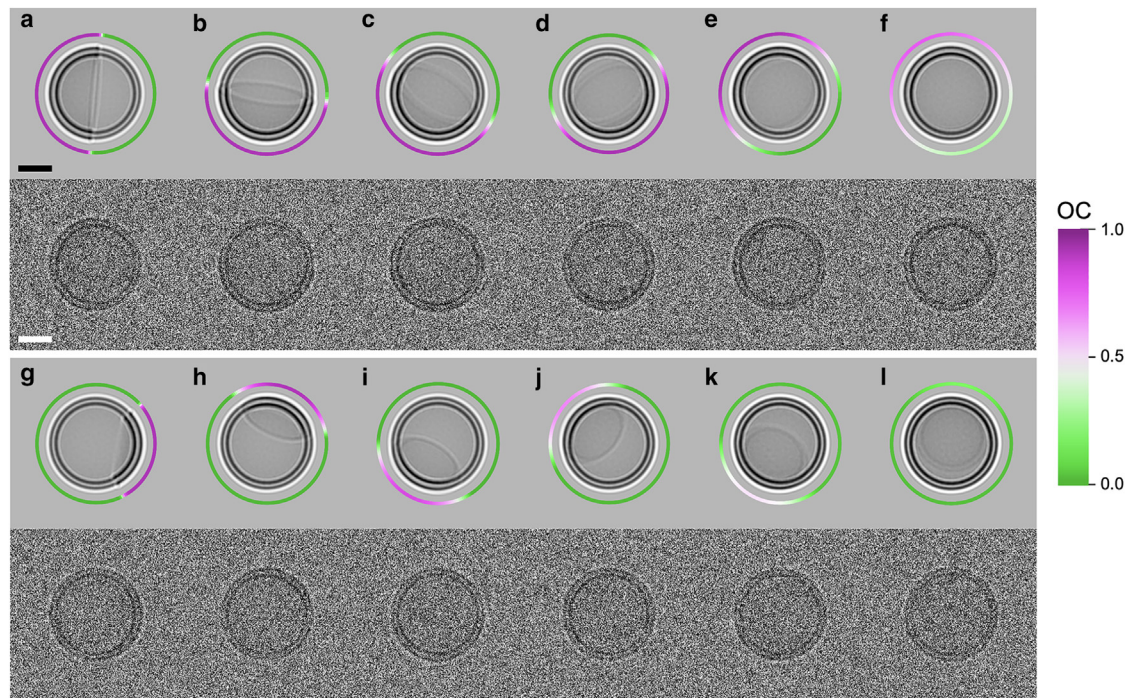


FIGURE 6 Selected images demonstrating the influence of vesicle orientation on apparent phase behavior. Shown are pairs of images for several orientations of ≈ 50 -nm-diameter vesicles from data sets S5 (a–f) and S2 (g–l). In each pair, the upper and lower images are before and after the addition of Gaussian noise, respectively. The ring surrounding the vesicle in the noise-free images plots the ground-truth OC value along the projected circumference as indicated by the color scale. Scale bars, 20 nm. To see this figure in color, go online.

segments (i.e., an Lo segment identified as Ld or vice versa), in this case when a binary cutoff of $OC^* = 0.5$ was used, as described in [materials and methods](#). A few such misclassified segments are found in each of the phase-separated vesicles shown in [Fig. 5](#). Unsurprisingly, the probability of misclassification is greater near domain boundaries where both phases contribute to the projected signal. Notably, the representative examples of pure Ld and Lo vesicles shown in [Fig. 5 b](#) and [g](#) contain no misclassifications, consistent with the absence of confounding boundary segments in these data sets, although misclassifications in the bulk do occasionally occur in both uniform and phase-separated vesicles and can be seen in [Fig. 5, c, e, and f](#).

In light of the previous discussion, the trends in accuracy shown in [Fig. 4](#) can be explained by the difficulty of classifying boundary segments. The lowest overall accuracy occurs in system S5, which has equal area fractions of the two phases and thus the largest number of boundary segments. Indeed, every projection image for this sample—regardless of the vesicle’s 3D orientation—is assured to have boundary segments, as demonstrated in [Fig. 6, a–f](#). However, as the vesicle composition moves toward the ends of the tieline in either direction, the size of one domain increases at the expense of the other, and the total domain boundary length in the projected images decreases. Consequently, there is a non-zero probability of observing only the majority phase at the edges of a projected vesicle, as demonstrated in [Fig. 6 l](#) for sample S2. Such projections

are relatively common for compositions near the ends of the tieline ([Fig. S9](#)) and, as a result, the accuracy of phase identification is higher than for samples near the middle of the tieline.

Further insight can be gained by visualizing segment classification accuracy for an entire data set. For this purpose, we generated matrix plots in which rows correspond to individual vesicles (arranged from smallest to largest diameter, top to bottom) and columns correspond to individual segments (arranged such that the middle of the minority phase domain is centered within the row), as shown in [Fig. 7](#) for system S4. Several interesting aspects of the data sets become apparent in these plots. For example, the ground-truth OC values plotted in [Fig. 7 b](#) reveal how the fraction of boundary segments (i.e., those with lighter shades) varies widely between vesicles due to differences in their random 3D orientations prior to projection (cf. [Fig. 6](#)). In a few cases, the vesicle was oriented such that the domain boundary was nearly parallel to the projection plane, resulting in every segment containing intensity contributions from both phases. [Fig. 7 a](#) plots the corresponding OC' values predicted by GBT, revealing a strong tendency for the model to predict values close to 0 or 1 and eschew intermediate values, even for boundary segments whose ground-truth OC values are in fact intermediate. This tendency is emphasized in [Fig. 7 c](#), which plots the signed residual $OC - OC'$ and thus makes clear the outsized contribution of boundary segments to overall accuracy. The relatively poor

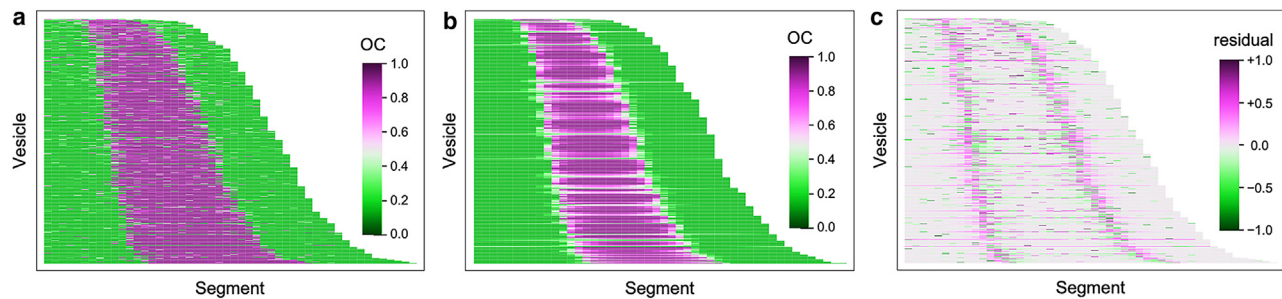


FIGURE 7 Matrix plots of ordered character. Each row represents one of 490 vesicles in data set S4, where the columns are individual 5-nm segments of the vesicle. Plotted are: (a) the OC' prediction from the GBT model with the size-training procedure; (b) the ground-truth OC value; and (c) the signed residual $OC - OC'$. To see this figure in color, go online.

performance in boundary identification is likely caused by the fact that the models were not trained to recognize boundary segments. Adding such segments to the training data set will almost certainly improve the predictions and may prove to be important for real-world data analysis, where the likelihood of encountering multiple smaller domains within vesicles is greater.

Average IPs of coexisting phases

Among the most important information obtained from cryo-EM analysis is the average IP of the bilayer. In principle, the bilayer's electron-scattering profile can be recovered from the IP by deconvolution (39), raising the possibility of using cryo-EM to complement well-established neutron- and X-ray-based methods for bilayer structure determination (46). A significant advantage of cryo-EM compared to traditional scattering methods is the ability to classify individual segments by their phase and thus separately determine the IPs of coexisting phases. Fig. 8 *a* compares the mean IPs for Ld and Lo phases (i.e., averaged over all segments classified as Ld or Lo) for samples S0–S10 using predictions from GBT. As expected, agreement with ground-truth profiles (Fig. 3) is better for the majority phase in each data set, although the overall agreement for both phases is reasonable in all data sets. Fig. 8 *b* shows that D_{TT} values calculated from the IPs of phase-separated samples S1–S9 are generally in excellent agreement with the values calculated from the pure Ld and Lo data sets (Fig. 8 *b*, dashed

lines), although deviations occur for the minority phase near tieline endpoints.

Isolating vesicle-size effects

It is well established that nanoscopic membrane curvature plays a vital role in many biological processes (47,48). Curvature is influenced by intrinsic factors such as lipid composition and the presence of integral membrane proteins as well as by external factors including cytoskeletal attachments and the binding of peripheral membrane proteins (49). Even in simple liposomes with nominally identical leaflet compositions, it is likely that substantial structural perturbations emerge in the presence of high membrane curvature. This has been observed in coarse-grained MD simulations of 20-nm-diameter vesicles, which showed changes in the headgroup and acyl-chain packing of inner and outer leaflets that depended on both lipid headgroup type and tail saturation (50). In this context, another major advantage of cryo-EM compared to neutron and X-ray scattering is the ability to isolate membranes of different average curvature within the sample and separately interrogate their structure.

As a proof-of-principle demonstration, we binned vesicles in each data set by size and calculated several properties for the different bins (we note that all vesicles in our simulated data sets have identical bilayer structure irrespective of their size). Fig. 9 shows the apparent area fraction of Ld and

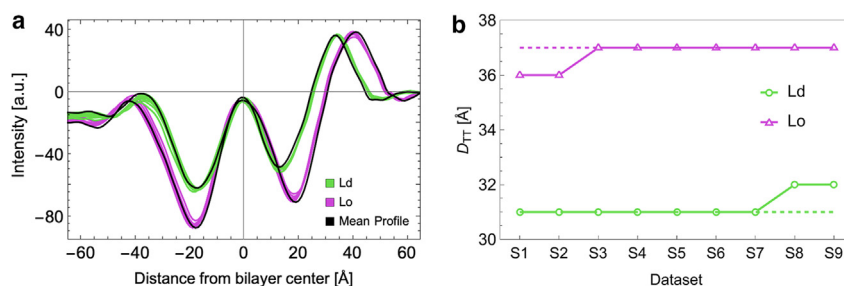


FIGURE 8 Average intensity profiles and D_{TT} values. (a) After classification by GBT with the size-training procedure, the mean intensity profile for all Lo and Ld segments is depicted in pink and green, respectively for each phase-separated data set (i.e., S1–S9), while the mean intensity profiles of pure Ld and Lo (data sets S0 and S10, respectively) are shown in black. (b) D_{TT} values evaluated from the mean intensity profiles shown in (a). Dashed pink and green lines represent the D_{TT} value of the pure Ld and Lo data sets. To see this figure in color, go online.

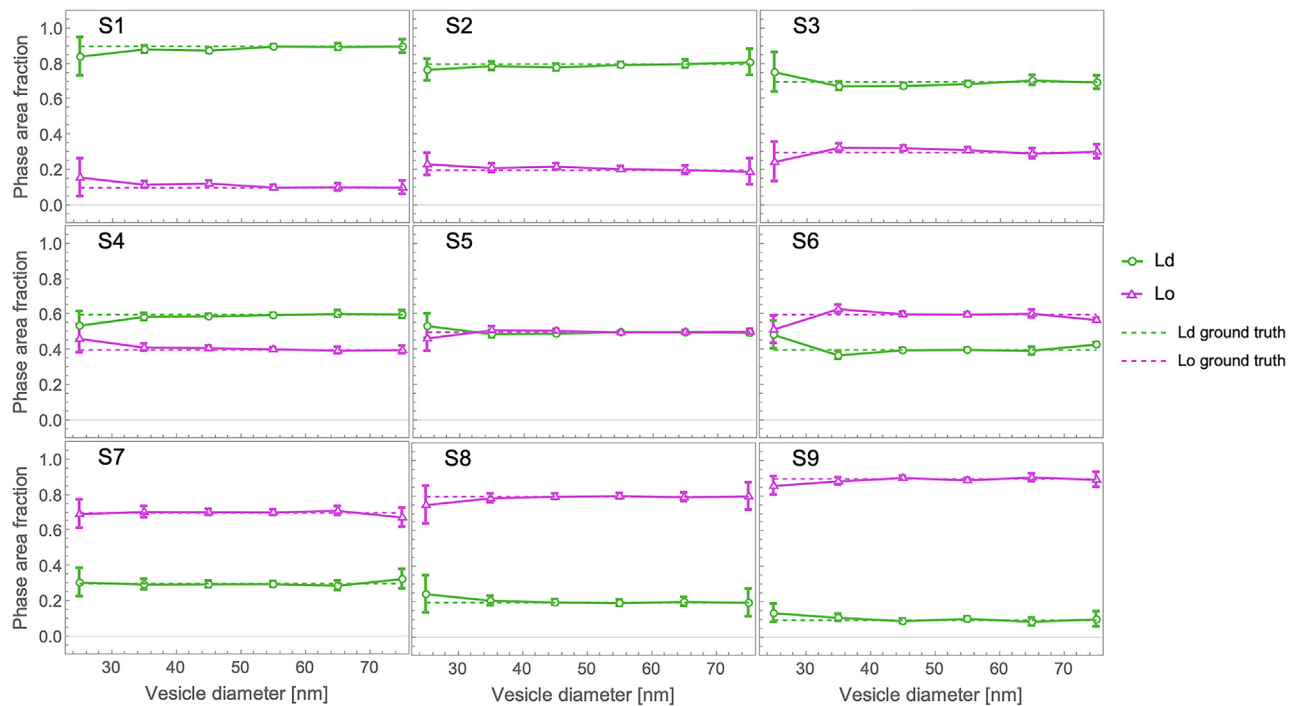


FIGURE 9 Influence of vesicle size on apparent area fraction of Ld and Lo phases. The area fraction of Ld (green) and Lo (pink) phases predicted by logistic regression with the size-training procedure is shown as a function of vesicle size for each of the phase-separated data sets. Error bars are 95% confidence intervals. To see this figure in color, go online.

Lo phases for systems S1–S9 calculated separately for each vesicle-size bin. We also attempted to determine the average linear extent and number of domains as a function of vesicle size. This was complicated by the non-negligible frequency of segment misclassifications far from domain boundaries (cf. Fig. 5), which have the effect of erroneously dividing a continuous domain into two or more discontinuous domains. To counteract this artifact, we implemented a correction that eliminated isolated segments (i.e., those whose phase classification was different than its two nearest neighbors) as illustrated in Fig. S10 *a*. With this correction, the average domain size calculated from images was similar to ground-truth values (Fig. 10). Reasonably good agreement between the average number of domains per vesicle calculated from images and the ground-truth value (i.e., 2 domains per vesicle) was also obtained from this analysis (Fig. S10, *b* and *c*).

Simulating noise in synthetic images

It is often noted that the signal-to-noise ratio (SNR) of cryo-EM is among the lowest of any imaging technique (11), with the noise power typically exceeding the signal power by more than an order of magnitude. When ground-truth data sets are used to assess the performance of classification methods, it is therefore reasonable to question whether the simulated noise characteristics accurately reflect real-world noise and how any simplifying assumptions might influence

the outcome. Baxter et al. have identified three independent sources of noise in cryo-EM projection images: 1) “structural noise” originating from variability in the local background structure (e.g., ice thickness) of individual vitrified particles as well as particle-to-particle conformational variability; 2) shot noise originating from the quantum nature of electron radiation; and 3) digitization noise arising from processes associated with image capture (e.g., CCD readout noise) (51). Through a careful analysis of paired projection images of ribosomes, the same authors were able to disentangle these noise sources and separately determine their magnitude, finding that shot noise is ≈ 10 -fold larger than structural noise, which in turn is over 10-fold larger than digitization noise (51). Clearly, shot noise is the dominant contributor to the low SNR of cryo-EM images.

In addition to having different magnitudes, the factors responsible for the independent noise sources act at different stages of the image-generation process. Because the factors that give rise to structural noise are present in the vitrified sample itself, structural noise is affected by the CTF and is thus frequency dependent. In contrast, shot noise and digitization noise originate at the detector and can be considered as white (i.e., frequency-independent) noise sources. When generating synthetic images, it follows that structural noise (typically modeled as a zero-mean Gaussian noise image) must be added to the projection image prior to convolution with the CTF, while the effects of shot noise and digitization noise can be combined in a single zero-mean Gaussian noise

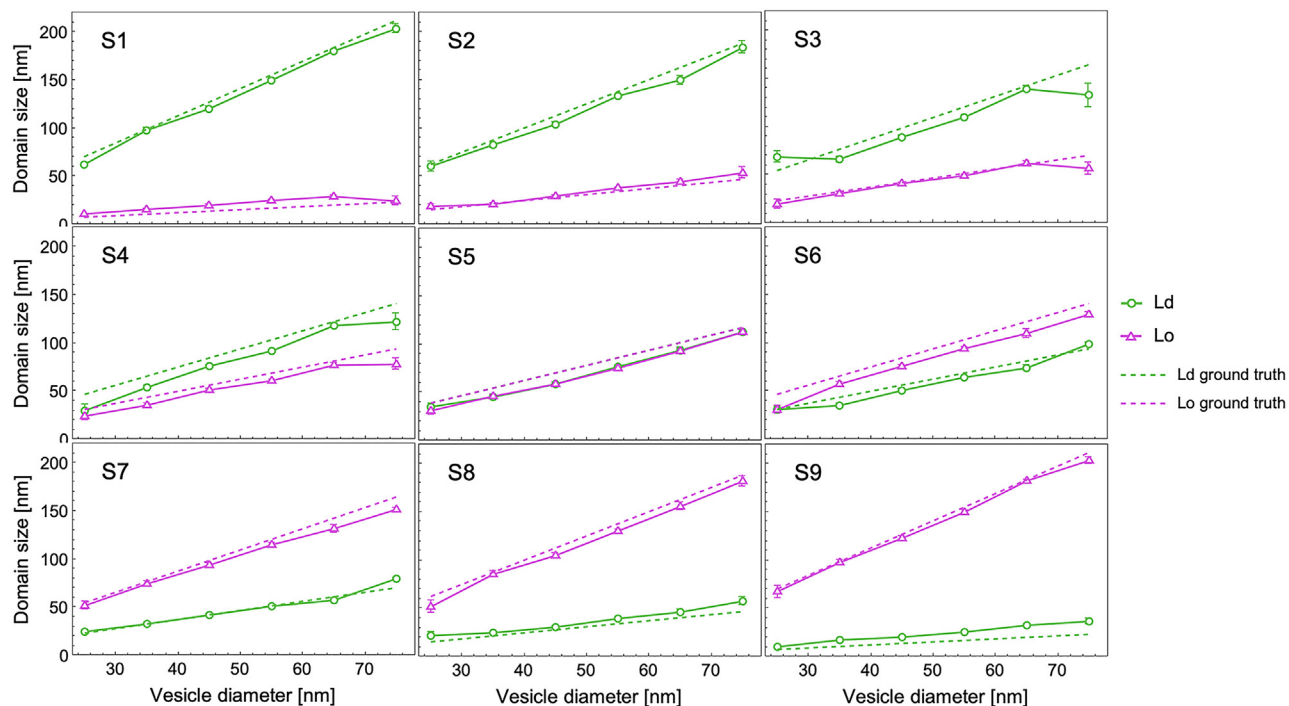


FIGURE 10 Influence of vesicle size on apparent domain size. The computed domain size across all nine systems is presented to examine the impact of vesicle sizes on area fractions. Logistic regression with the size-training procedure is employed as the classification technique. Error bars are 95% confidence intervals. To see this figure in color, go online.

image that is added to the CTF-corrupted particle image (S1). We note that, although shot noise is inherently governed by Poisson statistics, the use of a Gaussian distribution is justified by the relatively high electron counts per pixel (S2). The use of zero-mean Gaussian noise in our simulated liposome data sets is further validated in Fig. S11, which shows that histograms of pixel intensities in particle-free regions of a representative experimental image are well described by Gaussian distributions.

The full procedure for generating synthetic images of liposomes in the presence of both frequency-dependent and frequency-independent noise sources is outlined schematically in Fig. S4. As described in materials and methods, we omitted the frequency-dependent structural noise in our simulated data sets and included only frequency-independent zero-mean Gaussian noise with a standard deviation that was adjusted to mimic experimentally observed contrast-to-noise ratios of vitrified liposomes. It has previously been suggested that for synthetic cryo-EM images of ribosomes to be “realistic,” a frequency-dependent noise component with $\text{SNR} \approx 1$ should be included (S1). Although the magnitude of the structural noise for liposomes has not been reported to our knowledge, we tested whether including a structural noise component with a magnitude similar to that of vitrified ribosomes would affect the classification of segment phase state. Table S1 shows that including structural noise in both the training and validation data sets resulted in negligible differences in predic-

tion accuracy compared to the case where training and validation data both lacked structural noise. We are therefore confident that our reported prediction accuracies reflect outcomes that would be obtained in real-world analyses of similar samples, provided that the training and testing data are subject to the same sources of noise with comparable SNR.

Sampling the vesicle orientational distribution

Because projection images capture only a relatively small portion of the total vesicle surface, phase fractions calculated from individual vesicles can vary substantially depending on the orientation of the vesicle (Fig. S9). This can result in large uncertainties in the mean area fraction calculated from a sparse data set in which the orientational distribution may be undersampled. To estimate the uncertainty in reported phase fractions, we simulated experiments in which the mean phase fraction was calculated from a data set of N vesicles with random orientations; repeating the experiment a large number of times provides an adequate sampling from which to calculate the standard deviation of the mean and thus estimate the uncertainty inherent to a single data set. Fig. S12 demonstrates this procedure for vesicles with a ground-truth Lo fraction of 0.3 and varying N from 1 to 1000, in each case simulating the experiment 1000 times. Consistent with Fig. S9, the Lo fraction measured from individual vesicles is broadly distributed between

0 and 0.37. As the number of vesicles in the data set increases, the distribution of the mean phase fraction narrows considerably; for $N = 500$ (Fig. S12 *e*), the standard deviation is ≈ 0.005 . Fig. S13 plots the standard deviation as a function of N for data sets S1–S5 (corresponding to domain area fractions 0.1–0.5). For data sets of the size used in this study, the standard deviation is less than 0.005.

CONCLUSIONS

We report on a new application of ML to classify the phase state of lipid bilayers in cryo-EM images. To mimic the scenario of lipid rafts in eukaryotic plasma membranes, we analyzed synthetic images of phase-separated vesicles constructed from MD simulations of Ld and Lo phases. In previous work, we showed how membrane thickness can be measured in cryo-EM projection images (21). When these measurements were made locally, the distribution of obtained thickness values (i.e., unimodal vs. bimodal) was used to infer phase coexistence. However, there are limitations to this approach: 1) because the local thickness measurement is quite noisy, thickness distributions of the separate Lo and Ld phases are broad and overlapping, such that a large portion of the bilayer cannot be unambiguously classified as Ld or Lo by thickness alone (see Fig. S6 *c*); and 2) even had we attempted classification using thickness or any other local measurement, we had no means of assessing accuracy. The current work represents a significant advance by addressing each of these limitations: 1) rather than using scalar thickness measurements, the full information content of the image—i.e., the local intensity variation in the direction normal to the membrane—is used to determine the local phase state; and 2) by analyzing ground-truth simulated data sets where the actual phase state is known, we are able to assess and compare the accuracy of different classification methods in addition to testing the performance of the models in the presence of heterogeneities in both vesicle size and domain size.

Our most important finding is that ML methods can classify the phase state of the bilayer at accuracies exceeding 90% for the conditions that were simulated. Supervised methods such as LR and GBT showed the highest accuracy, while unsupervised methods such as k -means and k -medoids closely followed. Notably, unsupervised methods exhibited greater accuracy than some supervised methods, showcasing their potential for analyzing stand-alone experimental data sets and characterizing nanodomains in the absence of training data. In comparing results from supervised and unsupervised ML methods, we do not wish to imply that one is intrinsically superior to another; indeed, compared to classification based on segment thickness alone, each of the ML methods showed significantly improved performance. Instead, the choice between methods depends on the scientific questions being asked and the data that are available.

We observed that prediction accuracy decreased for samples positioned in the middle of a tieline where approximately equal fractions of each phase are present, owing to the presence of domain boundaries that are particularly prone to misidentification in a binary classification scheme. Extensive domain interface is a hallmark of nanoscopic phase separation, and the ability to correctly classify boundary segments is thus a key target for future development. It must also be recognized that additional challenges are encountered in vitro and in vivo that are not accounted for in our in silico data sets; these include variability in physical parameters such as vesicle composition and ice thickness, as well as instrumental parameters such as defocus length (26). Despite these limitations, the ability to directly image and decode nanoscopic membrane features will likely position cryo-EM at the leading edge of membrane biophysics in the years to come.

SUPPORTING MATERIAL

Supporting material can be found online at <https://doi.org/10.1016/j.bpj.2024.04.029>.

AUTHOR CONTRIBUTIONS

K.D.S., M.D., M.N.W., and F.A.H. designed the research. M.D. performed molecular simulations. F.A.H. generated the synthetic data sets. K.D.S. and F.A.H. analyzed the data sets. K.D.S., M.D., M.N.W., and F.A.H. wrote the manuscript.

ACKNOWLEDGMENTS

We are grateful to Dr. Tristan Bepler for providing the machine-learning algorithm, MEMNET, for automated contouring of vesicles. We also thank the anonymous reviewers for valuable comments and suggestions that improved the paper.

This work was supported by NSF grant CHE-2204126 (to F.A.H. and M.N.W.). M.N.W. acknowledges the William Wheless III Professorship.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Sezgin, E., I. Levental, ..., C. Eggeling. 2017. The mystery of membrane organization: composition, regulation and roles of lipid rafts. *Nat. Rev. Mol. Cell Biol.* 18:361–374.
- Eggeling, C., C. Ringemann, ..., S. W. Hell. 2009. Direct observation of the nanoscale dynamics of membrane lipids in a living cell. *Nature.* 457:1159–1162.
- Heberle, F. A., R. S. Petruzielo, ..., J. Katsaras. 2013. Bilayer thickness mismatch controls domain size in model membranes. *J. Am. Chem. Soc.* 135:6853–6859.
- Kollmitzer, B., P. Heftberger, ..., G. Pabst. 2015. Bending rigidities and interdomain forces in membranes with coexisting lipid domains. *Biophys. J.* 108:2833–2842.

5. Das, C., K. H. Sheikh, ..., S. D. Connell. 2010. Nanoscale mechanical probing of supported lipid bilayers with atomic force microscopy. *Phys. Rev.* 82:041920.
6. Levental, I., and E. Lyman. 2023. Regulation of membrane protein structure and function by their lipid nano-environment. *Nat. Rev. Mol. Cell Biol.* 24:107–122.
7. Lorent, J. H., K. R. Levental, ..., I. Levental. 2020. Plasma membranes are asymmetric in lipid unsaturation, packing and protein shape. *Nat. Chem. Biol.* 16:644–652.
8. Levental, I., K. R. Levental, and F. A. Heberle. 2020. Lipid rafts: controversies resolved, mysteries remain. *Trends Cell Biol.* 30:341–353.
9. Cornell, C. E., A. D. Skinkle, ..., S. L. Keller. 2018. Tuning length scales of small domains in cell-derived membranes and synthetic model membranes. *Biophys. J.* 115:690–701.
10. Pfab, J., N. M. Phan, and D. Si. 2021. DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proc. Natl. Acad. Sci. USA.* 118: e2017525118.
11. Bepler, T., K. Kelley, ..., B. Berger. 2020. Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nat. Commun.* 11:5208.
12. Tegunov, D., L. Xue, ..., J. Mahamid. 2021. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nat. Methods.* 18:186–193.
13. Ronneberger, O., P. Fischer, and T. Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18 Springer.
14. Sanchez-Garcia, R., J. Gomez-Blanco, ..., J. Vargas. 2021. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* 4:874.
15. Zhong, E. D., T. Bepler, ..., J. H. Davis. 2021. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods.* 18:176–185.
16. Yao, X., X. Fan, and N. Yan. 2020. Cryo-EM analysis of a membrane protein embedded in the liposome. *Proc. Natl. Acad. Sci. USA.* 117:18497–18503.
17. Baker, M. R., G. Fan, ..., I. I. Serysheva. 2021. Cryo-EM structure of type 1 IP3R channel in a lipid bilayer. *Commun. Biol.* 4:625.
18. Doyle, M. T., J. R. Jimah, ..., H. D. Bernstein. 2022. Cryo-EM structures reveal multiple stages of bacterial outer membrane protein folding. *Cell.* 185:1143–1156.e13.
19. Tao, X., C. Zhao, and R. MacKinnon. 2023. Membrane protein isolation and structure determination in cell-derived membrane vesicles. *Proc. Natl. Acad. Sci. USA.* 120, e2302325120.
20. Biou, V. 2023. Lipid-membrane protein interaction visualised by cryo-EM: a review. *Biochim. Biophys. Acta Biomembr.* 1865, 184068.
21. Heberle, F. A., M. Doktorova, ..., I. Levental. 2020. Direct label-free imaging of nanodomains in biomimetic and biological membranes by cryogenic electron microscopy. *Proc. Natl. Acad. Sci. USA.* 117:19943–19952.
22. Ward, A. E., D. Sokovikova, ..., L. K. Tamm. 2023. Serinc5 restricts HIV membrane fusion by altering lipid order and heterogeneity in the viral membrane. *ACS Infect. Dis.* 9:773–784.
23. Cornell, C. E., A. Mileant, ..., S. L. Keller. 2020. Direct imaging of liquid domains in membranes by cryo-electron tomography. *Proc. Natl. Acad. Sci. USA.* 117:19713–19719.
24. Heftberger, P., B. Kollmitzer, ..., G. Pabst. 2015. In situ determination of structure and fluctuations of coexisting fluid membrane domains. *Biophys. J.* 108:854–862.
25. Bleecker, J. V., P. A. Cox, ..., S. L. Keller. 2016. Thickness mismatch of coexisting liquid phases in noncanonical lipid bilayers. *J. Phys. Chem. B.* 120:2761–2770.
26. Sharma, K. D., F. A. Heberle, and M. N. Waxham. 2023. Visualizing lipid membrane structure with cryo-EM: past, present, and future. *Emerg. Top. Life Sci.* 7:55–65.
27. Heberle, F. A., D. Welsch, ..., M. N. Waxham. 2023. Optimization of cryo-electron microscopy for quantitative analysis of lipid bilayers. *Biophys. Rep.* 3, 100090.
28. Phillips, J. C., R. Braun, ..., K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.
29. Klauda, J. B., R. M. Venable, ..., R. W. Pastor. 2010. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B.* 114:7830–7843.
30. Klauda, J. B., V. Monje, ..., W. Im. 2012. Improving the CHARMM force field for polyunsaturated fatty acid chains. *J. Phys. Chem. B.* 116:9424–9431.
31. Jo, S., T. Kim, ..., W. Im. 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 29:1859–1865.
32. Jo, S., J. B. Lim, ..., W. Im. 2009. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* 97:50–58.
33. Lee, J., X. Cheng, ..., W. Im. 2016. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Biophys. J.* 110, 641a.
34. Doktorova, M., M. V. LeVine, ..., H. Weinstein. 2019. A new computational method for membrane compressibility: Bilayer mechanical thickness revisited. *Biophys. J.* 116:487–502.
35. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33.
36. Sachs, J. N., P. S. Crozier, and T. B. Woolf. 2004. Atomistic simulations of biologically realistic transmembrane potential gradients. *J. Chem. Phys.* 121:10847–10851.
37. Gurtovenko, A. A., and I. Vattulainen. 2009. Calculation of the electrostatic potential of lipid bilayers from molecular dynamics simulations: Methodological issues. *J. Chem. Phys.* 130:215107.
38. Kirkland, E. J. 1998. *Advanced Computing in Electron Microscopy*. Springer.
39. Wang, L., P. S. Bose, and F. J. Sigworth. 2006. Using cryo-EM to measure the dipole potential of a lipid membrane. *Proc. Natl. Acad. Sci. USA.* 103:18528–18533.
40. Kiewisz, R., G. Fabig, ..., T. Bepler. 2023. Automated Segmentation of 3D Cytoskeletal Filaments from Electron Micrographs with TARDIS. Oxford University Press US.
41. Zhao, J., J. Wu, ..., G. W. Feigenson. 2007. Phase studies of model biomembranes: complex behavior of DSPC/DOPC/cholesterol. *Biochim. Biophys. Acta.* 1768:2764–2776.
42. Heberle, F. A., J. Wu, ..., G. W. Feigenson. 2010. Comparison of three ternary lipid bilayer mixtures: FRET and ESR reveal nanodomains. *Biophys. J.* 99:3309–3318.
43. Konyakhina, T. M., S. L. Goh, ..., G. W. Feigenson. 2011. Control of a nanoscopic-to-macroscopic transition: modulated phases in four-component DSPC/DOPC/POPC/Chol giant unilamellar vesicles. *Biophys. J.* 101:L8–L10.
44. Enoki, T. A., F. A. Heberle, and G. W. Feigenson. 2018. FRET detects the size of nanodomains for coexisting liquid-disordered and liquid-ordered phases. *Biophys. J.* 114:1921–1935.
45. Usery, R. D., T. A. Enoki, ..., G. W. Feigenson. 2017. Line tension controls liquid-disordered+ liquid-ordered domain size transition in lipid bilayers. *Biophys. J.* 112:1431–1443.
46. Heberle, F. A., J. Pan, ..., J. Katsaras. 2012. Model-based approaches for the determination of lipid bilayer structure from small-angle neutron and X-ray scattering data. *Eur. Biophys. J.* 41:875–890.
47. McMahon, H. T., and E. Boucrot. 2015. Membrane curvature at a glance. *J. Cell Sci.* 128:1065–1070.
48. Jarsch, I. K., F. Daste, and J. L. Gallop. 2016. Membrane curvature in cell biology: An integration of molecular mechanisms. *J. Cell Biol.* 214:375–387.

49. Kozlov, M. M., and J. W. Taraska. 2023. Generation of nanoscopic membrane curvature for membrane trafficking. *Nat. Rev. Mol. Cell Biol.* 24:63–78.
50. Risselada, H. J., and S. J. Marrink. 2009. Curvature effects on lipid packing and dynamics in liposomes revealed by coarse grained molecular dynamics simulations. *Phys. Chem. Chem. Phys.* 11:2056–2067.
51. Baxter, W. T., R. A. Grassucci, ..., J. Frank. 2009. Determination of signal-to-noise ratios and spectral SNRs in cryo-EM low-dose imaging of molecules. *J. Struct. Biol.* 166:126–132.
52. Scheres, S. H. W., R. Núñez-Ramírez, ..., J. M. Carazo. 2007. Modeling experimental image formation for likelihood-based classification of electron microscopy data. *Structure.* 15:1167–1177.