



UNIVERSITÀ POLITECNICA DELLE MARCHE

DIPARTIMENTO DI SCIENZE DELLA VITA E DELL'AMBIENTE

**Corso di Laurea Magistrale
Biologia Molecolare e Applicata**

VALIDAZIONE DEL MODELLO ASDP PER L'ANALISI SAXS
DI MEMBRANE LIPIDICHE ASIMMETRICHE

VALIDATION OF THE ASDP MODEL FOR SAXS ANALYSIS
OF ASYMMETRIC LIPID MEMBRANES

Tesi di Laurea Magistrale:

Andrea TRAVAGLINI

Matricola 1120565

Relatore:

Prof. Francesco SPINOZZI

Sessione Autunnale

Anno Accademico 2025/2026

Contents

1	Introduction	4
1.1	Biological membranes, transverse asymmetry, and model systems	4
1.2	Small-angle scattering from bilayers	12
1.3	Chemical-group models and non-idealities of the signal	18
1.4	The inverse problem	23
1.5	Molecular dynamics as an external criterion	26
1.6	Aims of the thesis	31
2	Materials and methods	34
2.1	Materials and samples	34
2.2	Vesicle preparation	36
2.3	SAXS measurements	39
2.4	Data reduction	41
2.5	Simultaneous fit	43
2.6	The ASDP model	45
2.7	Data weighting	55
2.8	Constraints	57
2.9	Optimization	64

2.10	Design of the analysis	66
2.11	Identifiability	69
2.12	Spread across runs	71
2.13	Molecular dynamics	72
2.14	Closure test	77
2.15	The MDtoSDP program	80
3	Results	92
3.1	From the reference to the final configuration	92
3.2	The final configuration	98
3.3	Parameters determined by the data	106
3.4	Convergence of the simulations	110
3.5	Group volumes from the simulations	111
3.6	Simulated distributions and their representation	115
3.7	Comparison between fit and molecular dynamics	118
3.8	Closure test on synthetic curves	123
4	Conclusions	129
A	Parameters of the final configuration	137
B	Checks of the simulations	140
B.1	Convergence	140
B.2	Solution for the group volumes	140
B.3	Electron density and amplitudes	140
C	Closure test	142
C.1	Convergence	142

C.2 Noise levels and bounds	142
C.3 Cholesterol as one group	142
List of symbols	144

Chapter 1

Introduction

1.1 Biological membranes, transverse asymmetry, and model systems

Lipids contain a polar head group and apolar chains, and in water this dual character does not lead to a unique form of organization. Depending on molecular geometry, lipid concentration, and the composition of the medium, lipids may remain dispersed in solution, associate into micelles, or assemble into extended phases such as bilayers [1, 2]. Which form prevails is determined by the balance of free energy. The same lipid may therefore adopt different forms of organization when these conditions change. Under the conditions found in cells and in the vesicle suspensions considered here, this balance favors the bilayer, a film about 5 nm thick that can extend laterally over micrometers.

Aggregation is driven primarily by the response of water rather than by a direct attraction between the lipids themselves. An apolar chain exposed to water induces a more ordered arrangement of the surrounding solvent, with an

entropic cost that increases with the exposed surface area. Bringing the chains together reduces their contact with water and thereby recovers part of this entropy. Dispersion forces between the chains further stabilize the aggregate, whereas hydration and repulsive interactions between the head groups oppose close packing. The balance between these contributions determines the area occupied by each molecule at the interface [3].

Aggregation also depends on concentration. For amphiphiles that form micelles, below a characteristic concentration, the critical micelle concentration, molecules are present predominantly as free monomers. Above it, monomers coexist with aggregates, with molecules continuously exchanging between the two populations. Aggregation is therefore an equilibrium process rather than an irreversible step.

Which type of aggregate forms is instead strongly influenced by molecular shape. The ratio between the volume of the apolar region, the length of the chains, and the area occupied by the head group determines the preferred curvature of the interface. A molecule whose head group is wide relative to its chains tends to form micelles, one whose head group and chains occupy comparable cross-sectional areas favors bilayers, whereas one with a relatively small head group favors inverted phases [2, 4] (Figure 1.1). Phosphatidylcholine and sphingomyelin broadly belong to the second group and therefore readily form bilayers. In contrast, phosphatidylethanolamine, with its smaller head group, has a tendency toward inverted phases. Within a bilayer, the hydrocarbon chains may in turn be ordered or disordered. The transition between these states occurs at a temperature that depends mainly on chain length and degree of saturation, ranging from below 0 °C for some monounsaturated phos-

phatidylcholines to above 40 °C for saturated sphingomyelins [5].

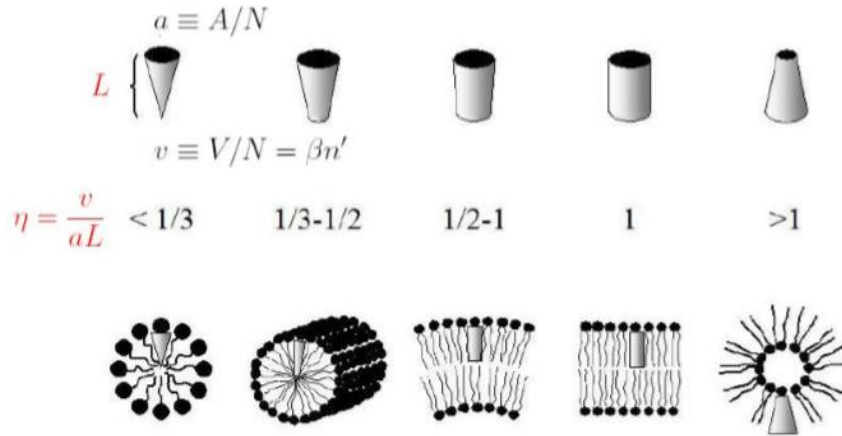


Figure 1.1: The packing parameter $\eta = v/(aL)$, with v the volume of the apolar region, a the area of the head group, and L the length of the chains, and the aggregate it favors: spherical micelles below $1/3$, cylindrical micelles between $1/3$ and $1/2$, curved bilayers and vesicles between $1/2$ and 1 , planar bilayers around 1 , and inverted phases above 1 . After [3].

The interest in membrane structure extends well beyond structural biology. Membrane proteins account for about 30% of the mammalian proteome and about 60% of known drug targets. Their activity depends on the surrounding lipid matrix, both through specific lipid-protein interactions and through bulk properties such as membrane thickness and lipid packing [6]. Alterations in lipid homeostasis are associated with conditions as diverse as type 2 diabetes, cancer, and neurodegeneration [7].

In pharmacology, membranes play a dual role: they are barriers that compounds must cross, but they can also serve as delivery vehicles, since liposomes form the basis of several formulations already in clinical use. In such systems, membrane composition influences both drug loading and release properties [8, 9], while the compositions of the two leaflets provide an additional degree of freedom for their design. Finally, in materials science, the selectivity

of biological membranes provides a benchmark for synthetic membranes used in desalination and water purification, including systems built around biologically derived water channels such as aquaporins [10].

The lipid matrix itself is far from uniform in composition. Eukaryotic cells devote about 5% of their genes to the synthesis of thousands of distinct lipids [1, 7]. This diversity appears at different levels of biological organization and concerns the polar head group, the backbone, and the acyl chains, which occupy different regions of the transverse membrane profile.

The two leaflets of a bilayer are not copies of one another. In mammalian plasma membranes, the outer leaflet is enriched in the choline-containing lipids phosphatidylcholine and sphingomyelin, whereas phosphatidylserine, phosphatidylethanolamine, and phosphatidylinositol are enriched in the cytosolic leaflet [1, 11] (Figure 1.2).

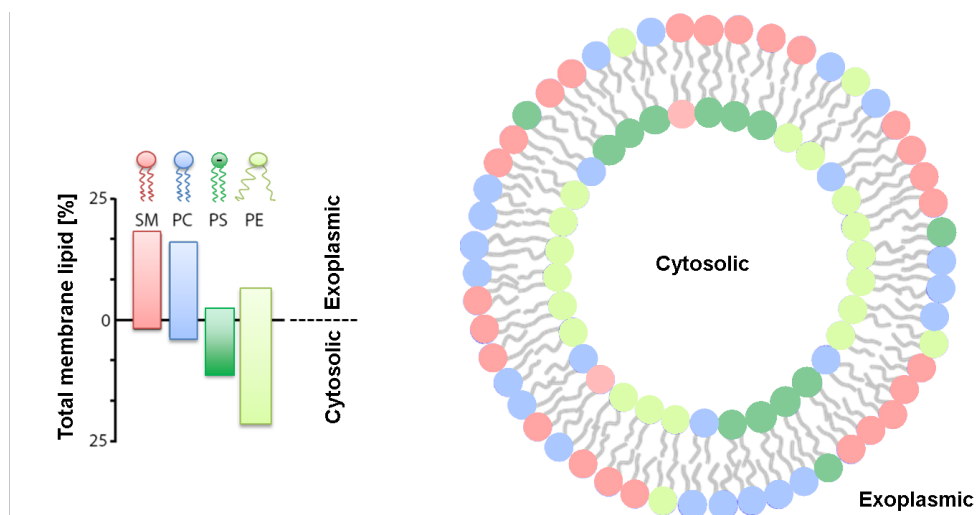


Figure 1.2: Left, the fractions of the total membrane lipid represented by sphingomyelin, phosphatidylcholine, phosphatidylserine, and phosphatidylethanolamine in the exoplasmic and cytosolic leaflets of the human red blood cell. Right, the same distribution represented on a vesicle. From [2], after the data of Verkleij and coworkers.

This arrangement is not an equilibrium state. It is maintained by trans-

porters that consume ATP to move lipids against their spontaneous distribution, whereas scramblases allow lipid head groups to cross the hydrophobic core [11]. Asymmetry is therefore a normal property of cell membranes rather than a laboratory construction, and the metabolic cost required to maintain it suggests that it has functional significance rather than being merely tolerated [11]. The transverse distribution of cholesterol, the most abundant individual membrane component at about 40 mol%, remains disputed, with experimental studies placing it preferentially in either leaflet [11].

Compositional asymmetry is accompanied by a corresponding physical asymmetry. The acyl chains of the outer leaflet are predominantly saturated, whereas those of the inner leaflet are predominantly polyunsaturated, with an average of 1.6 double bonds per lipid in the outer leaflet compared with 3.4 in the inner leaflet [12]. Because saturation affects how closely lipids pack, the two leaflets differ in packing, lateral diffusion, and the occurrence of transient defects at the head-group level, with the outer leaflet being the more ordered of the two [12]. The shapes of protein transmembrane domains reflect the same asymmetry, and this pattern is conserved across eukaryotes [12]. The two leaflets therefore differ not only in composition but also in their transverse structure, that is, in the positions and widths of the different chemical groups along the membrane normal. Structural measurements can provide direct access to these quantities.

A cell membrane, however, is difficult to characterize directly in its native context because of its heterogeneous and variable composition. Measurements on intact biological vesicles may also contain contributions from material enclosed within the vesicle and from other associated cellular components. For

this reason, a common experimental strategy is to reconstitute membranes from a limited number of known lipid species, so that composition, size, and geometry can be controlled by the preparation [13, 14]. Such model systems trade the biological complexity of the cell for control over the variables of the experiment. Model membranes fall broadly into two families that are not interchangeable, since membranes suspended in solution and membranes deposited on a support are accessible to different experimental techniques [14] (Figure 1.3).

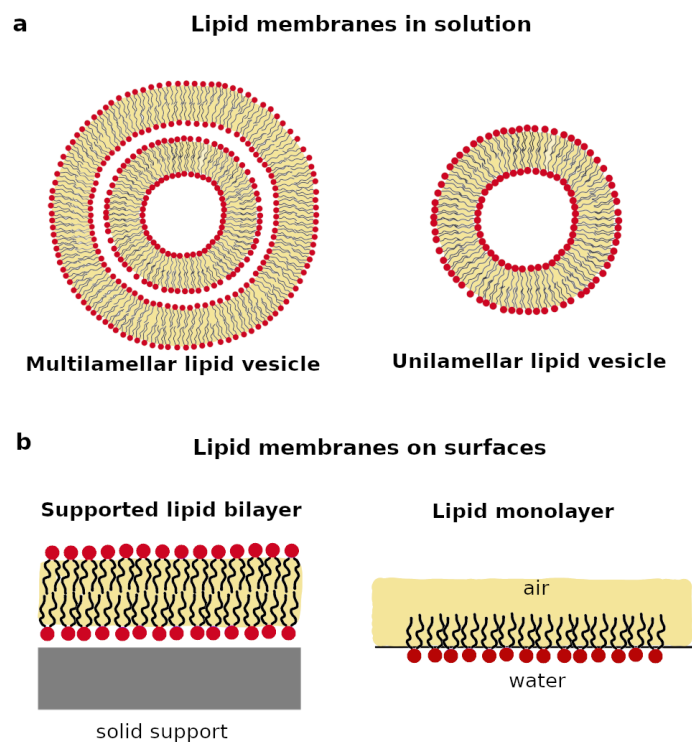


Figure 1.3: (a) Membranes in solution: a multilamellar and a unilamellar vesicle. (b) Membranes on surfaces: a bilayer on a solid support and a monolayer at the air-water interface. From [14].

Vesicles in suspension are commonly classified by size: small vesicles have radii below about 100 nm, large vesicles lie above this range, and giant vesicles exceed about 1 μm [14]. Giant vesicles can be observed individually under a microscope, making the response of a single membrane to mechanical, elec-

trical, or chemical perturbations directly accessible [15]. Large unilamellar vesicles, by contrast, are too small for such observations, since a diameter of about 200 nm lies close to the resolution limit of optical microscopy. This same size range, however, makes them well suited to small-angle X-ray and neutron scattering, SAXS and SANS respectively, or SAS in general.

The volume of suspension illuminated by the X-ray or neutron beam contains a very large number of vesicles in random orientations, so that the measured intensity represents an average over the entire population. The vesicles are nevertheless large enough that, with radii more than an order of magnitude greater than the bilayer thickness, the membrane can be regarded as locally flat. Under this approximation, its transverse structure can be analyzed independently of the overall vesicle geometry (Sec. 1.2).

Large unilamellar vesicles also provide a convenient experimental system for reproducing the compositional asymmetry of biological membranes. Starting from initially symmetric vesicles, an asymmetric bilayer can be prepared by cyclodextrin-mediated exchange, in which lipids in the outer leaflet are replaced by lipids supplied by a donor population while the inner leaflet remains largely unchanged. The phospholipid asymmetry produced in this way is sufficiently stable to persist for hours to days [2, 16].

Different model-membrane preparations provide access to different aspects of membrane structure. Giant vesicles can be observed individually by fluorescence and confocal microscopy, revealing their lateral organization and its response to external perturbations [15]. In supported bilayers, atomic force microscopy provides information on surface morphology and local thickness, whereas reflectometry probes the structure along the membrane normal [14].

For vesicles in suspension, corresponding structural information is obtained by scattering techniques: X-rays and neutrons probe length scales comparable with the bilayer thickness and thereby provide access to its transverse structure [14]. Lateral organization is not entirely inaccessible in suspension either. Small-angle neutron scattering, when combined with fluorescence resonance energy transfer, can constrain the size of nanoscopic membrane domains [17].

The present work focuses on the organization of the membrane along its normal, resolved separately for the two leaflets. Rather than representing the bilayer as a small number of uniform regions, the model describes its transverse structure in terms of the chemical groups that constitute the lipid molecules (Sec. 1.3). Each group contributes according to its scattering length density (SLD), which for X-rays is directly related to its electron density. Differences in SLD between neighboring groups provide the contrast needed to distinguish their contributions to the scattering signal. Structural changes between different conditions can therefore be localized, within the resolution allowed by the data and the model, to particular regions of the lipid molecules rather than being assigned only to the membrane as a whole.

The measurements analyzed in this thesis were obtained by small-angle X-ray scattering from symmetric and asymmetric large unilamellar vesicles. The work was carried out within the Molecular Biophysics group of the Università Politecnica delle Marche, which has long studied membrane systems by small-angle scattering and in which both the structural model used here and the program implementing it were developed (Sec. 1.3). Describing the bilayer at the level of individual chemical groups necessarily introduces a relatively large number of structural parameters. The central question of this thesis is therefore

how much of this detailed structure is actually determined by the experimental SAXS curves, and which parameters, or combinations of parameters, can be recovered reliably from them (Secs. 1.4 and 1.6).

1.2 Small-angle scattering from bilayers

1.2.1 Source and elastic scattering by electrons

In a small-angle scattering experiment on a dilute suspension, the useful signal is weak and is recorded in the vicinity of the much more intense transmitted beam. Reliable measurements therefore require both high photon flux and a well-collimated beam, so that sufficient intensity is collected without compromising access to the smallest scattering angles. These conditions are provided by synchrotron sources, whose flux greatly exceeds that of conventional laboratory sources and whose beam divergence can be kept sufficiently small [18]. For this reason, the measurements analyzed in this work were performed at a synchrotron beamline (Sec. 2.3).

When X-rays interact with the sample, their electric field drives the electrons into oscillation. Each electron then acts as a secondary source, emitting radiation with the same wavelength λ as the incident beam (Figure 1.4). In the elastic scattering considered here, no energy is transferred to the sample, and this process provides the dominant contribution to X-ray scattering [19]. The intensity recorded at each point of the detector results from the constructive and destructive interference of the waves scattered by all electrons within the illuminated volume.

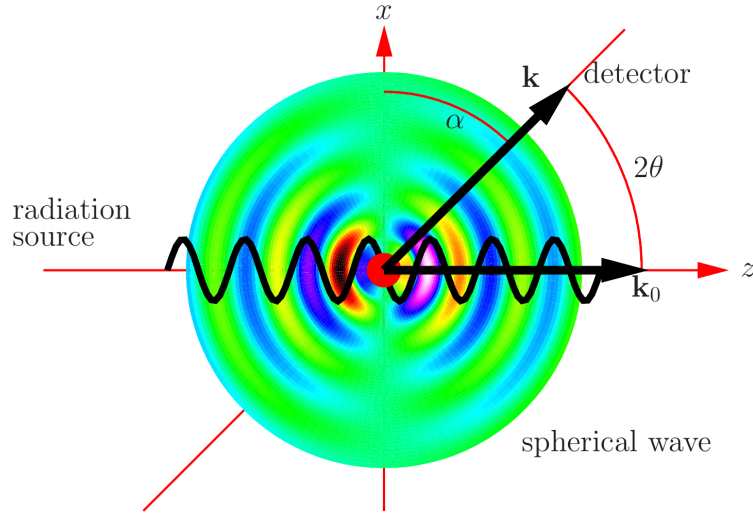


Figure 1.4: Elastic scattering of X-rays by an electron. The incident wave, with wave vector $\mathbf{k}_0$ along z , sets the electron oscillating, and the electron re-emits a spherical wave of the same wavelength. $\mathbf{k}$ is the wave vector of the wave scattered towards the detector, at an angle 2θ from the incident direction; α is the angle between $\mathbf{k}$ and the direction x of the incident electric field.

1.2.2 The scattering vector

The waves scattered by different electrons have the same wavelength as the incident wave, but they do not in general reach the detector with the same phase. Consider two scattering centers separated by a vector $\mathbf{r}$. Because the incident wave reaches the two centers at different positions, and because the scattered waves then travel toward the detector along the same observation direction, the two contributions acquire a relative phase that depends on both the incident and the scattered wave vectors.

This phase difference can be written as

$$\mathbf{q} \cdot \mathbf{r},$$

where

$$\mathbf{q} = \mathbf{k}_0 - \mathbf{k} \tag{1.1}$$

is the scattering vector. Here $\mathbf{k}_0$ and $\mathbf{k}$ are the wave vectors of the incident and scattered waves, respectively. Since the scattering is elastic, their moduli are equal,

$$|\mathbf{k}_0| = |\mathbf{k}| = \frac{2\pi}{\lambda}.$$

If the scattered beam forms an angle 2θ with the incident beam, the geometry of the two wave vectors gives

$$q = \frac{4\pi}{\lambda} \sin \theta. \quad (1.2)$$

The pattern recorded on the two-dimensional detector is the result of that interference. For scattering centers arranged on planes a distance d apart, the waves scattered by successive planes interfere constructively when their difference in path is a wavelength, which is Bragg's law [20],

$$\lambda = 2d \sin \theta. \quad (1.3)$$

The variable q has the dimensions of an inverse length, and combining (1.3) with (1.2) shows that a periodicity d in the sample contributes at $q = 2\pi/d$. Small angles therefore correspond to large distances, and the q -range between the smallest and the largest accessible value sets the range of distances on which a measurement carries information.

1.2.3 Amplitude, contrast and the measured cross section

For a continuous distribution of electrons, the total scattering amplitude is obtained by integrating over the electron density $\rho_e(\mathbf{r})$ within the illuminated

volume. Since the scattering amplitude of a single electron is proportional to the classical electron radius r_e , the corresponding scattering length density is $r_e \rho_e(\mathbf{r})$. The solvent contributes to the scattering as well. The relevant quantity is therefore not the electron density of the sample alone, but its excess over the solvent electron density ρ_0 . The excess scattering length density, or contrast, is thus $r_e[\rho_e(\mathbf{r}) - \rho_0]$.

After reduction to absolute units, the measured quantity is the macroscopic differential scattering cross section. It represents the scattering into the solid angle $d\Omega$ per unit incident flux and per unit irradiated volume (Figure 1.5) [19, 21],

$$\frac{d\Sigma}{d\Omega}(\mathbf{q}) = \frac{r_e^2}{V} \left\langle \left| \int_V [\rho_e(\mathbf{r}) - \rho_0] e^{i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r} \right|^2 \right\rangle, \quad (1.4)$$

where V is the irradiated volume and $\langle \cdot \rangle$ denotes the average over the configurations sampled during the measurement. The role of contrast is immediately apparent from (1.4): a region whose electron density is equal to that of the solvent does not contribute to the scattering, regardless of its shape.

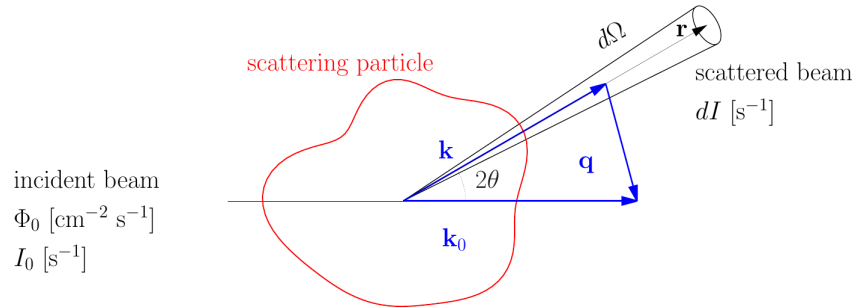


Figure 1.5: The geometry of the differential scattering cross section. The incident beam, of flux Φ_0 and intensity I_0 , travels along $\mathbf{k}_0$; the particle scatters into the solid angle $d\Omega$ around $\mathbf{k}$, at a distance $\mathbf{r}$, the intensity dI . The scattering vector is $\mathbf{q} = \mathbf{k}_0 - \mathbf{k}$.

For a suspension of particles, the integral in (1.4) can be decomposed into contributions from the individual particles. The scattering amplitude of a single

particle is

$$F_p(\mathbf{q}) = \int [\rho_e(\mathbf{r}) - \rho_0] e^{i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r}, \quad (1.5)$$

and for identical particles the cross section becomes

$$\frac{d\Sigma}{d\Omega}(q) = r_e^2 n_p \langle |F_p(\mathbf{q})|^2 \rangle_\omega S_M(q), \quad (1.6)$$

where n_p is the number of particles per unit volume, $\langle \cdot \rangle_\omega$ denotes the average over orientations, and $S_M(q)$ accounts for correlations between the positions of different particles. In a dilute suspension, the particles rotate freely and have no preferred orientation. The measured scattering is therefore isotropic and depends only on the modulus q of the scattering vector. If correlations between different particles are negligible, then $S_M(q) = 1$.

1.2.4 The orientational average and the reduction to one dimension

A bilayer extends laterally over distances much larger than its thickness. This separation of length scales makes it possible to reduce the scattering problem to the structure along the membrane normal. Consider a flat membrane represented as a disk of radius R , with an electron density that varies only along the normal. In cylindrical coordinates, the integral in (1.5) separates into an in-plane contribution and a Fourier transform along the normal. The in-plane integration gives a Bessel function,

$$F_p(\mathbf{q}) = 2\pi R^2 F(q \cos \psi) \frac{J_1(qR \sin \psi)}{qR \sin \psi}, \quad F(q) = \int [\rho_e(z) - \rho_0] e^{iqz} dz, \quad (1.7)$$

with ψ the angle between $\mathbf{q}$ and the normal. The function $F(q)$ is the form factor associated with the transverse structure of the bilayer and contains the information carried by the scattering curve about the electron-density profile along the normal.

When R is large compared with the bilayer thickness, the factor containing $J_1(qR \sin \psi)$ is strongly concentrated around $\sin \psi = 0$. Significant scattering therefore arises mainly when $\mathbf{q}$ is nearly parallel or antiparallel to the membrane normal. Carrying out the orientational average in this limit gives

$$\langle |F_p(\mathbf{q})|^2 \rangle_\omega = \frac{2\pi}{q^2} A_p |F(q)|^2, \quad (1.8)$$

with $A_p = \pi R^2$ the area of the particle in its plane, that is, the area of the whole membrane and not the area per lipid of the model (Sec. 2.6).

The result does not depend on the choice of a circular membrane. The same calculation for a square slab or an oblique prism gives the same expression, with A_p replaced by the corresponding in-plane area. The factor $2\pi/q^2$ is therefore a consequence of the orientational average of a flat, extended particle rather than of the particular disk geometry. Substituting (1.8) into (1.6), with $S_M(q) = 1$ for unilamellar vesicles, gives

$$\frac{d\Sigma}{d\Omega}(q) = r_e^2 n_p A_p \frac{2\pi}{q^2} |F(q)|^2. \quad (1.9)$$

The lateral geometry enters only through the product $n_p A_p$, which is the total bilayer area per unit volume of sample. The Luzzati thickness d_B is the volume of dry lipid per unit area of bilayer [22], so that $n_p A_p = \phi_L / d_B$, with ϕ_L the volume fraction of lipid, fixed by the concentration. The factor $2\pi/(q^2 d_B)$

is therefore the same factor that appears in the intensity of the model used here (Sec. 2.6).

A multilamellar particle contains a stack of parallel bilayers whose positions are correlated along the membrane normal. These correlations belong to the particle itself and are distinct from the correlations between different particles described by $S_M(q)$, which remains unity between independent stacks. Their orientational average introduces a lamellar structure factor multiplying $|F(q)|^2$ in (1.9) (Secs. 1.3 and 2.6). From (1.7) onward, the structural problem is therefore reduced to one dimension, and the remaining task is to choose a suitable representation of $\rho_e(z)$ (Sec. 1.3).

1.3 Chemical-group models and non-idealities of the signal

The model developed in this work is based on the locally flat membrane description introduced in Sec. 1.2. It therefore returns the transverse profile of a bilayer averaged over the vesicle population, while the overall geometry of the vesicle does not enter explicitly into that profile.

An alternative description treats a vesicle as a set of concentric spherical shells, with radial coordinates measured from the center of the particle and the scattering amplitude evaluated in spherical coordinates. This is not simply another representation of the same model. The two descriptions use different geometrical origins, the center of the vesicle in the spherical case and the bilayer midplane in the planar one, so that their structural parameters are not directly equivalent. A concentric-shell description is most appropriate when

vesicle curvature cannot be neglected, for example when the vesicle radius is not much larger than the membrane thickness. The flat description is adopted here because the vesicles considered in this work are large compared with the thickness of the membrane (Sec. 1.1).

The simplest models of the transverse profile divide the membrane into a small number of regions of uniform SLD, for example a core-shell representation of the vesicle or a set of slabs across a planar bilayer (Figure 1.6, right). Such models contain only a limited number of parameters, typically the SLD and thickness of each region. Their simplicity is an advantage when only coarse structural information is needed, but they cannot resolve how the different chemical groups are distributed within each region.

Chemical-group models are introduced precisely to recover this finer level of structural detail. Instead of assigning a single uniform SLD to a broad membrane region, they represent the bilayer as the superposition of contributions from distinct chemical groups. The price of this additional resolution is a larger number of model parameters, which makes the fit less well constrained by the scattering data alone. Additional a priori information is therefore required to restrict the physically meaningful parameter space. For example, the volumes assigned to the different chemical groups are commonly constrained using independent information from densitometry, molecular simulations, or complementary scattering measurements [23, 24].

An asymmetric membrane requires the two leaflets to be described independently, since they may differ in area per lipid, thickness, and the positions of their chemical groups. The resulting transverse profile is no longer symmetric with respect to the bilayer midplane. Models of this kind have already been

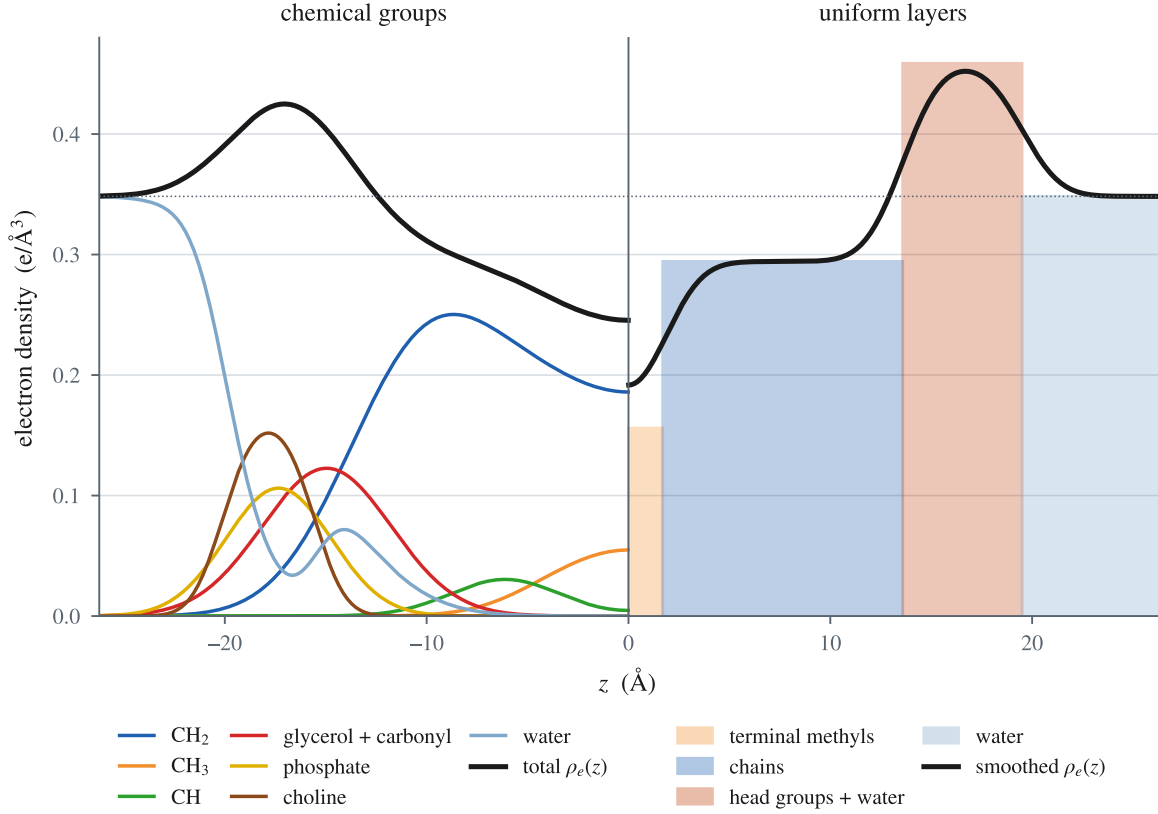


Figure 1.6: The electron density profile of POPC from the fit of this work (Sec. 2.6), with the midplane at $z = 0$. Left, the contributions $\phi_g(z)\rho_g$ of the chemical groups, volume fraction times electron density of the group, and their sum $\rho_e(z)$. Right, a schematic of the same lipid as uniform layers: the two terminal methyls, the rest of the chains (28 CH_2 and 2 CH), the head groups with the water that fills their layer, and the solvent. The height of each bar is the electron density of what the layer contains and its width holds the volume per lipid; the curve is the same profile with its steps smoothed ($\sigma = 1.5 \text{ \AA}$).

applied to asymmetric vesicles. A chemical-group model based on Gaussian distributions was developed for the joint analysis of X-ray and neutron scattering data from vesicles made asymmetric by lipid exchange, allowing the area per lipid and the hydrocarbon thickness to be determined separately for the two leaflets [25]. Earlier, small-angle neutron scattering data had been analyzed using a simpler four-slab representation to resolve the transverse structure of an asymmetric model membrane [26].

The model used in this work originates from the modified scattering density

profile (MSDP) model [27], which provides a chemical-group description of the transverse structure of a bilayer. In MSDP, the distribution of each chemical group along the membrane normal is represented by an erf slab: a region with an approximately uniform volume fraction in its central part and smoothly decaying edges described by error functions. Its position, half-width, and edge smoothing determine the location and spatial extent of the group. The hydrophobic region is described in the same way by an error-function envelope, whose midpoint defines the hydrophobic thickness. Once the volume fractions of the lipid groups have been specified, the local water volume fraction is obtained as the complement to unity.

MSDP is implemented in GENFIT, the program developed by the Molecular Biophysics group of the Università Politecnica delle Marche for the analysis of small-angle X-ray and neutron scattering data [28]. This implementation is particularly useful for models of this kind because GENFIT allows parameters to be linked by explicit constraints and functional relations, both within a single dataset and across several datasets. Independent information, common structural parameters, and dataset-specific quantities can therefore be incorporated directly into a global analysis.

In its original formulation, however, MSDP describes a symmetric bilayer: each chemical group in one leaflet has an equivalent group at the mirror-related position in the opposite leaflet. The two leaflets therefore share the same structural parameters, and the model cannot independently describe their transverse structures.

The asymmetric scattering density profile (ASDP) model extends MSDP to bilayers that lack that symmetry, and it is a new model for the analysis of

SAXS data from asymmetric membranes. It assigns to each monolayer its own composition, its own area, and its own group positions and widths, and it describes every chemical group of both monolayers as an erf slab; Sec. 2.6 gives its form. The other major software packages for the analysis of SAXS data available online do not offer this approach. The implementation of ASDP in GENFIT was developed alongside the work reported here, and no application of it to experimental data has been published.

A measured curve, however, is not the form factor of one isolated bilayer. Vesicle suspensions contain a fraction of particles with more than one lamella. The bilayers of such a particle are separated by a nearly constant repeat distance and scatter in phase, which produces lamellar correlation peaks and, more generally, a structure factor that redistributes intensity around them [29, 30]. Structures much larger than the membrane thickness, such as aggregates and residual multilamellar material, raise the intensity at the lowest q with a signal that carries no information on the bilayer.

Each of these contributions has to be represented explicitly in the model, and what is not represented is attributed to the membrane, where it appears as a distortion of the profile. How the signal is decomposed in this work is described in Sec. 2.6.

The information in a SAXS curve of finite q -range and finite precision, however, does not grow with the number of parameters. Whether the data determine what the model is free to vary is the subject of Sec. 1.4.

1.4 The inverse problem

The structural information contained in a SAXS curve is limited in two fundamental ways. First, the cross section in (1.9) depends on the squared modulus of the amplitude, $|F(q)|^2$, rather than on $F(q)$ itself. The phase of the amplitude is therefore not measured, and the real-space profile cannot be recovered directly and uniquely from the scattering intensity alone. Second, the experiment samples only a finite range of q . The corresponding real-space resolution is consequently limited, with the smallest length scale that can be distinguished from the data alone being of the order of $2\pi/q_{\max}$, about 13 Å for the measurements considered here (Sec. 2.5).

The membrane structure must therefore be inferred through a model. A set of structural parameters defines the transverse profile, from which the corresponding SAXS curve is calculated. The parameters are then varied until the calculated curve reproduces the experimental data as closely as possible. The discrepancy between model and experiment is quantified through a weighted sum of squared residuals, with each residual scaled by the uncertainty assigned to the corresponding data point. The objective function minimized in this work has this form, although it should not be interpreted as a reduced χ^2 in the strict statistical sense. The reasons for this distinction, which arise from the treatment of experimental uncertainties, are discussed in Sec. 2.7.

A good agreement between model and data does not, however, imply that all model parameters are independently determined. Two parameters may affect the calculated curve in nearly the same way, so that a change in one can be compensated by a change in the other with little effect on the fit. A recurring example in this work is provided by the two parameters that describe

a chemical-group distribution, its half-width and the smoothing of its edges. Increasing one while decreasing the other can leave the overall extent of the distribution nearly unchanged and, over the accessible q -range, produce almost the same scattering curve. Other combinations of changes alter the profile more strongly and are therefore more readily detected by the data. A fit may thus converge reproducibly even when the individual parameter values remain poorly determined. The position reached along such a weakly constrained direction may depend appreciably on the initial values and on the optimization path, rather than being uniquely selected by the experimental data.

This behavior can be quantified through the local sensitivity of the model to its parameters. Near a minimum, small changes in the calculated curve are described by the first derivatives of the model with respect to those parameters. Arranging these derivatives in a matrix and applying singular value decomposition identifies orthogonal directions in parameter space with different sensitivities of the calculated curve. Large singular values correspond to parameter combinations that strongly affect the observable and are therefore well constrained by the data. Small singular values identify directions along which substantial changes in the parameters produce only minor changes in the calculated curve. In a highly parameterized model, such weakly constrained directions may involve a substantial fraction of the available parameters.

Non-identifiability can arise for two different reasons. It is structural when different parameters enter the observable only through the same combination, so that no measurement of that observable can determine them separately. It is practical when the parameters are distinguishable in principle, but the available data lack the q -range, contrast, or precision required to resolve their different

effects. The procedure used in this work to quantify non-identifiability is described in Sec. 2.11.

Local flat directions are not the only source of ambiguity. The objective function may also contain several distinct minima separated by regions of higher value. Independent optimization runs can then converge to different solutions, depending on their initial conditions and on the stochastic trajectory of the optimizer. This situation is conceptually different from local non-identifiability. A flat direction describes weak sensitivity in the neighborhood of a particular minimum, whereas multiple minima reflect the global structure of the objective-function landscape. The two cases therefore require different forms of analysis.

For this reason, the results reported here do not rely on a single optimization. The final configuration was optimized in thirty independent runs (Sec. 2.10). The resulting solutions separate into two distinct groups, which differ in the extent to which the choline distribution protrudes beyond the phosphate distribution (Chapter 3).

The problem remains tractable because the model parameters are not treated as completely unconstrained quantities. Physical information enters through bounds on parameter ranges, restrictions on the ordering and spatial extent of the chemical groups, molecular volumes obtained from independent measurements, and parameters shared across datasets. A chemical-group model is particularly well suited to this approach because many of its parameters correspond to quantities that can be constrained by independent physical information. The inferred structure therefore results from the combination of the SAXS data with these external constraints, rather than from the scattering data

alone.

This distinction is essential when interpreting the results. Demonstrating that a model is compatible with the data is a weaker statement than assigning a specific value and uncertainty to an individual parameter. The latter requires showing that the parameter, or the relevant combination of parameters, is actually constrained by the measurement rather than primarily by the imposed physical restrictions. Establishing which structural features support which level of inference is therefore itself part of the analysis.

Two complementary approaches are used for this purpose. The derivative-based analysis is local and identifies the combinations of parameters to which the SAXS curves are sensitive around a given minimum. The ensemble of independent optimizations probes the problem more globally by showing which structural features persist across the different minima reached by the optimizer. Both analyses are nevertheless based on the same model, the same experimental data, and the same weighting scheme, and therefore share the assumptions of the SAXS analysis. Molecular dynamics provides an external point of comparison by describing membranes of the same composition without relying on the scattering model or on the parameterization used in the SAXS analysis (Sec. 1.5).

1.5 Molecular dynamics as an external criterion

A molecular dynamics (MD) simulation numerically integrates the equations of motion for all atoms within a hydrated bilayer patch enclosed in a simulation box with periodic boundary conditions in all three dimensions. The inter-

atomic forces derive from an empirical force field, which models the potential energy as a function of atomic coordinates. The parameters of this force field are calibrated using quantum-mechanical calculations on model compounds and experimental data from simpler systems [31]. In the additive formulation standard for lipids, the total energy is decomposed into bonded interactions (internal to the covalent structure) and non-bonded pairwise interactions between non-adjacent atoms [32].

$$\begin{aligned}
U(\mathbf{r}^N) = & \underbrace{\sum_{\text{bonds}} K_b (b - b_0)^2}_{\text{stretching}} + \underbrace{\sum_{\text{angles}} K_\theta (\theta - \theta_0)^2}_{\text{bending}} + \underbrace{\sum_{\text{UB}} K_{\text{UB}} (S - S_0)^2}_{\text{1-3 coupling}} \\
& + \underbrace{\sum_{\text{dihedrals}} K_\phi [1 + \cos(n\phi - \delta)]}_{\text{torsion}} + \underbrace{\sum_{\text{impropers}} K_\omega (\omega - \omega_0)^2}_{\text{planarity}} \quad (1.10) \\
& + \sum_{i < j} \left\{ \underbrace{\epsilon_{ij} \left[\left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^6 \right]}_{\text{dispersion and repulsion}} + \underbrace{\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}}_{\text{electrostatics}} \right\}
\end{aligned}$$

where b , θ , S , ϕ , and ω represent the bond lengths, valence angles, distances between 1–3 atom pairs (separated by two bonds), torsion angles, and improper dihedral angles, respectively; r_{ij} denotes the interatomic distance and q_i the partial atomic charges. The bonded terms (■) constrain the molecular geometry, whereas the non-bonded terms (■), summed over atom pairs separated by more than two bonds, govern chain packing, polar group hydration, and ion interactions, thereby determining the transverse organization of the bilayer along the normal. The simulation yields a trajectory (a sequence of atomic coordinates stored at regular time intervals) from which time averages of physical

quantities are computed. For a lipid membrane, these include the area per lipid, bilayer thickness, group volume, and the trans-bilayer distributions of chemical groups, matching the observables used to describe the transverse profile in Secs. 1.2 and 1.3. For a fluid bilayer comprising a few hundred lipids, trajectories of a few hundred nanoseconds provide sufficient sampling for these structural properties, though slow processes such as inter-leaflet lipid flip-flop require significantly longer timescales.

The simulation does not use the SAXS curves and does not share the assumptions of the model used to analyze them. No functional form is imposed on the group distributions, which are obtained simply by counting the atoms along the bilayer normal; the simulation takes only the composition and the temperature from the experimental conditions. The distributions obtained in this way can be converted into volume fractions and electron density [33]. Represented with the erf slabs of the model, they are reduced to the same structural parameters returned by the fitting procedure, and the residual of that representation measures how well the erf slabs reproduce distributions on which no functional form has been imposed (Sec. 2.15.3).

The comparison is therefore carried out on these quantities, between the values obtained from the fit and those calculated from the trajectory, and not between calculated and measured scattering curves. What is being evaluated is the asymmetric scattering density profile model (ASDP, Sec. 2.6), specifically whether the structure it returns from a curve corresponds to the real structure of a membrane with that composition. Everything that separates a measured intensity from the form factor of a single bilayer, such as the multilamellar component, the non-specific low- q excess intensity, the background, and the

instrumental resolution, belongs to the experiment and not to the model. Therefore, a comparison directly in q -space would include these experimental effects along with the model itself. Both agreement and disagreement are informative, and the goal is neither to validate the simulation against scattering nor scattering against the simulation. This independence applies to the data in this thesis and not to scattering as a general technique, since the force-field parameters for lipids were also tuned using form factors of pure lipid bilayers measured with X-rays and neutrons [31, 34].

The main advantage of the simulation is that it provides structural parameters that SAXS data alone cannot separate. Quantities such as the distance between phosphate and choline groups, the overlap of terminal methyls at the bilayer midplane, and the individual group volumes are directly determined in a simulation trajectory, as they are calculated from atomic positions without relying on an analytical model.

In contrast, the distance between phosphate and choline is poorly determined by the SAXS curves. Without a constraint on the envelope of the two groups, part of the solutions place the outer edge of the choline well inside that of the phosphate, which is physically implausible, and with the constraint the independent runs still divide into two groups that differ in the extent of the choline. This is the behavior of a parameter that the data do not determine, described in general in Sec. 1.4 and in detail in Chapter 3.

The value of this comparison, however, is limited by what the simulation cannot provide. First, the accuracy of the results depends heavily on the force-field parametrization, which cannot be assessed internally by the simulation itself. Second, the time averages are meaningful only if the trajectory is long

enough for the system to sample all relevant configurations, a condition that must be carefully verified. Furthermore, accessible timescales do not cover the flip-flop of phospholipids between leaflets; therefore, in a simulated asymmetric membrane, the distribution of lipids between the two sides reflects the initial setup rather than an equilibrium state [35, 36]. Finally, not all experimentally measured samples have a simulated counterpart; Sec. 2.13 lists the systems that were actually simulated.

The same distributions also allow for a further use of the simulation: as a source for a dataset in which the true underlying structure is exactly known. From the distributions extracted along a trajectory, a synthetic scattering curve is calculated and then analyzed with the same model applied to the experimental data. This curve contains only the signal of a single bilayer, free from any of the experimental contributions that make a measured intensity and a calculated one directly incommensurable. Thus, here too, the comparison is made between structural parameters rather than scattering curves.

By comparing the recovered parameters with those of the representation of the simulated distributions (Sec. 2.15.3), this approach measures how accurately the fit can reconstruct the parameters when the true answer is known. This allows us to separate the non-identifiability inherent to the model from the error of its functional form. The setup of this test is described in Sec. 2.14, while its results are presented in Chapter 3.

1.6 Aims of the thesis

The subject of this thesis is the ASDP model (Sec. 1.3). The work asks what the structure of the measured membranes is and, at the same time, how much of the structure the model returns is established by the measurement.

The model, developed alongside this work, is proposed here as a new description of asymmetric membranes by erf slabs for the analysis of SAXS data, together with its first application to experimental data. The present work also contributes the simultaneous fit of the seven SAXS curves, the molecular dynamics simulations that provide the external criterion, and the MDtoSDP program (Sec. 2.15). From a trajectory, MDtoSDP computes the group volumes and the volume-fraction distributions of the chemical groups, represents the distributions with the erf slabs of the model, and generates the SAXS curves that follow from them.

The data on which the assessment rests are seven SAXS curves from vesicle suspensions, measured in a single session (Sec. 2.3). The compositions can be ordered by increasing complexity. A suspension of a single phosphatidylcholine provides the reference, since its bilayer is symmetric. Adding a sphingomyelin introduces a second species, with a sphingosine backbone and predominantly saturated chains. Cholesterol enters with the three equimolar ternary mixtures, which differ in the source of the sphingomyelin, so that the chain composition of that species varies. Two further preparations, one of them with cholesterol, are made asymmetric by exchanging the lipids of the outer leaflet, and they are the only samples in which the transverse asymmetry that the model is designed to describe is expected. The first five samples, and the inner lamellae of the last two, are described by the symmetric form of

the model, built from the same erf slabs; the asymmetric form enters with the outermost bilayer of the two asymmetric samples (Sec. 2.6).

The assessment is organized around three questions.

1. Which of the structural quantities returned by the model are determined by the SAXS curves, and which are fixed by the physical knowledge introduced alongside them? Physical constraints on the parameters make that boundary visible, since a constraint that excludes configurations without worsening the agreement with the data is acting where the data do not reach (Chapter 3).
2. Do the erf slabs with which the model describes the chemical groups reproduce the distributions of the same groups in a simulated membrane, where those distributions are known? Molecular dynamics provides them without imposing any functional form (Sec. 1.5), and each is represented with the erf slab of the model. The residual of the representation measures how much of the distribution the erf slab cannot reproduce, and the amplitudes computed from the distribution and from its representation show how large an imprint that part leaves on the SAXS curve (Sec. 2.15.3, Chapter 3). The same representation is the reference of a closure test, in which the SAXS curve computed from the simulated distributions is fitted with the model used for the experimental data (Sec. 2.14).
3. Does the structure obtained from the experimental curves agree with a description of membranes of the same nominal composition obtained without using them? Molecular dynamics provides that description, and

the comparison is made on the output parameters of the model (Secs. 2.13 and 2.15.3, Chapter 3).

Chapter 2

Materials and methods

2.1 Materials and samples

2.1.1 Lipids and buffer

The phospholipids were obtained from Avanti Polar Lipids (Alabaster, AL, USA) and comprise 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) and three sphingomyelins of natural origin, extracted from chicken egg yolk (EggSM), from bovine milk (MilkSM) and from porcine brain (BrainSM), all with a stated purity above 99 %. The third component of the mixtures is cholesterol (Chol). Each preparation contains 1 mol% of rhodamine-labeled phosphatidylethanolamine (Rh-PE) as a fluorescent probe; the compositions refer to the structural lipids only. The buffer, common to the hydration of the films, to the exchange and to the measurements, is 25 mM HEPES with 150 mM KCl at pH 7.4 (HK buffer). Lipid exchange is mediated by methyl- β -cyclodextrin (m β CD).

2.1.2 Natural sphingomyelins

Each sphingomyelin is a mixture of species with the same headgroup and the same sphingosine backbone and with acyl chains of different length and unsaturation (Table 2.1). Egg sphingomyelin is 86 % palmitoyl (16:0); brain sphingomyelin is half stearoyl (18:0), with a substantial fraction of nervonoyl (24:1); milk sphingomyelin consists of long-chain saturated species, mostly tricosanoyl (23:0).

Acyl chain	EggSM	MilkSM	BrainSM
16:0	86	16	2
18:0	6	–	50
20:0	–	–	5
22:0	3	20	7
23:0	–	34	–
24:0	–	21	5
24:1	3	3	21
other	2	6	10

Table 2.1: Percentage distribution of the acyl chains in the three natural sphingomyelins, from the supplier’s technical data sheets (Avanti Polar Lipids, products 860061, 860063 and 860062).

In the analysis of the SAXS curves each sphingomyelin is approximated by its most abundant species (palmitoyl, stearoyl, tricosanoyl), which determines the number of methylenes with which it enters the model (Sec. 2.6). In the simulations egg sphingomyelin is represented as a mixture of its four most abundant species (Sec. 2.13).

2.1.3 Samples

The seven measured samples are listed in Table 2.2. Sample 1 is the single-component reference; sample 2 adds egg sphingomyelin in a 2:1 ratio; samples 3, 4 and 5 are equimolar ternary mixtures with cholesterol, which at 37 °C lie in

the coexistence region of the liquid-ordered and liquid-disordered phases [5], and differ in the source of the sphingomyelin [17]. Samples 6 and 7 are the asymmetric preparations, obtained by lipid exchange between an acceptor and a donor population (Sec. 2.2); sample 6 without cholesterol, sample 7 with cholesterol in the acceptor. The measured suspensions have a lipid concentration of 2 mg/mL.

Sample	Nominal composition	Preparation	Profiles
1	POPC	extrusion	1
2	POPC/EggSM 2:1	extrusion	2
3	POPC/MilkSM/Chol 1:1:1	extrusion	3
4	POPC/BrainSM/Chol 1:1:1	extrusion	4
5	POPC/EggSM/Chol 1:1:1	extrusion	5
6	POPC POPC/EggSM 0.80:0.20	exchange, acceptor 1	6 ₂
7	POPC/Chol 0.66:0.34 POPC/EggSM/Chol 0.63:0.20:0.17	exchange, acceptor POPC/Chol	7, 7 ₂

Table 2.2: Nominal composition of the seven samples. For samples 6 and 7 the symbol || separates the inner leaflet from the outer one; the per-leaflet compositions are those expected from the exchange (Sec. 2.2), with POPC/EggSM 0.60:0.40 as donor and, as acceptor, sample 1 and POPC/Chol 0.66:0.34 respectively. The last column lists the labels of the independent transverse profiles that each sample contributes to the fit (Sec. 2.5); the subscript 2 denotes the outer leaflet.

2.2 Vesicle preparation

2.2.1 Symmetric vesicles

The lipids, taken from stock solutions in chloroform in the desired molar ratios, were dried to a film by solvent evaporation and desiccation under vacuum; the film was hydrated with HK buffer at a temperature above the sphingomyelin transition, with agitation. The vesicle suspensions were obtained by extrusion through polycarbonate membranes with 400 nm pores [37, 38].

Samples 1 to 5 were measured in this form; the extruded suspensions of POPC and of POPC/Chol 0.66:0.34 constituted the acceptor populations of samples 6 and 7. Preliminary dynamic light scattering (DLS) measurements on the vesicle suspensions indicated a dominant population with a hydrodynamic diameter close to 200 nm and a relatively narrow size distribution.

2.2.2 Cyclodextrin-mediated exchange

Methyl- β -cyclodextrin, at a concentration below that which dissolves the vesicles, forms reversible complexes with the lipids and carries them between membranes, acting only on the surfaces exposed to the solution [16, 39]. When acceptor vesicles are incubated with a donor population of different composition, the outer leaflet of the acceptors exchanges lipids with the accessible leaflets of the donors, while the inner leaflet remains unchanged on the time scale of the preparation [2].

The preparations follow the heavy-donor protocol of Doktorova and co-workers [16], carried out on the ESRF site immediately before the measurements. The film of the donor vesicles, POPC/EggSM 0.60:0.40, was hydrated in HK buffer with 20 % (w/v) sucrose to 4 mM, with agitation at 55 °C; the suspension of multilamellar vesicles was diluted with buffer and centrifuged at $20000 \times g$ for 30 min, and the pellet was resuspended in 32 mM methyl- β -cyclodextrin in HK buffer, lipid/cyclodextrin molar ratio 1:8, and incubated for 2 h at 50 °C with agitation (600 rpm). The acceptor vesicles were added at a donor/acceptor molar ratio of 2:1 and incubated for 15 min at 50 °C. The exchange was stopped by an eightfold dilution with buffer; the donors were removed by centrifugation at $20000 \times g$ for 30 min at room temperature, and

the supernatant was concentrated in ultrafiltration units with a 100 kDa cutoff at $5000 \times g$ and washed twice with 2 mL of buffer, to remove the cyclodextrin and its complexes.

2.2.3 Expected leaflet compositions

If the exchange brings the lipids of the outer leaflet of the acceptors to equilibrium with the accessible lipids of the donors, present in equal amounts, without selectivity of the cyclodextrin and without redistribution between the leaflets, the outer leaflet takes on the average composition of the two populations and the inner leaflet retains that of the acceptor. For sample 6 the outer leaflet is then POPC/EggSM 0.80:0.20 and the inner leaflet POPC; for sample 7 the outer leaflet is POPC/EggSM/Chol 0.63:0.20:0.17 and the inner leaflet POPC/Chol 0.66:0.34. These are expected, not measured, compositions, and the analysis adopts them as such (Chapter 4).

The SAXS curves of all seven preparations show the lamellar correlation peaks of a multilamellar component. This component was therefore already present in the acceptor suspensions before the exchange, which typically happens upon extrusion of neutral lipids [16]. Since cyclodextrin can only reach the surfaces directly exposed to the solution, it is assumed that, in multilamellar vesicles, the exchange modifies only the outer leaflet of the outermost bilayer, while the inner lamellae remain symmetric (Figure 2.1). The model used to analyze samples 6 and 7 is built on this hypothesis, featuring an inner symmetric stack and an outer asymmetric bilayer (Sec. 2.6). The assignment of the two compositions to the inner and outer leaflets is based on the preparation protocol rather than the scattering curves, since a flat bilayer profile cannot be

distinguished from its mirror image in SAXS experiments (Sec. 2.6).

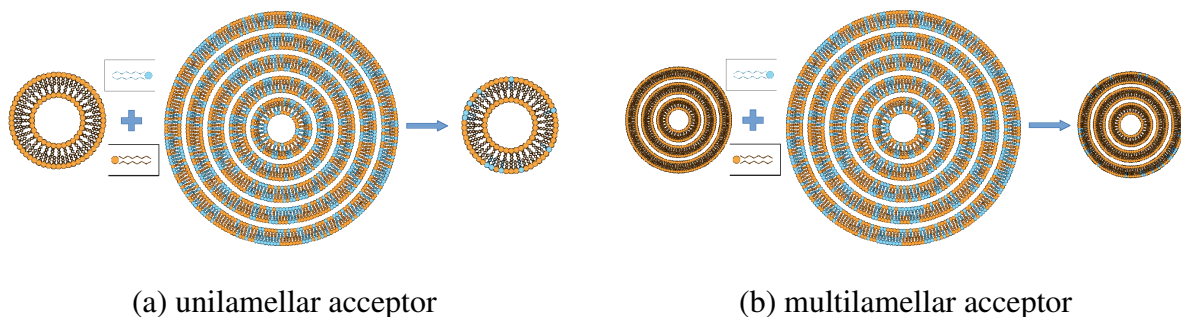


Figure 2.1: Cyclodextrin-mediated exchange on a unilamellar acceptor (a) and on a multilamellar one (b), where only the outer leaflet of the outermost lamella is modified.

The asymmetry obtained by exchange is a non-equilibrium state, which relaxes through spontaneous flip-flop of the phospholipids between the leaflets on time scales from hours to days [2, 16]; preparing immediately before the measurement keeps the delay within the window over which the asymmetry must persist. Cholesterol crosses from one leaflet to the other on much shorter time scales [2, 40], and its distribution between the leaflets, in the samples whose composition includes it, is the one that the system adopts spontaneously during and after the preparation.

2.3 SAXS measurements

The measurements were carried out at beamline ID02 of the European Synchrotron Radiation Facility (ESRF) in Grenoble, in April 2026, as part of experiment LS-3662. The beamline [18] operates at a wavelength of 1 Å with a variable sample-to-detector distance, and records the scattering on an Eiger2 4M detector. The sample is held in a thermostated flow-through cell with a quartz capillary of 2 mm path length (Figure 2.2), in which samples and buffer

are measured in the same capillary. All measurements were performed at 37 °C (310 K).

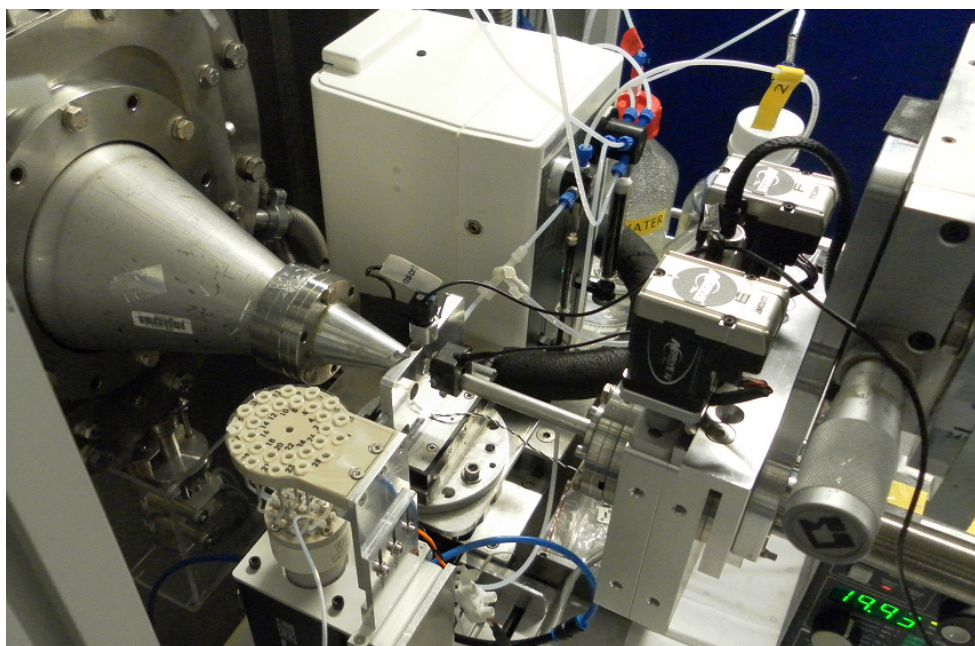


Figure 2.2: The sample environment at ID02, with the thermostated flow-through cell and the quartz capillary in which samples and buffer are measured in sequence.

Each sample was measured at two sample-to-detector distances. The 1.5 m configuration covers the q -range from about $4 \cdot 10^{-3}$ to $0.5 \text{ \AA}^{-1}$, where the form factor of the bilayer carries the information on the transverse profile. The 7 m configuration covers the q -range from about $8 \cdot 10^{-4}$ to $0.11 \text{ \AA}^{-1}$, giving access to the larger length scales of the whole vesicles and to the first orders of the lamellar correlation peak. In both configurations, the azimuthal integration of the 2D detector image yields a 1D scattering curve of approximately 1590 points.

The measurements were carried out first at 1.5 m and then at 7 m. In each configuration the empty capillary and water were measured at the beginning, as instrument diagnostics, followed by the HK buffer, used for solvent subtraction. The samples were loaded manually into the cell, which was rinsed

with buffer between consecutive samples. Each measurement consists of multiple 300 ms exposures, compared with one another before averaging, to exclude those with alterations attributable to radiation damage. During the same beamtime, measurements of the vesicles after manual mixing with the actinoporin Fav were also acquired, together with time-resolved measurements after rapid mixing in a stopped-flow device. Fav is a protein from *Orbicella faveolata* [41]. This work analyzes only the static measurements on the seven samples listed in Table 2.2.

2.4 Data reduction

2.4.1 Integration and absolute scale

The images of each exposure were processed by the automatic pipeline of the beamline [18]. A mask excludes the shadow of the beamstop, the edges of the detector modules and defective pixels; the image is corrected and normalized for flux, transmission and solid angle according to the beamline calibrations and integrated over the azimuthal angle into a curve $I(q)$; division by the sample path length in the capillary, 0.2 cm, gives the differential scattering cross section per unit volume, in cm^{-1} . The absolute scale allows the curves of the two distances to be merged without a scale factor; in the fit the scale of each curve is determined by GENFIT (Sec. 2.6). Consecutive exposures were averaged and the average curve rebinned; each configuration thus yields one curve per sample and one for the buffer.

2.4.2 Solvent subtraction and merging of the two configurations

The buffer curve was subtracted from the sample curve at each distance. To estimate the length scale of the unilamellar component, the hydrodynamic diameter of about 200 nm indicated by the preliminary DLS (Sec. 2.2) is used, with the vesicle idealized as a thin spherical shell of non-zero net contrast. The radius of gyration is then of the order of the vesicle radius, $R_g \simeq 100$ nm. The geometric scale of the Guinier regime, $qR_g \lesssim 1$, is therefore $q \lesssim 10^{-3} \text{ \AA}^{-1}$; the 7 m configuration, which reaches about $8 \cdot 10^{-4} \text{ \AA}^{-1}$, approaches this region, whereas the 1.5 m configuration does not cover it. The long-distance measurement also served to search for the contribution of the vesicle as a whole [42].

In the subtracted curves, however, no Guinier plateau attributable with confidence to the vesicles was identified, since at 7 m only the upper edge of the theoretical region is accessible.

At the lowest q some curves show negative intensities after subtraction, others a positive excess. Residual buffer scattering and contributions from larger structures can therefore hinder the attribution of the unilamellar signal. The structural analysis is restricted to $q \geq 0.005 \text{ \AA}^{-1}$ because of the poor reliability of the lower region. The empirical power-law term of the fit (Sec. 2.6) describes the low- q excess within the analyzed range, without identifying its origin or demonstrating a Porod-like regime.

For samples 2, 3, 5 and 7 the subtracted curves at the two distances were merged without a scale factor, because they were already on an absolute scale and coincident in the overlap region, where the junction point was chosen. For samples 1, 4 and 6 the 7 m configuration added no information above the

lower limit of the analysis, and only the 1.5 m curves were used. Subtraction and merging were performed with RAW [43], the final rebinning with SAX-Sutilities2 [44]; the processing chain is recorded in the header of each file. The result is seven curves $I(q)$ on an absolute scale (Table 2.3), each with a column of uncertainties propagated from the counting statistics; their use in the fit is described in Sec. 2.7.

Sample	Composition	Points	$q_{\min}$ ($\text{\AA}^{-1}$)	$q_{\max}$ ($\text{\AA}^{-1}$)	Distances
1	POPC	484	$4.27 \cdot 10^{-3}$	0.494	1.5 m
2	POPC/EggSM	646	$8.49 \cdot 10^{-4}$	0.439	merged
3	POPC/MilkSM/Chol	613	$8.49 \cdot 10^{-4}$	0.475	merged
4	POPC/BrainSM/Chol	477	$4.27 \cdot 10^{-3}$	0.467	1.5 m
5	POPC/EggSM/Chol	632	$8.49 \cdot 10^{-4}$	0.455	merged
6	asymmetric	511	$3.84 \cdot 10^{-3}$	0.467	1.5 m
7	asymmetric with Chol	708	$8.49 \cdot 10^{-4}$	0.468	merged

Table 2.3: Number of points and q range of the seven reduced curves, with the distance configurations that contribute to each.

2.5 Simultaneous fit

The seven curves of Table 2.3 were analyzed with a single parameter vector (Sec. 2.6), minimized over all of them simultaneously, with some parameters shared among systems and others specific to each system.

2.5.1 Independent profiles

In systems 1–5 the two leaflets have the same composition and the model describes only one of them, mirrored about the bilayer midplane; each system contributes one transverse profile. In systems 6 and 7 each leaflet is described by its own set of positions and widths, with the composition fixed at the nom-

inal value. Case 7 contributes two profiles, one per leaflet; case 6 contributes only one, because its inner leaflet is pure POPC and its parameters are identified with those of case 1, under the assumption that the acceptor leaflet remains unperturbed by the exchange. There are therefore eight independent profiles against seven curves, with the labels 1, 2, 3, 4, 5, 6₂, 7, 7₂ of Table 2.2, and the statistics built on the repeated runs (Sec. 2.9) count eight profiles per run.

2.5.2 Shared parameters

As a working hypothesis and to minimize the number of free parameters, the volumes of the chemical groups are shared among the seven systems. This is justified because the sensitivity of the SAXS data to small volume variations (a few percent) across different compositions is comparable to the magnitude of the variation itself. Allowing the group volumes to vary freely for each system would introduce strongly correlated parameters that the data cannot uniquely determine. In contrast, the parameters describing the structural organization of the membrane (area per lipid, hydrophobic thickness, and the positions and widths of the chemical groups) remain specific to each system, as they depend directly on composition and physical phase. Under this scheme, curves 6 and 7 provide the primary access to transverse asymmetry, while the symmetric systems help constrain the shared molecular quantities.

2.5.3 q range and resolution

The scattering vector range used in the fit is $0.005 \leq q \leq 0.5 \text{ \AA}^{-1}$, where the lower limit is determined by the data reduction process (Sec. 2.4). The upper

limit sets the nominal spatial resolution, defined as:

$$\delta \simeq \frac{2\pi}{q_{\max}} \simeq 13 \text{ \AA}, \quad (2.1)$$

which represents the length scale below which two unmodeled details of the profile cannot be resolved. However, the position of a distribution with a known functional shape can be determined on a finer scale, especially when the signal-to-noise ratio is high and parameter correlations are weak; for the widths of the chemical groups this is examined in Sec. 3.3. The extra decimal digits returned by the optimization beyond this nominal limit depend on the applied restraints and parameter correlations. For this reason, comparisons with the simulations are made on spatial features defined to within a few ångströms, such as group positions, widths, and bilayer thicknesses.

2.6 The ASDP model

2.6.1 Molecular unit and groups

A flat bilayer dispersed in water is considered, with z the axis normal to the membrane plane and the origin at the midplane. The two monolayers occupy the regions $z < 0$ and $z > 0$ and are indexed by $\alpha = 1, 2$. The membrane is composed of N_S molecular species. For species j , $x_{j,\alpha}$ is the fraction of its molecules residing in monolayer α ($\sum_{\alpha} x_{j,\alpha} = 1$) and C_j its overall molar concentration; the stoichiometric coefficient of the species in the molecular

unit of the monolayer is

$$\zeta_{j,\alpha} = \frac{x_{j,\alpha} C_j}{\sum_{j'} x_{j',\alpha} C_{j'}}, \quad \sum_j \zeta_{j,\alpha} = 1. \quad (2.2)$$

The molecular unit is the average composition of a monolayer normalized to one molecule; associated with it are the area A_α , the thickness D_α and the volume $V_\alpha = A_\alpha D_\alpha$.

Each molecular unit is partitioned into united groups, following the scattering density profile approach [34]. Each group g is defined by its volume $v_{g,\alpha}$ and its scattering length $b_{g,\alpha}$ (which, for X-rays, is proportional to the number of electrons). These properties yield the SLD $\rho_{g,\alpha} = b_{g,\alpha}/v_{g,\alpha}$, proportional to the electron density in the X-ray case. The number of groups g in the molecular unit is determined by stoichiometry: $n_{g,\alpha} = \sum_j \zeta_{j,\alpha} m_{g,j}$, where $m_{g,j}$ is the number of groups g in molecule j . The groups are ordered sequentially from the hydrophobic core toward the solvent: the first N_h groups describe the hydrocarbon chains, the next N_p describe the polar headgroup region, and the final group ($g = N_h + N_p + 1 \equiv N$) represents the hydration water, whose SLD $\rho_{N,\alpha}$ is allowed to differ from that of bulk water. The first group ($g = 1$, the methylene) occupies most of the hydrophobic region and its distribution is not modeled explicitly. Finally, the total volume of the hydrated unit is the sum of the volumes of the three regions: $V_\alpha = V_{h,\alpha} + V_{p,\alpha} + V_{w,\alpha} = \sum_g n_{g,\alpha} v_{g,\alpha}$.

The adopted partition is shown in Figure 2.3. Cholesterol constitutes a single group, and the position of the hydroxyl, the only quantity for which measurements exist, therefore cannot be constrained. The two carbons of the trans double bond of sphingomyelin are merged into the sphingosine backbone, whereas in POPC the two unsaturated carbons remain a separate group; the

backbone so defined does not coincide with any group for which the literature reports volumes or positions.

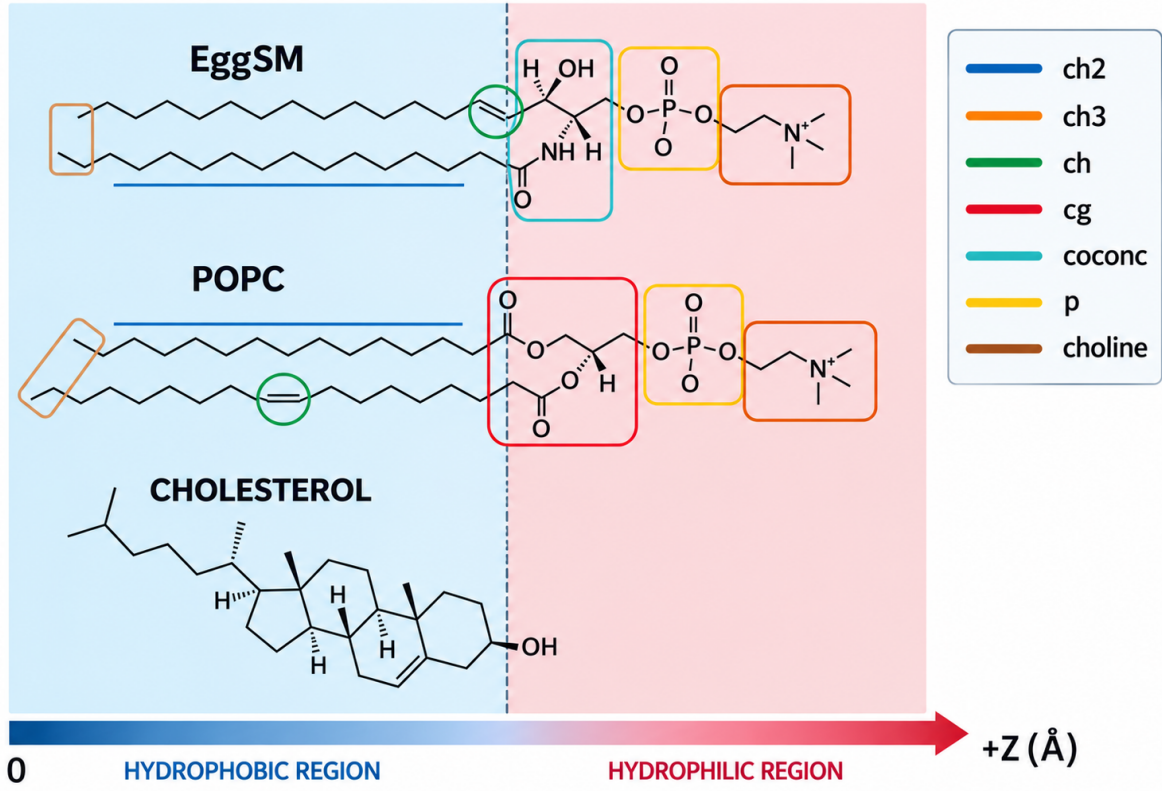


Figure 2.3: The united groups marked on the chemical structures of sphingomyelin, POPC and cholesterol, ordered along the normal z from the hydrophobic region to the polar region and the solvent. The abbreviations of the legend are the methylene (ch2), the terminal methyl (ch3), the unsaturated carbons (ch), the glycerol (cg), the sphingosine backbone (coconc), the phosphate (p) and the choline.

2.6.2 Volume fractions

The central quantity of the method is the volume-fraction distribution $\phi_{g,\alpha}(z)$, the fraction of the volume available at coordinate z that is occupied by group g of monolayer α . The distributions are subject to the space-filling constraint over the whole bilayer,

$$\sum_{\alpha=1}^2 \sum_{g=1}^N \phi_{g,\alpha}(z) = 1, \quad (2.3)$$

and their integral is fixed by the group volume and the area:

$$\int_{-\infty}^{+\infty} \varphi_{g,\alpha}(z) dz = \frac{n_{g,\alpha} V_{g,\alpha}}{A_\alpha}. \quad (2.4)$$

For the groups from $g = 2$ to $g = N_h + N_p$ a combination of two error functions is adopted,

$$E_2(z; z_{g,\alpha}, w_{g,\alpha}, \sigma_{g,\alpha}) = \frac{1}{2} \left[\operatorname{erf} \left(\frac{z - z_{g,\alpha} + w_{g,\alpha}}{\sqrt{2} \sigma_{g,\alpha}} \right) - \operatorname{erf} \left(\frac{z - z_{g,\alpha} - w_{g,\alpha}}{\sqrt{2} \sigma_{g,\alpha}} \right) \right], \quad (2.5)$$

where $z_{g,\alpha}$ is the position of the center, $w_{g,\alpha}$ the half-width of the slab and $\sigma_{g,\alpha}$ the smoothing of its edges. The Fourier transform is analytic,

$$\int_{-\infty}^{+\infty} E_2(z; z_{g,\alpha}, w_{g,\alpha}, \sigma_{g,\alpha}) e^{iqz} dz = 2w_g \frac{\sin(qw_g)}{qw_g} e^{-q^2 \sigma_g^2/2} e^{iqz_g}, \quad (2.6)$$

and the three parameters separate into three factors, since the half-width governs the oscillations, the smoothing a Gaussian damping at high q and the position only the phase. For $q = 0$ the integral equals $2w_g$, and (2.4) fixes the prefactor,

$$\varphi_{g,\alpha}(z) = \frac{n_{g,\alpha} V_{g,\alpha}}{2A_\alpha w_{g,\alpha}} E_2(z; z_{g,\alpha}, w_{g,\alpha}, \sigma_{g,\alpha}). \quad (2.7)$$

2.6.3 Hydrophobic envelope

The methylene and the hydration water are obtained by difference, and the construction guarantees (2.3). The hydrophobic envelope is described by a pair of error functions, with a thickness $D_{h,\alpha}$ and a smoothing $\sigma_{h,\alpha}$ specific to

each monolayer and a parameter σ_{mid} that describes the interpenetration of the chains at the midplane:

$$\varphi_{h,\alpha}(z) = \frac{(-1)^\alpha}{2} \left[\text{erf}\left(\frac{z}{\sqrt{2}\sigma_{\text{mid}}}\right) - \text{erf}\left(\frac{z - (-1)^\alpha D_{h,\alpha}}{\sqrt{2}\sigma_{h,\alpha}}\right) \right]. \quad (2.8)$$

The methylene is what remains of the envelope after subtracting the other apolar groups, and the hydration water what remains of the monolayer after subtracting the hydrophobic region and the polar groups:

$$\varphi_{1,\alpha}(z) = \varphi_{h,\alpha}(z) - \sum_{g=2}^{N_h} \varphi_{g,\alpha}(z), \quad (2.9)$$

$$\varphi_{N,\alpha}(z) = H_\alpha(z) - \lim_{\sigma_{\text{mid}} \rightarrow 0} \varphi_{h,\alpha}(z) - \sum_{g=N_h+1}^{N_h+N_p} \varphi_{g,\alpha}(z), \quad (2.10)$$

where $H_\alpha(z)$ equals one inside the monolayer and zero outside, and the total thickness is defined by the outermost group, $D_\alpha = \max_g \{|z_{g,\alpha}| + w_{g,\alpha} + 3\sigma_{g,\alpha}\}$. The limit in (2.10) makes the interpenetration terms of the two monolayers, of opposite sign, cancel in the sum over the bilayer. Integrating (2.10) gives the number of hydration water molecules per molecular unit, $n_{N,\alpha} = (A_\alpha D_\alpha - V_{h,\alpha} - V_{p,\alpha})/v_{N,\alpha}$; the hydration follows from the geometry and the volumes and is not a free parameter.

2.6.4 Amplitude

The amplitude is the Fourier transform of the excess SLD relative to the solvent, whose SLD is ρ_0 . Substituting (2.9) and (2.10) into the sum over all

groups yields four contributions:

$$\begin{aligned}
 F(q) = \sum_{\alpha=1}^2 \left\{ \right. & (\rho_{1,\alpha} - \rho_0) \frac{i(-1)^\alpha}{q} \left[e^{-q^2 \sigma_{\text{mid}}^2/2} - e^{-q^2 \sigma_{h,\alpha}^2/2} e^{i(-1)^\alpha q D_{h,\alpha}} \right] \\
 & + (\rho_{N,\alpha} - \rho_0) \frac{i(-1)^\alpha}{q} \left[e^{-q^2 \sigma_{h,\alpha}^2/2} e^{i(-1)^\alpha q D_{h,\alpha}} - e^{i(-1)^\alpha q D_\alpha} \right] \\
 & + \frac{1}{A_\alpha} \sum_{g=2}^{N_h} (\rho_{g,\alpha} - \rho_{1,\alpha}) n_{g,\alpha} v_{g,\alpha} \frac{\sin(q w_{g,\alpha})}{q w_{g,\alpha}} e^{-q^2 \sigma_{g,\alpha}^2/2} e^{i q z_{g,\alpha}} \\
 & + \frac{1}{A_\alpha} \sum_{g=N_h+1}^{N_h+N_p} (\rho_{g,\alpha} - \rho_{N,\alpha}) n_{g,\alpha} v_{g,\alpha} \frac{\sin(q w_{g,\alpha})}{q w_{g,\alpha}} e^{-q^2 \sigma_{g,\alpha}^2/2} e^{i q z_{g,\alpha}} \left. \right\} \quad (2.11)
 \end{aligned}$$

■ **Hydrophobic envelope**, with the contrast of the methylene relative to the solvent; the two exponentials are the midplane and the interface with the polar region.

■ **Hydration water**, from the interface between the hydrophobic and the polar region to the edge of the monolayer, with the contrast relative to bulk water; it vanishes if $\rho_{N,\alpha} = \rho_0$.

■ **Apolar groups** (terminal methyls, unsaturated carbons, cholesterol), with the contrast relative to the methylene they replace.

■ **Polar groups** (glycerol or backbone, phosphate, choline), with the contrast relative to the hydration water they displace.

The two contributions in $1/q$ are transforms of smoothed steps and describe edges; the two under summation derive from (2.6) and describe peaks, each weighted by the group volume per unit area, $n_{g,\alpha} v_{g,\alpha}/A_\alpha$. A group whose SLD equals that of the medium it replaces produces no contrast. The modulus of (2.11) for the POPC sample is in Figure 2.4.

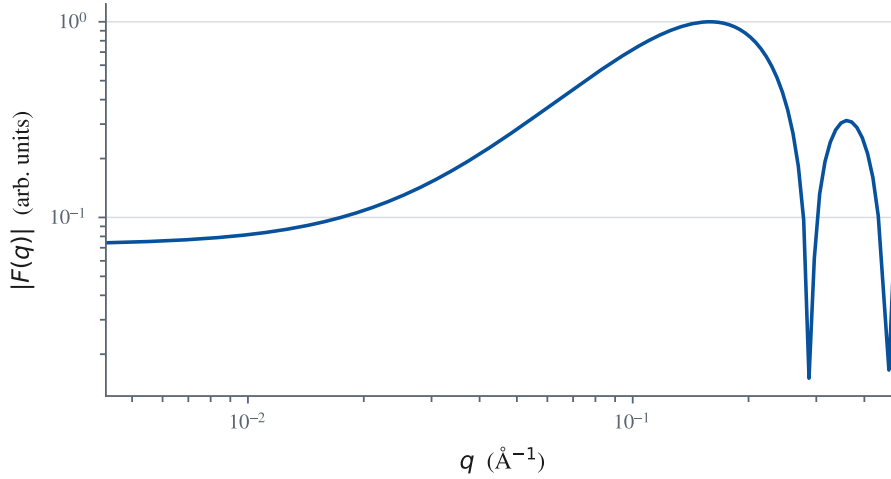


Figure 2.4: The modulus of the amplitude (2.11) for the POPC sample, from one run of the final set, in arbitrary units. The minima fall at about 0.29 and $0.46 \text{ \AA}^{-1}$ and separate the lobes that carry the information on the transverse profile; their depth on the plot is set by the spacing of the grid in q . Both axes are logarithmic.

2.6.5 Symmetric bilayer

The symmetric and the asymmetric bilayer are two distinct models in GENFIT, built with the same combination of error functions (2.5) and (2.8). In the symmetric one the two monolayers share composition, area, hydrophobic thickness, volumes and shape parameters, with positions reflected about the midplane, so that a single set of z_g , w_g , σ_g describes both leaflets and the bilayer amplitude is twice the real part of the amplitude of one monolayer. In the asymmetric one two sets are needed, and the sum in (2.11) runs over both leaflets. In what follows $F(q)$ always denotes the amplitude of the whole bilayer.

When the positions are reflected the contributions of the two monolayers are complex conjugates of each other and $F(q)$ is real; in the asymmetric case the cross term between the two leaflets in $|F(q)|^2$ is no longer fixed by one leaflet alone and depends on the difference between their profiles. The anal-

ysis of systems 6 and 7 exploits this sensitivity of the intensity to the loss of symmetry. The amplitude is however the transform of a real profile, so that $F(-q) = F^*(q)$, and the reflection $z \rightarrow -z$ of the whole profile, which maps $F(q)$ to $F(-q)$, leaves $|F(q)|^2$ unchanged. For a flat bilayer, in which nothing distinguishes the two faces, an asymmetric profile and its mirror image therefore produce the same curve. The data fix the difference between the two leaflets, not which of the two is the outer one; the assignment of the compositions to the leaflets comes from the preparation (Sec. 2.2).

2.6.6 Intensity and structure factor

The conversion from amplitude to intensity is the one implemented in GENFIT [28]:

$$I(q) = K |F(q)|^2 \frac{2\pi}{q^2 d_B} S(q), \quad (2.12)$$

where $2\pi/(q^2 d_B)$ is the factor derived in (1.9) for randomly oriented flat particles, $1/q^2$ being the Lorentz factor, and $d_B = d_{B,1} + d_{B,2}$ is the Luzzati thickness of the bilayer [22], with $d_{B,\alpha} = V_{\text{dry},\alpha}/A_\alpha$ and $V_{\text{dry},\alpha}$ the dry volume of the molecular unit; in the symmetric case $d_{B,\alpha} = d_B/2$. The constant K is computed for each curve from the nominal concentration, the molecular weight, the square of the classical electron radius, the dry volume of the molecular unit and the weight f_{ML} or $1 - f_{\text{ML}}$ of the contribution it refers to, and it fixes the units of the computed intensity. The effective scale of each curve is instead determined by the factor described below. At each evaluation of the model, GENFIT multiplies the computed curve by a scale factor specific to that curve and adds a flat background to it, both obtained in closed form by weighted least squares

against the data and external to the parameter vector. The absolute scale of the measurement therefore does not constrain the structural parameters, and the concentration enters only through K , which is reabsorbed by the scale factor.

For stacked membranes $S(q)$ is the structure factor of the modified Caillé theory [29, 30], averaged over a distribution of the number of lamellae:

$$S(q) = 1 + \sum_n g_{N_{\text{lam}},n} \cos(nqd) e^{-\gamma_E \eta_q (\pi n)^{-\eta_q}}, \quad \eta_q = \left(\frac{qd}{2\pi}\right)^2 \eta, \quad (2.13)$$

with γ_E the Euler–Mascheroni constant and $g_{N_{\text{lam}},n}$ the weights of the pairs of lamellae separated by n repeat distances. The free parameters are three per curve, the number of lamellae N_{lam} , the repeat distance d and the relative fluctuation σ_d/d ; the Caillé parameter is derived from the latter as $\eta = (\sigma_d/d)^2/0.087$, the relation adopted in the implementation [28]. Compared with a paracrystal, the Caillé theory produces peaks with power-law tails, and the description of the tails determines how much intensity in the peak region is attributed to the correlation between lamellae and how much to the single bilayer. Figure 2.5 shows the resulting shape for the POPC sample.

2.6.7 Signal composition

Each curve is described by the sum of two weighted contributions and an empirical term:

$$\frac{d\Sigma}{d\Omega}(q) = f_{\text{ML}} \left. \frac{d\Sigma}{d\Omega} \right|_{\text{multi}}(q) + (1 - f_{\text{ML}}) \left. \frac{d\Sigma}{d\Omega} \right|_{\text{single}}(q) + \kappa q^{-\alpha_p}. \quad (2.14)$$

The first term is the multilamellar contribution, with (2.13) active; the second an isolated bilayer, with $S(q) = 1$; both use the same constant K , so that

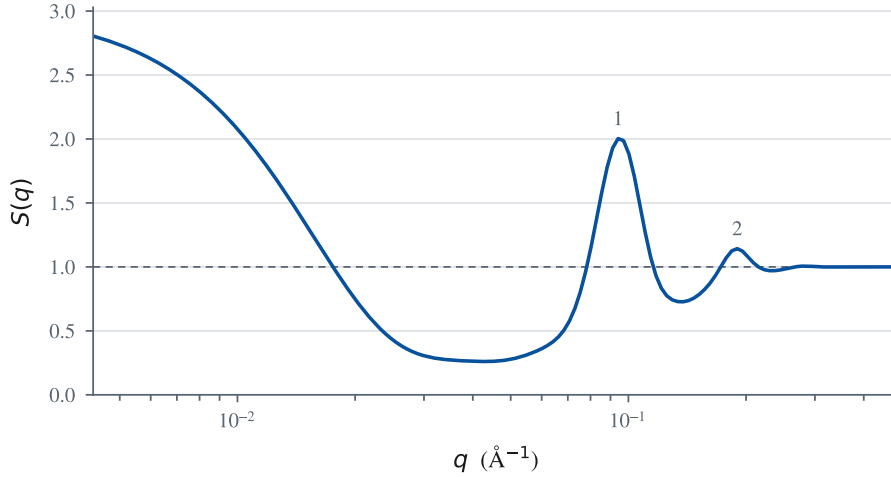


Figure 2.5: The structure factor (2.13) of the multilamellar contribution for the POPC sample, from one run of the final set, with the first two orders of the lamellar reflection marked. For the contribution of the isolated bilayer $S(q) = 1$ over the whole range. The q axis is logarithmic and the ordinate linear.

f_{ML} partitions the intensity without introducing a second scale. The third term describes the low- q excess produced by structures much larger than the membrane thickness, with its own scale κ , and serves to separate that background from the bilayer signal; the per-curve flat background is added to the sum.

In systems 1–5 the two bilayer terms use the symmetric model. In systems 6 and 7 the first remains symmetric and represents the inner lamellae; the second uses the asymmetric amplitude and represents the outermost lamella of the same particle. The two contributions are summed in intensity and the interference term between the outer lamella and the inner ones is neglected. For systems 6 and 7, (2.14) is therefore an effective decomposition of the signal of a multilamellar particle in which only the outer lamella is asymmetric.

2.6.8 Free parameters

The vector of minimized parameters has 215 components (Table 2.4). Each parameter is assigned a range of variation fixed before the optimization and

inherited from the preliminary analysis of the same data. For the 42 parameters of the structure factor, of f_{ML} and of the background the ranges are narrow, 2 to 10 % of their central value, and these parameters can vary only slightly. For the structural quantities the ranges are operational limits (Sec. 2.8), and their effect on the solutions is examined in Sec. 3.1.

Category	Number
Positions, half-widths and smoothing parameters of the groups and the hydrophobic envelope	155
Area per molecular unit, one per transverse profile	8
Group volumes, shared among the systems	10
Weight of the multilamellar contribution (f_{ML}), one per curve	7
Lamellar structure factor (d , N_{lam} , σ_d/d)	21
Empirical background (κ , α_p)	14
Total	215

Table 2.4: Breakdown of the free parameters of the simultaneous fit.

2.7 Data weighting

2.7.1 Error law and residual grid

The uncertainties propagated through the data reduction (Sec. 2.4) do not enter the fit. GENFIT reconstructs them point by point from a reference relative-error coefficient and from the measured intensity, according to the power law

$$\sigma_i = k_e I_1^{1-\beta} |I_i|^\beta, \quad \beta = 0.8, \quad (2.15)$$

where $I_1 > 0$ is the intensity of the first point of the curve and k_e is the relative error assigned to that point [28]. The weighted residual is the relative deviation multiplied by $(I_i/I_1)^{1-\beta}$. With $\beta = 1$ all points count equally; with the exponent of counting statistics, $\beta = 1/2$, the fit is dominated by the first points and

poorly describes the region at intermediate and high q , where the form factor carries the structural information. The value $\beta = 0.8$, common to the seven experiments, gives all points comparable importance, with a slight preference for those with higher intensity. The seven coefficients k_e were chosen so as to equalize the assumed relative error at a common reference q , $q \simeq 0.05 \text{ \AA}^{-1}$, rather than at the first point, because the curves start at different q .

Each curve is re-interpolated onto a logarithmic grid of 150 points, with linearly interpolated weights; the residuals evaluated within the fit range number 872, distributed among the seven experiments as 145/111/113/145/109/141/108. The grid makes the number of residuals per decade of q uniform and the same for all curves; the exponent sets how much weight each residual carries. The grid points are interpolated values, not independent measurements; the grid is a rule for weighting the regions of q .

2.7.2 Definition of χ_{data}^2

The quantity minimized in the data term of the objective function is

$$\chi_{\text{data}}^2 = \frac{1}{7} \sum_{e=1}^7 \frac{1}{n_{q,e}} \sum_{i=1}^{n_{q,e}} \left[\frac{I_{ei}^{\text{obs}} - I_{ei}^{\text{mod}}}{\sigma_{ei}} \right]^2, \quad (2.16)$$

where $n_{q,e}$ is the number of grid points of experiment e within the fit range. The inner average makes the contribution of an experiment independent of the number of its points, and the outer average gives the seven cases the same weight. Since the uncertainties are those constructed from (2.15), the scale of χ_{data}^2 is conventional, and its value can be used for comparisons between models evaluated on the same data with the same weighting, not as an absolute measure of

agreement nor as a basis for confidence intervals derived from the shape of the objective function. The structural quantities are therefore accompanied by the spread across independent runs (Sec. 2.12).

A common rescaling of the k_e would not shift the minimum of χ^2_{data} but would change its ratio to the restraint terms (Sec. 2.8), whose weights are calibrated on the scale of the adopted k_e . The entire analysis was carried out with a single weighting, and the conclusions hold for that weighting (Chapter 4).

2.8 Constraints

The 215 parameters are not free to take any combination of values. The data alone do not exclude combinations without chemical meaning, such as a choline deeper than the phosphate to which it is bound or a molecular volume far from densitometry, and the exclusions are formulated as described below. Some of the constraints transfer information from the literature, measured on systems analogous but not identical to those of this work; the rest are operational regularizations, ranges and weights chosen to prevent unphysical compensations. The origin of each is reported in Table 2.5.

2.8.1 Hard bounds and soft restraints

The range of variation of each parameter is a hard bound that the optimizer cannot cross, and it delimits the domain that the optimizer explores. A soft restraint is a term added to the objective function, zero in the allowed region and increasing outside it; it can be violated if the data require it, at an explicit cost. This is the form adopted for all chemical and geometric conditions, be-

cause none is known with the certainty that a hard bound would imply. The two tools are not used together on the same quantity. The area of the pure-POPC system is handled with the range $62 \leq A^{(1)} \leq 72 \text{ \AA}^2$ alone, whereas in the systems with cholesterol the lower bound is reduced to $45 \text{ \AA}^2$, because the condensation produced by the sterol must emerge from the data.

For a quantity x to be kept within $[L, U]$ a flat-bottom window is used, for an inequality $x \leq 0$ a one-sided threshold:

$$W(x; L, U, s_0) = \begin{cases} [(L-x)/s_0]^2, & x < L, \\ 0, & L \leq x \leq U, \\ [(x-U)/s_0]^2, & x > U, \end{cases} \quad W_+(x; s_0) = \left[\frac{\max(0, x)}{s_0} \right]^2, \quad (2.17)$$

with s_0 a scale that converts the violation into a dimensionless number. In the allowed region the penalty and its derivative vanish, so that a window does not attract the parameter toward the center and a threshold does not push it toward the limit, and a soft restraint is not an imposed measurement. The minimized function is

$$\Phi = \chi_{\text{data}}^2 + \sum_k \lambda_k P_k, \quad P_k = \frac{1}{n_k} \sum_{j=1}^{n_k} c_{k,j}, \quad (2.18)$$

with $c_{k,j}$ the cost of the individual residual of family k and n_k the number of its residuals; the average prevents a condition imposed on eight profiles from weighing four times as much as the same condition imposed on two.

Restrained quantity	Source	Form	Allowed region	Scale	λ
Dry molecular volumes (5)	densitometry ^a	window	see below	25–65 Å ³	1
Polar headgroup volumes (2)	partition from the literature ^b	window	321–341; 264–284 Å ³	15 Å ³	10 ^{−2}
Relations among the areas (9)	expected condensation	threshold	(2.19)	2 and 3 Å ²	3 · 10 ^{−2}
Order of the polar groups (22)	chemical connectivity	threshold	(2.20)	0.5 Å	1
Glycerol anchoring (8)	hydrophobic thickness ^c	window	0.45–1.50 Å	0.5 Å	10 ^{−2}
Backbone anchoring (6)	analogy	window	−1.50–1.50 Å	0.5 Å	3 · 10 ^{−3}
Choline extension (8)	volume balance	threshold	(2.22)	1 Å	10 ^{−2}
Cholesterol edge (5)	interfacial hydroxyl ^d	window	−2–2 Å	1 Å	3 · 10 ^{−3}
Effective widths (5 families)	literature profiles	window	1.8–3.8 Å	0.25 Å	3 · 10 ^{−2}
Phosphate–choline envelope (8)	this work	threshold	(2.26)	1 Å	3 · 10 ^{−3}

Dry molecular volumes: POPC 1239–1289; cholesterol 611–649; egg sphingomyelin 1096–1188; brain 1152–1248; milk 1252–1384 Å³. Effective widths: glycerol 2.0–3.2; sphingosine backbone 2.0–3.5; phosphate 1.8–3.2; unsaturated carbons 1.8–3.2; cholesterol 2.0–3.8 Å; maximum spread across profiles 0.8 or 1.0 Å depending on the group.

^a [45–47]; ^b [24, 48]; ^c [25]; ^d [49].

Table 2.5: The ten families of soft restraints, with the source of the information, the form of the penalty (2.17), the allowed region, the scale s_0 and the weight λ in (2.18). In parentheses, the number of residuals of each family. The weights are not comparable across rows with different scales.

2.8.2 Volumes and areas

The group volumes are shared among the systems (Sec. 2.5), and the dry molecular volumes of the five species follow from them as stoichiometric combinations; it is these, and not the individual v_g , that are comparable with densitometry. The ranges are centered on the literature values for lipids and sterols in the fluid phase [45], for egg sphingomyelin [46] and for milk sphingomyelin [47], and widened (from ± 19 to ± 66 Å³) to cover the temperature, phase and composition of the mixtures, which no source reproduces exactly. A further condition concerns the sums $v_{PC}^{\text{head}} = v_{CG} + v_P + v_{\text{choline}}$ and $v_{SM}^{\text{head}} = v_{\text{amide}} + v_P + v_{\text{choline}}$, the only combinations comparable with the polar headgroups of the literature, around 331 Å³ for phosphocholine [24] and 274 Å³ for sphingomyelin [48]; the windows are regularizations, since different parametrizations of the same system give different values [47].

The area A_α is the projected area per average molecular unit, cholesterol

included, and is not comparable with the single-species areas of the literature. With the superscript labeling the transverse profile (Sec. 2.5), the conditions imposed express a weak ordering:

$$\begin{aligned} A^{(i)} &\leq A^{(1)} + 2 \text{ \AA}^2, & i = 2, 3, 4, 5, 7, \\ A^{(2)} - 1 \text{ \AA}^2 &\leq A^{(6_2)} \leq A^{(1)} + 3 \text{ \AA}^2, \\ A^{(5)} - 3 \text{ \AA}^2 &\leq A^{(7_2)} \leq A^{(6_2)} + 3 \text{ \AA}^2. \end{aligned} \tag{2.19}$$

No mixture with sphingomyelin or cholesterol has an area larger than that of pure POPC, and the asymmetric outer leaflets lie between the compositions from which they derive; their areas remain independent parameters, because the condensation is not linear in the composition.

2.8.3 Polar region and cholesterol

The group positions are parametrized as fractions of the hydrophobic thickness, $z_g = D_{h,\alpha} f_g$; the geometric conditions are expressed in ångströms and recomputed at each evaluation from the current thickness. From the center toward the solvent the phosphate follows the glycerol and precedes the choline,

$$z_P - z_{CG} \geq 0.5 \text{ \AA}, \quad z_P - z_{amide} \geq 1.0 \text{ \AA}, \quad z_{choline} - z_P \geq 0.5 \text{ \AA}, \tag{2.20}$$

a connectivity condition that receives the highest weight of all and imposes no upper limits on the separations. The sphingosine backbone is not ordered with respect to the glycerol. The sequence is anchored to the hydrophobic thickness

by

$$0.45 \leq z_{\text{CG}} - D_{h,\alpha} \leq 1.50 \text{ \AA}, \quad -1.50 \leq z_{\text{amide}} - D_{h,\alpha} \leq 1.50 \text{ \AA}. \quad (2.21)$$

The glycerol window derives from the distance of 0.45–1.1 Å between the glycerol and the edge of the hydrocarbon region measured by Eicher and co-workers [25]; in the model that edge is the hydrophobic thickness $D_{h,\alpha}$, the center of the error function that delimits the envelope. The backbone window, for a group that corresponds to no measurement, is wider and symmetric, with one third of the weight.

The center of the choline cannot lie beyond the Luzzati thickness of its own monolayer,

$$z_{\text{choline},\alpha} \leq d_{B,\alpha}, \quad (2.22)$$

with $d_{B,\alpha}$ computed from the dry volume and the area of the single leaflet. Since both sides contain the area in the denominator, the condition is equivalent to $z_{\text{choline},\alpha}/D_{h,\alpha} \leq V_{\text{dry},\alpha}/V_{h,\alpha}$ and does not constitute an independent determination of A_α . For cholesterol, a single group, the nominal edge is restrained,

$$-2 \leq z_{\text{Chol}} + w_{\text{Chol}} - D_{h,\alpha} \leq 2 \text{ \AA}, \quad (2.23)$$

consistent with the hydroxyl located near the interface between water and lipid [49]; the margin is a model margin imposed by lumping the molecule into a single group.

2.8.4 Effective widths

The results are reported not as the parameters w_g and σ_g of (2.5) but as the effective width of the envelope of a group,

$$\sigma_{\text{eff}} = \sqrt{\sigma_g^2 + w_g^2/3}, \quad (2.24)$$

which is the standard deviation of the distribution (2.5), the convolution of a uniform distribution of half-width w_g , with variance $w_g^2/3$, with a Gaussian of variance σ_g^2 ; the two variances add. Equation (2.24) makes the profiles comparable with the Gaussian distributions with which the literature describes the same groups [34], since a Gaussian of standard deviation σ_{eff} has the same second moment, and it allows the width of a group to be restrained with a literature value without imposing w_g and σ_g separately. The widths are restrained through (2.24), with an absolute window for each group and a limit on the spread $\Delta\sigma_{\text{eff},g} = \max_i \sigma_{\text{eff},g,i} - \min_i \sigma_{\text{eff},g,i}$ across the eight profiles, which allows for differences in composition but excludes incompatible widths of the same group in different samples. The windows cover glycerol, sphingosine backbone, phosphate, unsaturated carbons and cholesterol; the terminal methyls, sensitive to mismatch and interdigitation, the choline, sensitive to hydration and orientation, and the interpenetration at the midplane are left out.

2.8.5 Phosphate–choline envelope

The relative position of phosphate and choline is measured between the outer edges of the two distributions,

$$\Delta_{\text{PC}}^{\text{out}} = (z_{\text{choline}} + \sigma_{\text{eff,choline}}) - (z_{\text{P}} + \sigma_{\text{eff,P}}), \quad (2.25)$$

positive when the choline protrudes toward the solvent beyond the phosphate and negative when the two envelopes are inverted; the choice of the edges rather than the centers makes the quantity sensitive to the widths and not only to the positions. The last restraint is the only one introduced in this work and concerns this quantity. The preceding conditions order the centers of phosphate and choline, but do not prevent the choline envelope, if wide, from remaining entirely contained within that of the phosphate. On each of the eight profiles the condition

$$\Delta_{\text{PC}}^{\text{out}} \geq -0.3 \text{ \AA}, \quad (2.26)$$

is imposed in threshold form with scale 1 Å; the margin allows the slight overlap that the uncertainty on the widths cannot rule out. The restraint was formulated after inspection of the solutions obtained without it (Sec. 2.10). For each of the eight profiles the effective widths of phosphate and choline, the positions of their outer edges $z_g + \sigma_{\text{eff},g}$ and $\Delta_{\text{PC}}^{\text{out}}$ are computed; these forty derived observables are the basis of the structural results.

2.9 Optimization

2.9.1 Objective function and minimizer

The function minimized is (2.18). Φ combines two heterogeneous quantities with weights chosen in this work, and two models that differ in the imposed constraints do not have the same objective function; the comparison between models is carried out on the two terms separately, with χ_{data}^2 as the cost on the data and the chemical plausibility of the solution, assessed on the structural quantities, as the benefit.

GENFIT [28] provides a random search within the parameter ranges, simulated annealing (SIMANN), the simplex method (SIMPLEX) and a quasi-Newton method (ZXMIN). In the production fits only SIMANN is active. Gradient-based methods were not employed because the penalties (2.17) have a discontinuous second derivative at the edges of the allowed region, where the solutions cluster; local refinement with SIMPLEX after SIMANN was tried and not adopted, because the improvement in Φ turned out to be orders of magnitude smaller than the spread across independent runs.

2.9.2 Annealing schedule

In the adopted implementation SIMANN proposes a move for each parameter, one at a time, within a step length adjusted so as to bring the acceptance rate toward 0.5, and accepts an increase $\Delta\Phi$ of the objective function with probability $e^{-\Delta\Phi/T_j}$. A cycle is a complete sweep over the parameters; the temperature

levels are the terms of a geometric progression

$$T_j = T_i \left(\frac{T_f}{T_i} \right)^{\frac{j-1}{N_T-1}}, \quad j = 1, \dots, N_T, \quad (2.27)$$

traversed once, without reinitialization between levels. The cost of one minimization is $1 + 215 N_c N_T$ evaluations of the objective function, with 215 parameters, N_c cycles per level and N_T levels; SIMANN has no stopping criterion and uses up the budget fixed at the outset.

Five $N_c \times N_T$ schedules were compared on ten independent runs each, on the model with the full set of constraints. For a fixed budget, adding temperature levels is more effective than adding cycles per level; between the cheapest and the most expensive schedule the median of χ_{data}^2 remains unchanged to within 0.2 % while the restraint term decreases by 93 %, and the χ_{data}^2 distributions of the five schedules overlap largely. The 20×100 schedule (430001 evaluations, about 800 s per run) was adopted because it is the only one that regularly reaches the region where the objective function is lowest.

The temperatures $T_i = 10^{-3}$ and $T_f = 10^{-8}$ are very low compared with the usual heuristic, because the fit starts near a reasonable solution, in a flat valley where the differences between neighboring configurations are small; an initial temperature of the order of the value of the objective function would make almost every proposal acceptable in the first levels and would carry the configuration into a different region of parameter space. Exploration is entrusted to the repetition of independent runs rather than to the temperature. The dependence on the starting point was checked with three sets of initial values, with distances between solutions not exceeding those between runs from the same

starting point, and the analysis retains a single starting point.

2.9.3 Repeated runs

Since SIMANN is not deterministic, each model configuration is minimized several times with different random-number sequences, ten per schedule in the comparison above and thirty for the final set on which the results are built. GENFIT derives the seed from the system clock at startup and writes it to the log; the seeds of the reported runs were extracted from the logs, verified to be distinct, and stored together with the configuration, ranges, weights and reduced curves. The set of runs is reproducible in distribution; a single run cannot be repeated with the program in the form employed.

2.10 Design of the analysis

A model with 215 parameters, ten families of constraints and a non-deterministic optimizer admits many variants in the windows, in the weights and in the ranges of variation, and the objective function is neither calibrated on the data nor comparable across models with different constraints. The analysis was therefore organized in successive stages (Table 2.6). In the two perturbation stages each variant differs from the reference by a single modification, so that its effect can be attributed to that modification. The optimization protocol was fixed first and not modified thereafter. The acceptance criterion was written before the fits of the first perturbation stage and revised once after them. The revision introduced the bands on the volume left to the water and admitted a data cost against a large chemical benefit.

Stage	Content	Runs
Reference	starting configuration, with the nine families of constraints taken from the literature or from the model	20
Single perturbations	nine configurations, each modifying one quantity of the reference. The ranges of the structure-factor parameters of all curves, from 2 to 5 % of their central value, and that of f_{ML} , the weight of the multilamellar against the unilamellar contribution, for curves 6 and 7, from 10 to 20 %; the weight of the headgroup-volume restraint and that of the choline-extension restraint, each lowered to 10^{-3} and raised to 10^{-2} from $3 \cdot 10^{-3}$; the upper bounds on the half-width (4.3 to 5.0 Å) and smoothing (2.3 to 3.0 Å) of cholesterol and the lower bound on its position (0.65 to 0.60 D_h), one at a time; the margins of the relations among the areas that involve the asymmetric leaflets, from 3 to 1 Å ²	9×10
Second-stage variants	seventeen configurations, again each modifying one quantity of the reference. The two sides of the area margin separately, at 0, 1 and 2 Å ² ; the weights of the order restraint (1 to 0.3) and of the effective-width restraint ($3 \cdot 10^{-2}$ to 10^{-2} and 10^{-1}); the upper bounds on the half-width (4.2 to 4.8 Å) and smoothing (3.4 to 4.0 Å) of the terminal methyls and the lower bound on the backbone position in profiles 6 ₂ and 7 ₂ (0.95 to 0.90 D_h); the ranges of the structure-factor parameters and of f_{ML} of curves 6 and 7, one family at a time; the range of the power-law amplitude κ of curve 7. In addition, the headgroup-volume and choline-extension weights raised together	5–10 each
Combination	four configurations, joining the two retained restraint sets (choline extension strengthened, alone or with the headgroup volumes) with either side of the area margin	4×10
Phosphate–choline envelope restraint	introduction of the restraint (2.26), suggested by the solutions of the previous stages, in the two combinations retained. The one with both restraint sets is the final configuration; the other, the conservative combination, serves as a check	5
Final set	independent runs of the final configuration	30
Identifiability	local analysis of the final configuration (Sec. 2.11)	–

Table 2.6: Sequence of the stages of the analysis. The reference measures the spread against which each modification is compared; each perturbation tests whether a quantity of the model is determined by the data or by a window, a weight or a range.

The stages up to the combination constitute the exploratory phase. The restraint on the phosphate and choline envelope was not planned at the outset. It arises from the observation that in part of the solutions, from the reference

onward, the choline envelope does not protrude beyond that of the phosphate. Since it was chosen after the solutions had been inspected, the results obtained with it are always reported alongside those of the configuration without it. The two subsequent stages leave the configuration unchanged.

The acceptance criterion is not a single score, which would conceal the trade-off between agreement with the data and chemical plausibility. For each variant the data cost, the chemical benefit, the robustness and the dependence on the constraints are kept separately. The data cost is χ_{data}^2 , overall and per experiment. The chemical benefit consists of the structural quantities that the modification is intended to improve, evaluated on the derived observables of Sec. 2.8.5. The robustness is the spread of the quantity under examination across the runs of the variant, compared with that of the reference. The dependence on the constraints comprises the value of each family of penalties, the frequency of parameters at a bound and the volume left to the hydration water. The latter is the minimum along z of $1 - \sum_{g < N} \phi_{g,\alpha}(z)$, which becomes negative where the groups fill more than the available volume. The bands refer to the median over the runs of a variant. A median overfilling up to 1 % is read against the reference, and between 1 and 2 % it is a warning. Beyond 2 %, or within the warning band with a worsening reproducible across runs, it excludes the variant. A modification is retained when its chemical benefit is reproducible across runs and its data cost is either not reproducible or accepted against the benefit, in which case both numbers are reported. A gain in χ_{data}^2 obtained with a parameter at a widened bound is not counted as a benefit, since the fit then has no optimum for that parameter within the explored range. The robustness criterion is necessary but not sufficient, because the spread mea-

sures the variability of the optimizer and not the support that the data give to the difference. The reference quantity for the chemical benefit is the one specific to each family in the perturbation stages, and $\Delta_{\text{PC}}^{\text{out}}$ from the stage in which the envelope restraint was introduced onward.

2.11 Identifiability

A model with 215 parameters may contain combinations that leave the calculated intensity unchanged, and the values returned for those combinations come from the constraints and from the ranges rather than from the data.

2.11.1 Jacobian and singular values

The basic tool is the Jacobian **J**, the matrix of the first derivatives of the residuals with respect to the parameters, with one row per point of the fit grid and one column per parameter. The derivatives are computed by finite differences, perturbing each parameter by a fixed fraction of its own range of variation, which makes the columns comparable; for parameters at a bound the derivative is taken with a one-sided difference, and columns not stable against the step are excluded. The Jacobian is built three times with the same columns, from the data residuals alone, weighted as in Sec. 2.7, from the restraint terms alone, and from the two stacked together; the conclusions on identifiability are read from the data matrix; the other two measure the stiffness added by the restraints.

A column of negligible norm belongs to a parameter that does not affect the intensity; two nearly parallel columns belong to two parameters for which

the data determine only one combination. A degeneracy may however involve more than two parameters, and it is identified by the singular value decomposition $\mathbf{J} = \mathbf{U} \text{diag}(s_1, s_2, \dots) \mathbf{V}^T$, in which the columns of $\mathbf{V}$ are orthogonal directions in parameter space and the singular value s_k measures how much the calculated curve changes along the k -th direction. The directions with singular value close to zero are not determined by the data; they are combinations of parameters, and the number of well-determined directions is not the number of well-determined parameters. The numerical rank is the number of directions whose singular value exceeds a fixed fraction of the largest one, and it is reported for several values of this cutoff.

2.11.2 Degenerate pairs

The degeneracy between half-width and smoothing involves two parameters per group and is measured on the two corresponding columns. Differentiating (2.24),

$$\sigma_g d\sigma_g + \frac{w_g}{3} dw_g = 0, \quad (2.28)$$

gives the direction along which the effective width remains unchanged to first order. The cosine between the two columns, which equals one when the two parameters act indistinguishably, is computed together with the ratio of the response of the model along the tangent direction to that along the orthogonal direction and the component of the tangent direction in the subspace of the directions discarded by the decomposition.

The derivatives describe the model in the neighborhood of the point at which they are evaluated. They are complemented by a check on the indepen-

dent runs. If the data determined the split of a degenerate pair, its two parameters would be reproduced with the same stability as their combination. Within a homogeneous group of solutions, the relative spreads of the two parameters and of the effective width, and the correlation between the two parameters, are therefore compared; a wide spread of the parameters, a narrow spread of the combination and a strong negative correlation indicate compensation. For degenerate pairs, only the effective width is reported.

2.12 Spread across runs

The structural quantities are accompanied by the spread of the values obtained from independent runs of the optimization, which share configuration, data and starting point and differ only in the random-number sequence (Sec. 2.9). The spread measures the variability of the minimum reached at fixed data and weighting; it is not a confidence interval, it does not incorporate the experimental noise, which would require a calibrated error model, and it does not measure the sensitivity to the weighting. Each value is reported with the number of runs on which it is computed and with the group of solutions to which it belongs.

The runs may fall into distinct groups of solutions, and a statistic over the whole set would mix different regimes of the optimizer. Within a group the mean and the standard deviation are reported; in comparisons between sets of runs the median is used. The groups are formed according to the minimum of $\Delta_{\text{PC}}^{\text{out}}$ over profiles 1 and 3, where the restraint (2.26) is at its threshold. The values are sorted, the cut is placed at the midpoint of the largest gap between

consecutive values, and the bipartition is accepted if each group contains at least two runs, the gap at which the cut is placed exceeds 0.10 Å, and the ratio of the largest gap to the second largest exceeds three. With thirty runs the procedure supports a bipartition but not a finer structure.

The curvature of the objective function at the minimum does not provide a usable uncertainty. Along the degenerate directions of Sec. 2.11 the curvature of the data term vanishes in the Gauss-Newton approximation and the normal matrix of the Jacobian is singular; the pseudoinverse truncated at the same cutoff used to define the numerical rank yields an uncertainty dominated by the weakest retained direction, and hence by the truncation cutoff, which is a numerical convention. The dependence on the cutoff is measured in Sec. 3.3.1.

2.13 Molecular dynamics

For the measured compositions, the simulations provide a description of the membrane obtained without using the SAXS curves; they share with the experiment the nominal composition of each leaflet and the temperature, and no quantity obtained from the fit enters the construction of the boxes or the protocol. The distributions against which the fit is compared are obtained from the trajectories by atom counting (Sec. 2.15).

2.13.1 Systems and box construction

The systems were prepared with the Membrane Builder module of CHARMM-GUI (version 3.7) [50–52], which also generated the input files for GROMACS [53]. The geometry is a planar bilayer in a periodic box, with the

normal along z . The force field is CHARMM36 for the lipids, with the parameters for phospholipids [31], sphingomyelin [54] and cholesterol [55], in the distribution that the interface designates as CHARMM36m [56], which differs only in the protein parameters; the water is TIP3P in the CHARMM version [32, 57]. The water layer has an initial thickness of 28 Å per side (22.5 Å for POPC) and contains 0.15 M KCl; the systems are electrically neutral even without ions.

The compositions are listed in Table 2.7. The size of each box is a compromise between representativeness and cost, with the constraint that the ratios among the species of each leaflet reproduce the nominal composition used in the fit; with 147–315 molecules per leaflet the boxes have a lateral size of 10–14 nm, four to eight times the size of the 72-molecule bilayers on which the lipid parameters were validated [31]. Egg sphingomyelin is represented as a mixture of four species in the proportions of the commercial product renormalized to 100, with palmitoyl (16:0, 86.7 %), stearoyl (18:0, 6.7 %), behenoyl (22:0, 3.3 %) and nervonoyl (24:1, 3.3 %) chains.

2.13.2 Equilibration and production

Each system was minimized (steep) and equilibrated with the standard six-step sequence of CHARMM-GUI [51, 53], three NVT and three NPT steps for 1.9 ns in total, with POSRES on the phosphorus atoms and on the hydroxyl oxygen of cholesterol and DIHRES on the chain dihedrals, with force constants reduced along the sequence from 1000 to 0 kJ mol⁻¹ nm⁻² and from 1000 to 0 kJ mol⁻¹ rad⁻². Production, 300 ns per system, was run with GROMACS 2024.3 [58, 59] on the Anna Tramontano cluster of the Department of Life

Composition		POPC	SM	Chol	Total	Water	KCl	Sample
symmetric bilayers (molecules per leaflet)								
POPC		147	–	–	147	13 164	34	1; inner of 6
POPC/EggSM 2:1		172	78/4/2/2	–	258	26 379	70	2
POPC/EggSM/Chol 1:1:1		86	77/5/2/2	86	258	22 732	61	5
POPC/EggSM 0.80:0.20		120	26/2/1/1	–	150	15 969	42	outer of 6
POPC/Chol 0.66:0.34		198	–	102	300	29 391	80	inner of 7
POPC/EggSM/Chol 0.63:0.20:0.17		189	52/4/2/2	51	300	29 793	80	outer of 7
asymmetric bilayers (molecules of the indicated leaflet)								
POPC	inner	287	–	–	287	32 145	86	6
POPC/EggSM 0.80:0.20	outer	240	52/4/2/2	–	300			
POPC/Chol 0.66:0.34	inner	208	–	107	315	29 560	80	7
POPC/EggSM/Chol 0.63:0.20:0.17	outer	178	49/4/2/2	48	283			

Table 2.7: Composition of the simulated systems. For the symmetric bilayers the counts are per leaflet; for the asymmetric ones each row is one leaflet. The SM column lists the four sphingomyelin species in the order 16:0/18:0/22:0/24:1. Water and KCl are the water molecules and the ion pairs of the whole box; the samples are those of Table 2.2.

and Environmental Sciences, Università Politecnica delle Marche, with GPU acceleration and the settings of Table 2.8. With the same reference pressure in the plane and along the normal, the surface tension imposed by the barostat is zero on average [60], and the lateral area relaxes to its equilibrium value. The production window is determined a posteriori with the criteria of Sec. 2.13.5, and the averages are taken over windows chosen within it (Sec. 3.4).

Parameter	Value	Reference
integrator, dt	md, 2 fs	
nsteps	$1.5 \cdot 10^8$ (300 ns)	
tcoupl, tau-t, ref-t	V-rescale, 1 ps, 310 K	[61]
pcoupl, pcoupltype	C-rescale, semiisotropic	[62]
tau-p, ref-p	5 ps, 1 bar in the plane and along the normal	
coulombtype, rcoulomb	PME, 1.2 nm	[63, 64]
vdw-modifier, rvdw-switch, rvdw	Force-switch, 1.0 nm, 1.2 nm	[65]
constraints, constraint-algorithm	h-bonds, LINCS	[66, 67]
water	rigid, SETTLE	[68]

Table 2.8: Production settings, common to all systems, with the parameter names of the .mdp file.

2.13.3 Asymmetric boxes

In a periodic box the two leaflets share the same lateral area; with the same number of lipids on both sides neither of them is at its own equilibrium area per lipid, and the resulting differential stress is an effect of the construction that is superimposed on the structural differences to be measured [35, 36]. The construction adopted assigns the two leaflets numbers of lipids that compensate for the ratio between the equilibrium areas,

$$\frac{N_{\text{in}}}{N_{\text{out}}} = \frac{A_{\text{out}}}{A_{\text{in}}}, \quad (2.29)$$

the area-based method in the classification of Park and co-workers [36]. The equilibrium areas were measured by simulating, for each leaflet composition, a dedicated symmetric bilayer with the same protocol (the systems listed in Table 2.7 as inner or outer leaflet of samples 6 and 7), as the average over the production window with the standard error from the number of independent samples. Among the pairs of integers with a total between about 590 and 610 lipids, those whose deviation from the ratio (2.29) is smaller than the uncertainty of the areas, which at these sizes corresponds to about one lipid, are equivalent, and among them the pair that best reproduces the nominal mole fractions was chosen. Since in an asymmetric bilayer the two leaflets constrain each other and (2.29) does not guarantee zero tension per leaflet [35], the construction is checked a posteriori for compatibility, by comparing the mean lateral area of each asymmetric bilayer with the area predicted from the numbers of lipids and the reference areas.

2.13.4 Simulated compositions

Of the seven measured compositions five have a simulation of the same composition, and for the two asymmetric systems both the compositions of the individual leaflets and the complete bilayer were simulated. Cases 3 and 4, with milk and brain sphingomyelin, are left out, because the CHARMM36 parameters of sphingomyelin were validated on the palmitoyl and stearyl species [54] and the longer-chain species would introduce into the comparison a parametrization uncertainty extraneous to the SAXS model. Egg sphingomyelin consists for more than 90 % of the two validated species and is the only one used in the asymmetric preparations. In the fit it is approximated by the palmitoyl species alone; in the simulations it is the four-species mixture.

2.13.5 Convergence checks

Each trajectory is checked for the expected number of frames, the regularity of the time steps and the absence of constraint-solver warnings in the logs. From the energy file the lateral area, the density, the mechanical surface tension $\gamma = L_z [P_{zz} - (P_{xx} + P_{yy})/2]/2$, with L_z the box dimension along the normal, and the in-plane anisotropy are monitored; a zero mean tension verifies the semi-isotropic coupling.

The production window is determined on the area per lipid, the lateral area divided by the number of molecules per leaflet, cholesterol included, by adopting the later of two times. The first is the end of the initial contraction, the first 2-ns block whose mean enters the band of two standard deviations of the block means of the last quarter of the series. The second is the time that maximizes the number of uncorrelated samples in the remaining part of the series, ac-

cording to the criterion of Chodera [69]. The number of independent samples is the number of points in the window divided by the statistical inefficiency $1 + 2\tau_{\text{int}}/\Delta t$, with τ_{int} the integrated autocorrelation time and Δt the sampling interval, and the standard error of a mean is the standard deviation divided by the square root of that number. The surface tension is averaged over the whole trajectory, with the standard error from four blocks. Stationarity is verified over the last two thirds of the trajectory, 100–300 ns. A linear drift of the area above 1 % of the mean is flagged, unless it is smaller than two standard errors, and so is a difference between the means of the two halves larger than two standard errors and 0.5 %.

In the systems with cholesterol it is checked that the constructed asymmetry is preserved. The two leaflets are defined by the phosphorus atoms of the phospholipids, the bilayer center is recomputed at each frame and each cholesterol molecule is tracked through its hydroxyl oxygen; a crossing is counted when the molecule moves beyond a dead band around the center and remains there for a minimum dwell time. The detector was verified on synthetic trajectories built for this purpose. For the profiles along the normal the PBC are removed and the membrane is centered in the box with `gmx trjconv`.

2.14 Closure test

The experimental fit does not distinguish a disagreement due to the functional form, which does not represent the structure, from one due to the data, which do not determine the parameters. On data for which the structure is known the two causes become separately measurable. The test is carried out on the two

asymmetric bilayers (Sec. 2.13.3), whose transverse structure is the object of the ASDP model.

2.14.1 Synthetic curves

Each curve is computed from the electron density that is recombined from the volume-fraction distributions of the fine partition (Sec. 2.15.4), and not from their representation, so that the error of the functional form remains in the test. In the fine partition the pieces of cholesterol have different electron densities (Sec. 3.5.1), while the model describes cholesterol as a single group. The curve has 100 points uniformly spaced in logarithm between 0.005 and $0.5 \text{ \AA}^{-1}$ and a constant background. Three levels of counting noise are used, with a relative error of 1, 0.3 and 0.1 % on the first point. GENFIT receives the uncertainties written with the curve and not those of the error law (2.15). These are the uncertainties of the noise drawn into the curve, so χ_{data}^2 , the mean over the 100 points of a single curve, is close to one for a model that reproduces the curve within the noise.

2.14.2 Model and true values

Each curve is fitted alone with the ASDP model of a single bilayer, without other populations or structure factor and without the restraints of Sec. 2.8. Only the background, fixed at its value in the synthetic curve, and the number of each group per molecular unit, taken from the simulated composition, are held fixed. Volumes, areas, envelope parameters and the positions and widths of the groups are free. The check of GENFIT that every volume fraction lies between zero and one remains active and adds its penalty to the objective function.

The true values are those of the simulation and of its representation with erf slabs (Sec. 3.6.1). Volumes are the simulated ones (Sec. 3.5.2) and areas those of the representation (Sec. 2.15.3), and the positions enter as $f_g = |z_g|/D_{h,\alpha}$. The smoothing $\sigma_{h,\alpha}$ of the envelope (2.8) comes from a fit of the envelope to the simulated hydrophobic distribution with $D_{h,\alpha}$ fixed. The interpenetration σ_{mid} does not enter the intensity, because its two terms in (2.11) cancel when the methylene of the two monolayers has the same SLD. It acts only through the check on the volume fractions. Its true value is the one that best describes the simulated monolayers, raised until the true profile satisfies the check. The terminal methyls are left out of the comparison, because their two slabs overlap at the midplane and the representation describes them only through their sum (Sec. 2.15.3).

For sample 7 the test is repeated with cholesterol divided into two groups, the ring (151 electrons, $375.4 \text{ \AA}^3$) and the chain (65 electrons, $278.3 \text{ \AA}^3$), each with the number per molecular unit of cholesterol. Their volumes add up to that of the single group, so the hydrophobic thickness and the envelope are unchanged. The true positions and widths of the two pieces come from one erf slab per monolayer fitted to their simulated distributions.

2.14.3 Protocol

The model is first evaluated at the true parameters, without minimization. The resulting χ_{data}^2 measures the error of the functional form against the noise of each curve. In the fits each parameter is bounded within $\pm 20 \%$ of its true value, and five starting points are drawn uniformly within the bounds. The same fits were also run with bounds at $\pm 10 \%$ and $\pm 5 \%$. The minimizer is

SIMANN with 20 cycles, 100 temperature levels and $T_f = 10^{-8}$, as in Sec. 2.9. The initial temperature is instead $T_i = 10$, because with $T_i = 10^{-3}$ about a third of the runs ends in a minimum far from the true parameters (Appendix C.1). A run is converged when its objective function is within 1.5 times the lowest of the runs. The best fit, the converged run with the lowest objective function, is reported together with the range of the converged runs.

A second test isolates the functional form. All parameters are held at their true values except the half-widths and smoothings of the terminal methyls and of cholesterol (or of its ring and chain), the positions of cholesterol and σ_{mid} , with bounds at ± 20 and ± 50 %.

The recovered quantities are the positions $|z_g|$, the half-widths and smoothings with the effective widths (2.24), the hydrophobic thicknesses, the areas and the volumes. For the transverse asymmetry they are the differences between the outer and the inner leaflet in area and in the positions of phosphate and choline (Sec. 3.7.2).

2.15 The MDtoSDP program

The comparison between the distributions obtained from the simulations and those obtained from the fit requires that the former be constructed in the same terms as the model of Sec. 2.6, with the same chemical groups, the same quantity (the volume fraction) and the same normalization. SIMtoEXP [33] computes the per-group profiles and the form factors from the trajectories, with the group volumes taken from the literature or declared by the user. The program written for this work, MDtoSDP, adds the partition into groups in the conven-

tion of the model, the measurement of the group volumes on the simulation itself with a diagnostic of how well they are determined, the representation of the profiles in the functional form of the fit and the generation of synthetic curves for the closure test.

2.15.1 Stages and products

The program is organized in three stages (Figure 2.6). The first assigns each atom to a chemical group, computes the mass-density profiles along the normal, derives the group volumes and converts the densities into volume fractions; the second represents the distributions with the functional form (2.5); the third generates a scattering curve from the electron density. Each stage reads the products of the previous one and writes its own, so that every intermediate product can be inspected. The two downstream stages both read the products of the first and exchange nothing with each other, because the synthetic curve must be generated from the measured distributions and not from their representation; generating the synthetic data with the same functional form with which they are refitted would leave the error of the functional form out of the test.

The link between the atoms of the simulation and the groups of the model is declared in two configuration files read by all three stages. The chemical-group file lists, for each species, which atom names form choline, phosphate, glycerol or sphingosine backbone, methylenes, unsaturated carbons and terminal methyls, and for cholesterol distinguishes the sterol ring system from the iso-octyl chain; the names are those of the CHARMM36m topology. The system file declares which chemical groups are aggregated into a single group of the fit, which whole molecules constrain the sum of the volumes of their groups,

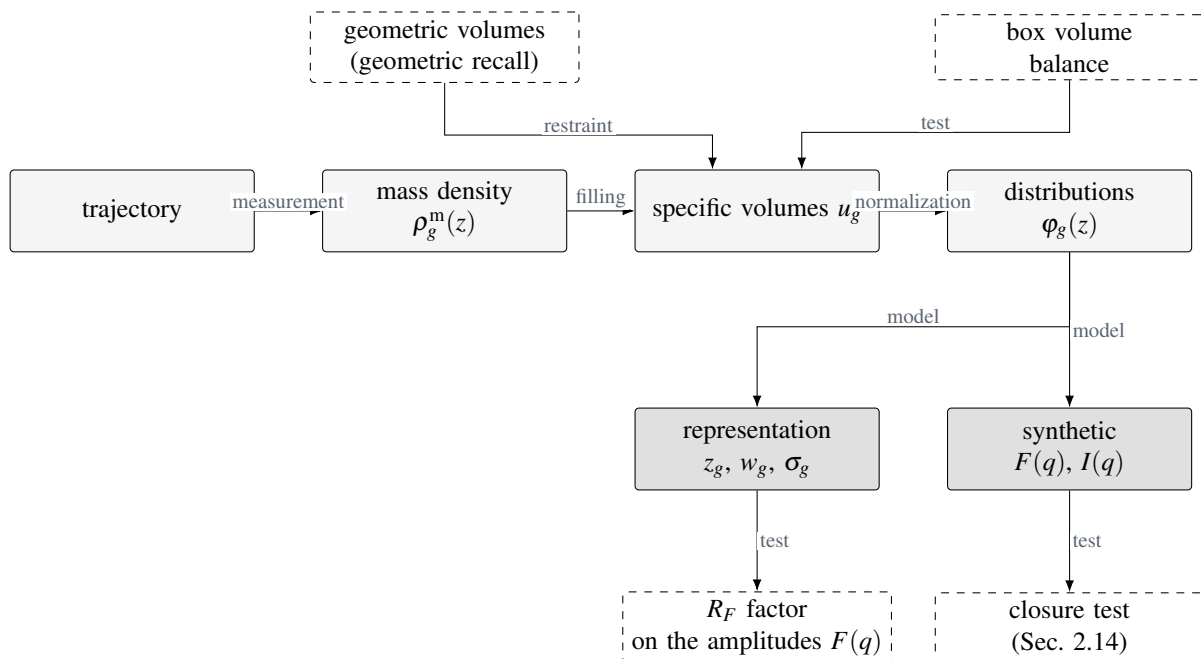


Figure 2.6: The chain from trajectory to synthetic curve. Light boxes are measurements on the trajectory, dark boxes models, dashed boxes constraints or independent tests; the labels on the arrows indicate the nature of each step.

which groups are anchored to the geometric measurement and with what tolerance, which window of the trajectory constitutes the production and with what step the normal is sampled. Before each run the program verifies that every atom not declared as extraneous (the ions) belongs to a group, and stops if the coverage is not complete, because an unassigned mass would disappear from the denominator of the normalization without warning.

The first stage delivers to the other two (Table 2.9) the volume-fraction distributions of all groups, tabulated on the same grid in z with their uncertainty, a condition for (2.3) to be defined point by point, and the volume record, which for each group reports the volume, the chemical groups merged into it, the uncertainty and the source that determined it. The areas per molecule of each monolayer and the numbers of units n_g expected from the composition are computed on input to the representation stage.

Stage	Receives	Produces
Partition and volumes	trajectory, structure, group declaration	$\varphi_g(z)$ on a common grid (nm); group volumes ($\text{\AA}^3$) with their source
Representation	$\varphi_g(z)$, volumes, areas ($\text{\AA}$)	z_g , w_g , σ_g per group and mono-layer
Generation	$\varphi_g(z)$ or electron density	q , $I(q)$, $\sigma(q)$

Table 2.9: Inputs and products of each stage, with units.

2.15.2 Group volumes

The model requires a volume for each group, and the published values are measured on pure-lipid bilayers and are reliable for the polar headgroup and for the chains as a whole, not for the individual groups into which the model decomposes them. The compositions contain sphingomyelin and cholesterol in high fractions, and adopting the tabulated values would introduce into the reference profiles the structure of a system different from the simulated one. The volumes are therefore measured on the simulation.

The starting point is the mass-density profile $\rho_g^m(z)$ of each group, computed by GROMACS in each slice of the box. With u_g the specific volume of the group, the volume of a chemical unit is $v_g = \mu_g u_g$, with μ_g the mass of the unit, and the product $\tilde{\varphi}_g(z) = \rho_g^m(z) u_g$ is the fraction of the slice volume occupied by the group. The filling factor

$$\varphi_{\text{tot}}(z) = \sum_g \rho_g^m(z) u_g \quad (2.30)$$

is the fraction of the volume of the slice at z covered by the volumes assigned to the groups, and equals one at every z if and only if the volumes fill space. The distributions delivered downstream are the normalized fractions $\varphi_g(z) = \tilde{\varphi}_g(z) / \varphi_{\text{tot}}(z)$, whose sum is one by construction; the deviation of $\varphi_{\text{tot}}(z)$ from

one measures how well the volumes fill space, whereas the normalized sum does not.

One measurement follows from the filling condition. Imposing $\phi_{\text{tot}}(z_k) = 1$ at all points z_k of the grid yields the overdetermined linear system

$$\sum_g \rho_g^{\text{m}}(z_k) u_g = 1, \quad k = 1, \dots, n_z, \quad (2.31)$$

in the specific volumes, solved by least squares; this is the procedure of Petracche and co-workers [70] adopted in the SDP model [34]. The system matrix $\mathbf{X}$, with elements $X_{kg} = \rho_g^{\text{m}}(z_k)$, is the Jacobian of the residuals with respect to the volumes, constant because the model is linear. Two groups distributed similarly along the normal contribute to $\mathbf{X}$ with nearly parallel columns, and can exchange volume without any change in the residual; the system determines their sum well and the two addends poorly. The degeneracy is measured by the condition number, the ratio of the largest to the smallest singular value of the system matrix, and by the correlations between the estimated volumes.

The geometric measurement partitions the space of the box among the atoms, and the volume of a group is the sum of the volumes of the cells of its atoms. It does not depend on the shape of the profiles and is complementary to the filling measurement, whose degeneracy it does not share; the two are not statistically independent, because they come from the same trajectory. The ordinary Voronoi tessellation ignores atomic sizes and is reliable on whole molecules but not on subgroups; for these the radical tessellation, or power diagram [71], is used; it assigns each point to the atom that minimizes

$$\pi_i(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_i\|^2 - r_i^2, \quad (2.32)$$

with r_i the van der Waals radius of atom i , taken from Bondi [72] (1.20 Å for hydrogen; for the ions the ionic radii are used). The result depends on the radii, which are an input and not a measurement, and the tessellation assigns the bulky interfacial groups part of the surrounding free volume. With all radii set equal, (2.32) reduces to the ordinary tessellation, and the sampled volumes must reproduce the exact ones computed with freud [73]; this is the internal verification of the code (Sec. 3.5.2).

The two measurements are combined in a single least-squares problem in the specific volumes u_g , whose minimized function is the sum of three terms:

$$\begin{aligned}\Phi_V(\mathbf{u}) = & \sum_{k=1}^{n_z} \left[\sum_g \rho_g^m(z_k) u_g - 1 \right]^2 \\ & + \sum_{\ell} \left[\sum_g \frac{M_{g,\ell} u_g}{V_{\ell}} - 1 \right]^2 \\ & + \sum_{g \in \mathcal{T}} \omega_g^2 (u_g - u_g^{\text{geom}})^2, \quad \omega_g = \frac{\sigma_{\varphi}}{\tau_g u_g^{\text{geom}}}.\end{aligned}\tag{2.33}$$

■ **Filling**: the rows (2.31), one per grid point z_k , with unit weight.

■ **Whole molecules**: for each declared molecule ℓ , the sum of the volumes of its groups must reproduce the volume V_{ℓ} of the whole molecule from the tessellation; $M_{g,\ell}$ is the mass of group g in the molecule, and the division by V_{ℓ} makes the residual comparable with that of a filling row.

■ **Geometric recall**: for each group of the set $\mathcal{T}$ with a declared relative tolerance τ_g , the volume is pulled toward the value u_g^{geom} of the radical tessellation, with weight ω_g .

The scale σ_{φ} is the root-mean-square deviation of $\varphi_{\text{tot}}(z) - 1$ over the filling rows, computed in a first solution of the system without the geometric terms (recall and imposed values), which also provides the value of each volume

determined by the profile alone; the complete system is then solved a second time. With this weight a deviation of one tolerance from the geometric value contributes to the minimized function as much as a single point of the profile, and the recall acts only along the directions in which the profile is indifferent, whereas along the directions determined by the profile its contribution is negligible; the final deviation from the geometric value, in units of tolerance, is reported with the volume. A zero tolerance τ_g imposes the geometric value, an undeclared tolerance leaves the group to the profile alone, and intermediate values of τ_g regularize the degenerate directions. The decision is made group by group. The geometric measurement is excluded for the choline, and for the terminal methyls in the systems with cholesterol and in the asymmetric bilayer of sample 6, for which it is distorted, imposed for the groups that the profile does not determine, because they are too poorly represented or degenerate with a neighboring group, and left as a recall with uniform tolerance $\tau_g = 0.10$ for the remaining ones; the volume of the water far from the polar region is obtained from the plateau of its own profile. The value of τ_g was calibrated by a scan on a development system, a bilayer of DPPC, POPC, POPE, POPS, palmitoyl sphingomyelin and cholesterol, as the tightest value for which the deviation of ϕ_{tot} from one remains unchanged from the solution without recall.

Each of the two systems is solved by singular value decomposition, which yields the condition number, the correlations between the volumes and the number of effective degrees of freedom. The latter is the trace of the projection matrix restricted to the filling rows and counts how many free volumes the profile, rather than the recall, determines. The normalization of the fractions alters the distributions by a relative amount of the order of σ_ϕ , reported for

each system in Table B.2.

The statistical uncertainty is measured by dividing the window over which the distributions are averaged (Sec. 3.4) into contiguous blocks of equal length, five for the symmetric bilayers and six for the asymmetric ones, of 20–50 ns, four to fourteen times the autocorrelation time of the lateral area (Sec. 2.13.5), and repeating the whole chain on each; the standard deviation across blocks provides the uncertainty on the volumes and the band on the profiles. A group determined by the geometric recall has by construction an almost zero spread across blocks, and the model error remains invisible to the resampling; the known systematic sources are the choice of the atomic radii, which in the development system shifts the phosphate by 7 % and the polar groups by 3 % when the hydrogen radius is changed from 1.20 to 1.09 Å [74], and the difference in definition between geometric and filling volume, which for water amounts to 2–3 %. The uncertainties reported on the volumes are the statistical ones only. The stage stops if the condition number, the deviation of φ_{tot} from one, the box volume balance or the deviation from the literature volumes exceed fixed thresholds; the volume balance, the sum of the volumes of all molecules compared with the volume of the box, is the only criterion independent of the least-squares solution.

The volumes are determined on a partition in which each group is chemically homogeneous, finer than the groups of the model where these are not. The ring system of cholesterol, with its two methyls and the hydroxyl, is a group of its own, and the atoms of its chain belong to the methylene and terminal-methyl groups and to a group of sp^3 carbons. The sp^2 carbons of the acyl chains and of the sphingosine are separate groups. Since volume fractions are additive,

the distribution of each group of the model is the sum of those of its members, and its volume is the mean of their volumes weighted by the numbers of units. The alternative, solving (2.33) directly on the groups of the model, requires a single specific volume for groups whose density changes along the normal, and was discarded (Sec. 3.5.1).

2.15.3 Profile representation

The second stage represents the measured distributions with (2.5), one per monolayer and per group, and derives from them position, half-width and smoothing; the simulations thus yield structural parameters in the same quantities as the experimental fit, and the residual of the representation measures how well the functional form reproduces the observed distributions.

In (2.7) the prefactor $n_{g,\alpha}v_{g,\alpha}/(2A_\alpha w_{g,\alpha})$ of each distribution is fixed by the number of units, the volume and the area per molecule. In the representation none of the three is unknown, since the number of units n_g is obtained by counting the molecules in the structure, the volume comes from the first stage and the area appears in (2.7) only through the ratio $n_{g,\alpha}/A_\alpha$ and is fixed at the value obtained by integrating the profiles, with the fraction of lipids of each leaflet measured rather than assumed. The integral of each distribution is therefore predicted without free parameters, and position, half-width and smoothing remain free.

The free parameters are fitted as logarithms if positive and as logits if proportions, so that the inner problem is unconstrained. The position is obtained from a proportion $p \in [0, 1]$ interpolated between two extremes that depend on

the half-width,

$$z_{g,\alpha} = z_{\min}(w_{g,\alpha}) + [z_{\max}(w_{g,\alpha}) - z_{\min}(w_{g,\alpha})] p, \quad (2.34)$$

so that the slab remains contained in the box by construction. On the asymmetric bilayers the inner edge of the terminal-methyl slab is placed at the bilayer center and the position ceases to be a parameter, because the methyls of the two monolayers meet at the center by construction and two free positions, with leaflets of different thickness, could overlap in a physically meaningless way. The minimization is carried out with a trust-region method for nonlinear least squares [75], with unweighted residuals, because the statistical band on the profiles is almost zero for the groups anchored to the geometric measurement and using it as a weight would give those groups an arbitrary importance.

The representation is assessed on the root-mean-square shape residual, computed after matching the integrals; when the two slabs of a group overlap along the normal, as for the terminal methyls, the comparison is carried out on the sum. Its imprint on a curve is measured by replacing, in the simulated electron density, each distribution with its slabs at the electron density of the group, with the rest of the density unchanged. The amplitudes $F(q)$ of the two densities are computed on 200 points between 0.05 and 0.6 Å⁻¹ and compared through the factor

$$R_F = \frac{\sum_q ||F_{\text{meas}}(q)| - c|F_{\text{rep}}(q)||}{\sum_q |F_{\text{meas}}(q)|}, \quad c = \frac{\sum_q |F_{\text{meas}}| |F_{\text{rep}}|}{\sum_q |F_{\text{rep}}|^2}, \quad (2.35)$$

where the scale c is optimized so that R_F measures the shape; the ratio of sums avoids the divergence near the zeros of the amplitude, and with it the positions

of the minima of $|F(q)|^2$ are compared. R_F is zero for amplitudes of the same shape and gives the summed deviation of the moduli as a fraction of their sum, so that 0.05 corresponds to a mean deviation of about 5 %. The shape residual measures the part of each profile that cannot be represented by (2.5), and the comparison on the amplitudes measures whether that part is visible in a curve. Half-width and smoothing are compared with the experimental fit through the effective width (2.24).

2.15.4 Synthetic curves

The third stage transforms an electron-density profile into a scattering curve with counting noise, readable by GENFIT, with the expressions of Sec. 2.6. The electron density $\rho_e(z)$ is built with the number of electrons of each atom or group, without atomic form factors. The direct source derives it from the whole trajectory, water and ions included, without assumptions on the groups or on the volumes, and is the check independent of the partition into groups; the recombined source weights the measured volume-fraction distributions by the electron content of each group, and provides the curve implied by the measured volumes. The two differ in the ions, excluded from the partition into groups, whose contribution to the density of the water layer at 0.15 M KCl is of the order of one percent.

The origin of z is placed at the bilayer center, identified as the centroid of the terminal-methyl profile above half its height; the Luzzati thickness in (2.12) is the total volume of the lipids divided by the in-plane area of the box.

The amplitude

$$F(q) = \int [\rho_e(z) - \rho_0] e^{iqz} dz \quad (2.36)$$

is computed by quadrature at the requested values of q and not with a discrete transform, which would return it only at the multiples of $2\pi/L_z$, imposed by the amount of water in the box. The solvent density ρ_0 is estimated from the edge regions of the profile, and the residual contrast at the edges, which would produce truncation artifacts, is quantified. A translation of the origin multiplies (2.36) by a phase factor and leaves $|F|^2$ unchanged; the centering is needed for the asymmetry indicators, the maximum and the root-mean-square value of $|\text{Im}F(q)|$ divided by the maximum of $|F(q)|$, which vanish for a symmetric profile (Sec. 3.7.2). The quadrature was compared with the closed form (2.6) on a profile composed of double error functions.

The grid in q is uniform in logarithm, specified by its endpoints and number of points. A constant background is added to the intensity, fixed as a fraction of the intensity in the last decade of q and calibrated on the background obtained from the fit of the experimental data, because the computed intensity has exact zeros that a measurement does not contain; the lamellar structure factor, in the paracrystalline and Caillé forms, is available and can be enabled. The noise is that of counting statistics. The intensity is converted into counts with a scale fixed by the prescribed relative error on the first point, each point is replaced by a Poisson-sampled value and converted back, so that the curve remains non-negative, and the uncertainty written next to each point is $\sigma(q) = a\sqrt{I(q)}$, with a the conversion scale. The libraries employed are NumPy [76] and SciPy.

Chapter 3

Results

3.1 From the reference to the final configuration

The stages of the analysis, the modifications tested in each and the numbers of runs are those of Table 2.6, and the acceptance criterion is that of Sec. 2.10. Each variant is compared with the reference, unless another configuration is named, through the difference between the medians of the two sets of runs. Since the scale of χ_{data}^2 is conventional (Sec. 2.7), the difference is expressed as a percentage of the reference median, and the same holds for the penalty P_k of a restraint, the unweighted term of (2.18). A difference is taken as reproducible when the central 95 % of 5000 bootstrap resamplings of the two sets [77] has a single sign. The resampling measures the variability of the optimizer, not the support that the data give to the difference (Sec. 2.12). Figure 3.1 shows the change in χ_{data}^2 of every variant, and the paragraphs below say what each group of variants shows and why it was retained or discarded.

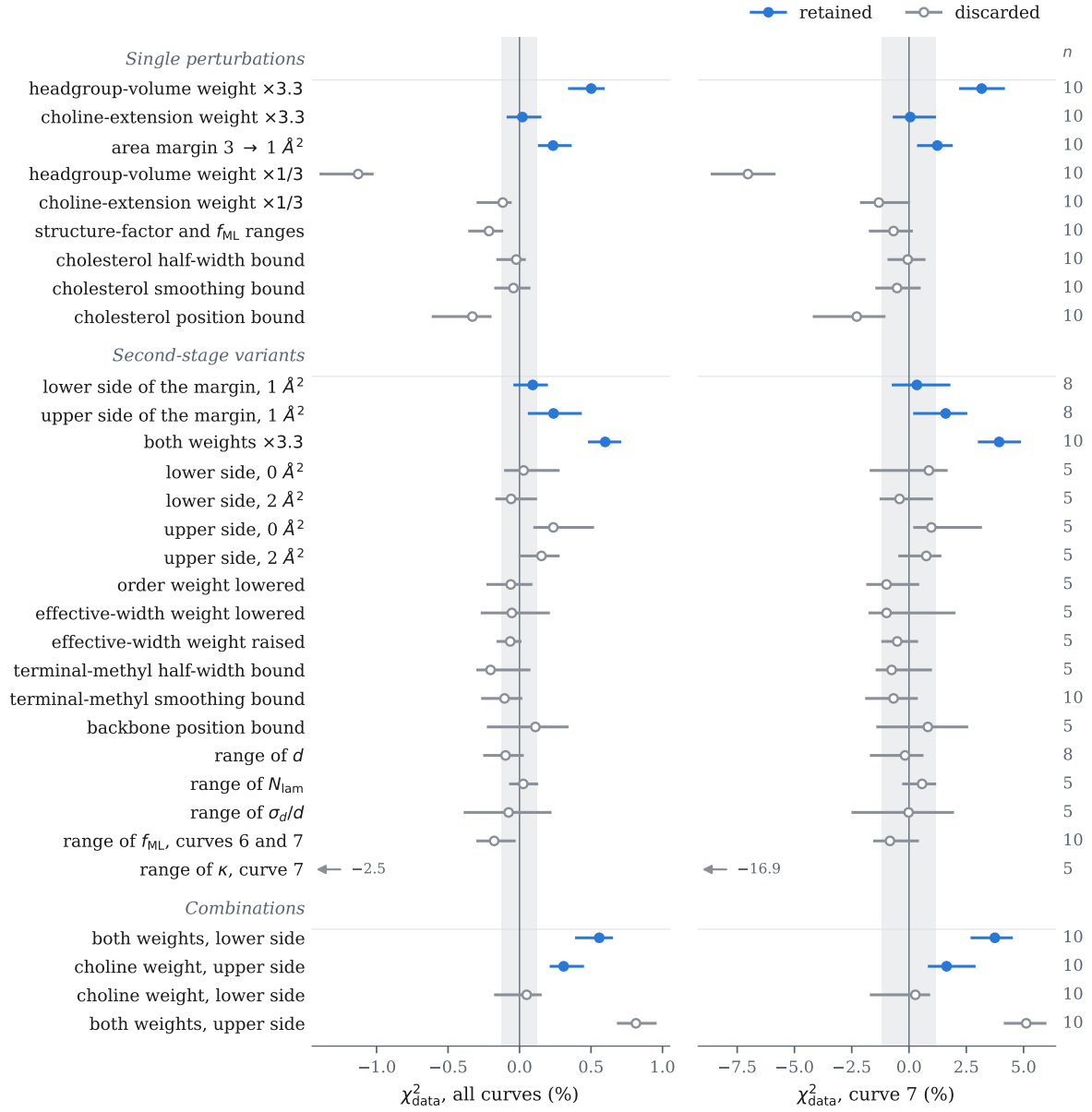


Figure 3.1: Change in χ^2_{data} of each variant with respect to the reference, over all curves and on curve 7, as a percentage of the reference median. Points are the differences between the medians; bars are the central 95 % of 5000 bootstrap resamplings of the two sets of runs, a measure of the variability of the optimizer and not a confidence interval. The shaded band is the relative median absolute deviation of the reference. Arrows mark values outside the axis, and n is the number of runs of each variant; the reference has 20. Curve 7 is shown separately because the cost of the headgroup-volume restraint and the gain of the low- q background fall almost entirely on it.

3.1.1 Reference configuration

The reference reproduces the curves, but not its own parameters, from one run to another. Across 20 runs χ^2_{data} varies by 0.12 % overall and by at most 1.3 %

on a single curve. Over the same runs 43 parameters spread over more than a quarter of their range, and about 60 of the 215 end within 0.1 % of a bound in a typical run, 31 of them in every run. The reference is therefore a distribution of solutions with many parameters held by their bounds, and each variant is compared with that distribution. In 19 of the 20 runs it also leaves a negative volume to the hydration water in profile 3.

3.1.2 Single perturbations

Weights of the restraints. Raising the weights of the headgroup-volume and choline-extension restraints from $3 \cdot 10^{-3}$ to 10^{-2} lowers their penalties by 94 and 91 % at a small cost in χ_{data}^2 , and both modifications were retained. The headgroup volumes move from 347 and 292 Å³ to 343 and 285 Å³, 1–2 Å³ above their windows, at a cost of 0.50 %, and 3.2 % on curve 7. In profile 7 the choline center exceeds its limit (2.22) by 0.09 instead of 0.31 Å, at no reproducible cost. Lowering either weight to 10^{-3} was discarded, because the penalties grow more than sixfold. With the lower weight the headgroup volumes reach 356 and 305 Å³ and χ_{data}^2 improves by 1.1 %, and by 7.0 % on curve 7. The curves, curve 7 above all, favor larger headgroups than the windows allow, and the volumes of the reference result from the balance between the two.

Area margin. Reducing from 3 to 1 Å² the margins of the relations (2.19) that involve the asymmetric leaflets costs 0.23 % in χ_{data}^2 , and the areas move until the relations are met. The modification was retained. The reference fills two of these relations, $A^{(6_2)} \geq A^{(2)} - s$ and $A^{(7_2)} \leq A^{(6_2)} + s$, with s the margin,

called below the lower and the upper side.

Wider ranges. Widening a range improves χ_{data}^2 only by moving the parameter to its new bound, and every such variant was discarded. Widening the ranges of the structure factor and of f_{ML} of curves 6 and 7 lowers χ_{data}^2 by 0.21 %, but f_{ML} ends at its new bound in all runs. The gain comes from the stacking of the lamellae rather than from the membrane. In the reference the position, half-width and smoothing of cholesterol are at their bounds in profiles 4, 7 and 7₂, and they compensate one another. Widening one of these bounds moves the parameter to the new bound and pushes another quantity out of its allowed region. The volume left to the water in profile 3 falls to -1.5 %, the penalty of the effective-width restraint grows 6.3-fold, or the cholesterol edge (2.23) of profile 7 reaches -2 Å, the limit of its window. The curves therefore ask for values outside the explored ranges, and the structure reported below is conditional on those ranges.

3.1.3 Second-stage variants

Sides of the area margin. A margin of 1 Å² was retained on each side. On the lower side the relation is met mainly by reducing $A^{(2)}$, the area of the symmetric POPC/egg SM sample, by 1.1 Å², at no reproducible cost. On the upper side it costs 0.24 %. Margins of 0 and 2 Å² add no reproducible gain and were discarded.

Both weights raised together. Raising the two weights together lowers both penalties as when they are raised separately, at the cost of the headgroup-

volume weight alone, 0.60 % in χ_{data}^2 and 3.9 % on curve 7. The modification was retained.

Variants without a reproducible gain. Lowering the weight of the order restraint from 1 to 0.3, widening the bounds of the terminal methyls and of the sphingosine backbone, and setting the weight of the effective-width restraint to 10^{-2} or 10^{-1} do not change χ_{data}^2 reproducibly. The widened bounds are reached again, and the cholesterol widths stay at their bounds whatever the weight. Of the structure-factor and f_{ML} ranges widened one family at a time, only that of f_{ML} gives a reproducible gain, 0.18 %, again with f_{ML} at its new bound. These variants were discarded.

Low- q background. Widening the range of the amplitude κ of the power-law term of curve 7 lowers χ_{data}^2 by 2.5 %, and by 17 % on curve 7, with κ again at its new bound. The gain is not confined to the background, since $A^{(1)}$ and $A^{(7)}$ change by more than $1 \text{ \AA}^2$ and the volume left to the water in profile 3 falls to $-1.6 \text{ \%}$. A background term should not reshape the membrane, so the modification was discarded, and the structure reported below is conditional on the background adopted.

3.1.4 Combinations

The retained modifications were joined in four combinations, each pairing the choline-extension weight, alone or with the headgroup-volume weight, with one side of the area margin. Two were retained. Both weights with the lower side form the combined configuration, with headgroup volumes of 343 and

285 Å³ at a cost of 0.56 %, and 3.7 % on curve 7. The choline-extension weight with the upper side leaves the headgroup volumes at those of the reference at a cost of 0.31 %, and 1.6 % on curve 7, and was kept as the conservative combination. The other two were discarded, one because the volume left to the water in profile 3 falls below -2 % in 4 runs of 10, the other because it costs 0.21 % more than the two weights alone.

3.1.5 Phosphate–choline envelope restraint

In 49 of the 160 profiles of the reference the outer edge of the phosphate, $z_P + \sigma_{\text{eff},P}$, lies beyond that of the choline, so that Δ_{PC}^{out} of (2.25) is negative, mostly in profiles 1 and 3. The inversion is present in every configuration tested, with a depth that depends on the other restraints. In the combined configuration it affects 24 of 80 profiles, 19 of them by more than 0.5 Å (Figure 3.2a).

The order restraint (2.20) does not prevent it. In profile 1 of the combined configuration the two centers stay 0.50 Å apart, on the threshold of that restraint, while the effective width of the phosphate, 2.4 Å, is twice that of the choline, 1.2 Å. The width of the choline is not restrained (Sec. 2.8.4), and the inversion comes from a narrow choline and a wide phosphate with ordered centers. We therefore introduced a restraint on the outer edges instead of the centers (2.26) as a hypothesis, to test whether the fit satisfies it and at what cost in χ_{data}^2 .

The restraint was added to the two retained combinations. The combined configuration with the restraint is the final configuration, chosen because it keeps the headgroup volumes closer to the windows at a higher cost on curve 7; its results are given in Sec. 3.2. Its runs fall into two groups of solutions, shown

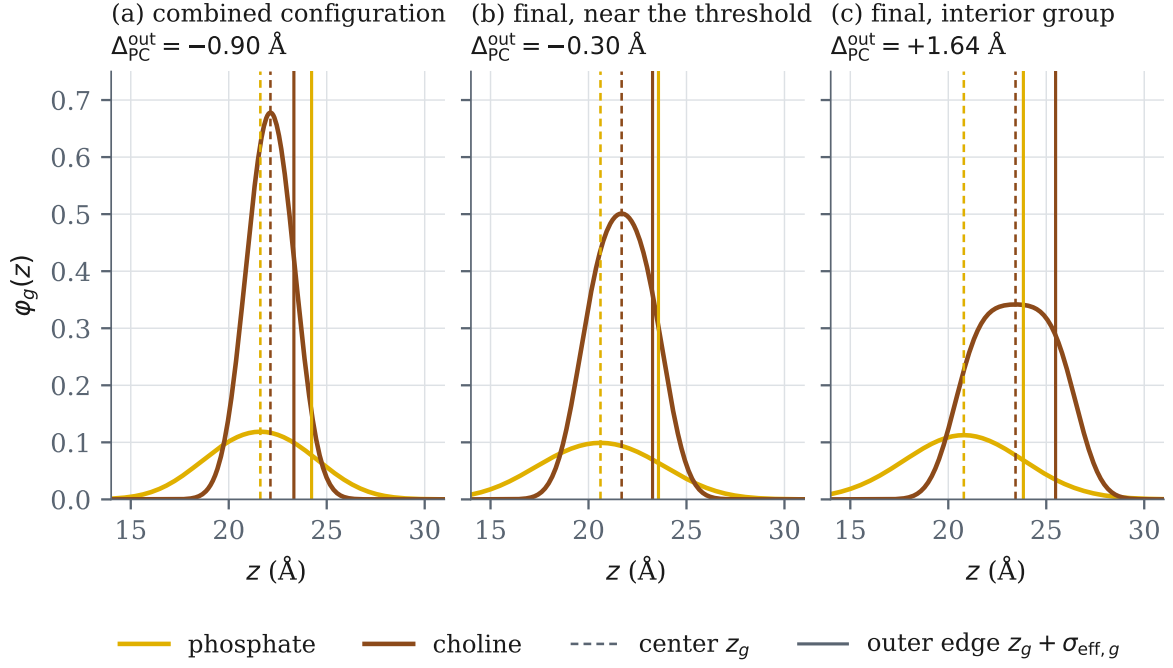


Figure 3.2: Volume fractions of phosphate and choline in profile 3, with the centers z_g (dashed) and the outer edges $z_g + \sigma_{\text{eff},g}$ (solid), whose distance is $\Delta_{\text{PC}}^{\text{out}}$. (a) Combined configuration, the run closest to the medians of $\Delta_{\text{PC}}^{\text{out}}$ in profiles 1 and 3 together. (b), (c) Final configuration, the representative runs of the group near the threshold and of the interior group (Sec. 3.2.2). The outer edge is defined through the second moment (2.24), not as the point where the two curves cross.

in Figure 3.2b–c and described in Sec. 3.2.2. In the conservative combination the restraint leaves no inversion deeper than $0.5 \text{ \AA}$ in 40 profiles, against 21 of 80 without it.

3.2 The final configuration

The final configuration (Sec. 3.1.5) differs from the combined configuration only by the envelope restraint. We therefore compare the two to isolate the effect of the restraint, whereas a comparison with the reference would also contain the effect of the weights and of the area margin. The final configuration was minimized in 30 runs, and structural quantities are given as mean $\pm$ stan-

dard deviation over the runs of one group of solutions (Secs. 2.12 and 3.3.1).

3.2.1 Agreement with the data

Figure 3.3 shows the seven curves with the model of the representative run of the group near the threshold (Sec. 3.2.2). Neither χ_{data}^2 , which is not a reduced χ^2 (Sec. 2.7), nor its value on curve 7 differs reproducibly from the combined configuration (Table 3.1). The envelope restraint has therefore no measurable cost in agreement with the data. The cost of 0.6 % with respect to the reference, 3.6 % on curve 7, is inherited from the headgroup-volume and choline-extension weights (Secs. 3.1.2 and 3.1.3).

Quantity	Combined (10 runs)	Final (30 runs)	Final – combined
Profiles with $\Delta_{\text{PC}}^{\text{out}} < -0.5 \text{ \AA}$	19/80	0/240	
χ_{data}^2	0.11606	0.11611	+0.04 % [−0.07, +0.17]
χ_{data}^2 , curve 7	0.13181	0.13167	−0.11 % [−0.66, +0.81]
Water, profile 3 (%)	−0.86	−0.02	+0.84 [+0.10, +1.22]
$v_{\text{PC}}^{\text{head}} (\text{\AA}^3)$	343.1	343.2	+0.09 [−0.03, +0.26]
$v_{\text{SM}}^{\text{head}} (\text{\AA}^3)$	285.1	285.1	+0.01 [−0.22, +0.23]

Table 3.1: Combined configuration, without the envelope restraint, and final configuration, with it. Counts are over all profiles; the other values are medians over the runs, and the last column gives the difference between the medians with the central 95 % of 5000 bootstrap resamplings of the two sets. Over the 30 runs of the final configuration χ_{data}^2 lies between 0.1158 and 0.1163. Water is the volume left to the hydration water, the minimum along z of $1 - \sum_g \phi_g(z)$ (Sec. 2.10).

3.2.2 Effect of the envelope restraint

The envelope restraint removes the inversions deeper than $0.5 \text{ \AA}$, present in 19 of the 80 profiles of the combined configuration and in none of the 240 of the

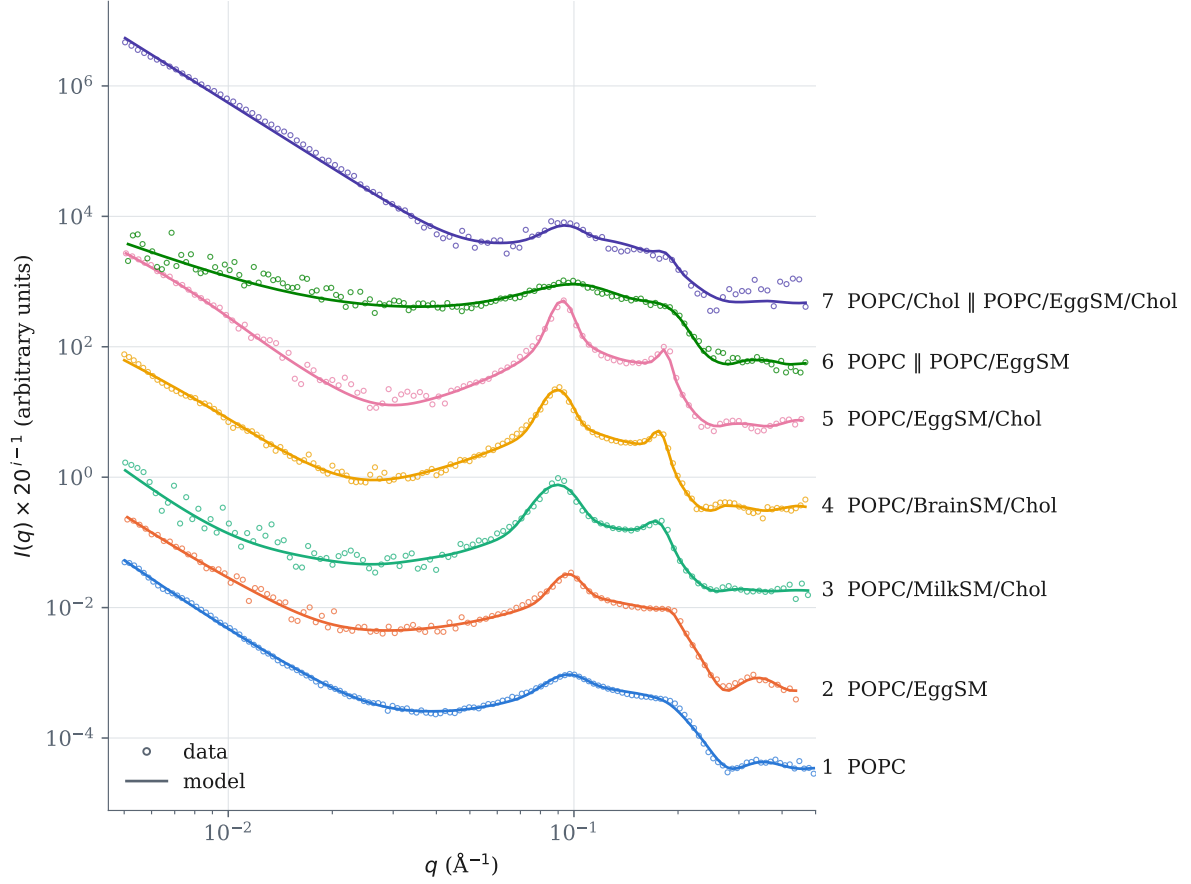


Figure 3.3: The seven SAXS curves (points) and the model of the final configuration (lines) in the fitted range $0.005 \leq q \leq 0.5 \text{ \AA}^{-1}$, for the representative run of the group near the threshold. Curve i is multiplied by 20^{i-1} ; the labels give the sample and its nominal composition (Table 2.2).

final one (Table 3.1). The headgroup volumes do not change, and the volume left to the hydration water in profile 3 returns to zero.

The minimum of $\Delta_{\text{PC}}^{\text{out}}$ over profiles 1 and 3 separates the 30 runs into two groups, by the criteria of Sec. 2.12 (largest gap $0.60 \text{ \AA}$, eight times the second). In the group near the threshold, 26 runs (95 % Wilson interval [78] 70–95 %), $\Delta_{\text{PC}}^{\text{out}}$ stops at the threshold, $-0.29 \text{ \AA}$ in profiles 1 and 3. In the interior group, 4 runs, it lies inside the allowed region. The fractions are frequencies of the optimizer, not populations of membranes in the samples. The two groups reach the same χ_{data}^2 with the same effective width of the phosphate in profile 1, 2.49 and $2.51 \text{ \AA}$, and different widths of the choline, 1.70 and $2.33 \text{ \AA}$. The rest of

the chapter refers to the group near the threshold, whose representative run, close to the medians of the group, is used in the figures.

The restraint acts on the choline. In profile 1 the centers of phosphate and choline stay at the threshold of the order restraint (2.20), and the effective width of the choline grows by 0.46 Å; in profile 3 the centers also separate (Figure 3.2a–b). In neither profile does the effective width of the phosphate change reproducibly.

Profile	A (Å ²)	D_h	σ_{eff}		$\Delta_{\text{PC}}^{\text{out}}$
			phosphate	choline	
1	68.7 ± 0.7	13.4 ± 0.1	2.49 ± 0.12	1.70 ± 0.12	$-0.29 \pm 0.02^{\ddagger}$
2	$56.9 \pm 0.4^{\ddagger}$	15.6 ± 0.1	2.22 ± 0.08	2.19 ± 0.10	3.01 ± 0.33
3	46.4 ± 0.6	18.3 ± 0.2	2.91 ± 0.18	1.70 ± 0.11	$-0.29 \pm 0.01^{\ddagger}$
4	$45.0 \pm 0.1^*$	17.8 ± 0.0	3.01 ± 0.10	2.93 ± 0.50	0.61 ± 0.41
5	$45.0 \pm 0.0^*$	17.4 ± 0.0	$2.07 \pm 0.08^{\ddagger}$	$4.03 \pm 0.06^{\dagger}$	3.29 ± 0.16
6 ₂	$56.0 \pm 0.3^{\ddagger}$	16.1 ± 0.1	$2.07 \pm 0.08^{\dagger\ddagger}$	1.58 ± 0.11	0.02 ± 0.14
7	51.5 ± 0.5	15.9 ± 0.2	$3.08 \pm 0.08^{\ddagger}$	$4.05 \pm 0.00^{\dagger}$	1.47 ± 0.08
7 ₂	$59.1 \pm 0.3^{\ddagger}$	14.4 ± 0.1	$2.75 \pm 0.52^{\ddagger}$	3.92 ± 0.12	4.61 ± 0.63

Table 3.2: Area, hydrophobic thickness, effective widths of phosphate and choline and $\Delta_{\text{PC}}^{\text{out}}$ in the final configuration, mean $\pm$ standard deviation over the 26 runs of the group near the threshold; lengths in Å. * At a bound in at least half of the runs. † Effective width with w and σ each at a bound in at least half of the runs. ‡ At the threshold of a restraint in at least half of the runs: the effective-width window or spread, the area relations (2.19), or the envelope restraint (2.26).

3.2.3 Profiles and derived observables

Figure 3.4 shows the volume-fraction profiles $\varphi_g(z)$ (Sec. 2.6) of the seven samples, computed from the parameters of the representative run. They give the structure returned by the fit in visual form. Across the samples they show how the distribution of the same chemical group shifts and widens with the composition, and how thick each monolayer is, from its hydrophobic thickness D_h . Samples 1–5 are symmetric by construction. For samples 6 and 7 the outer

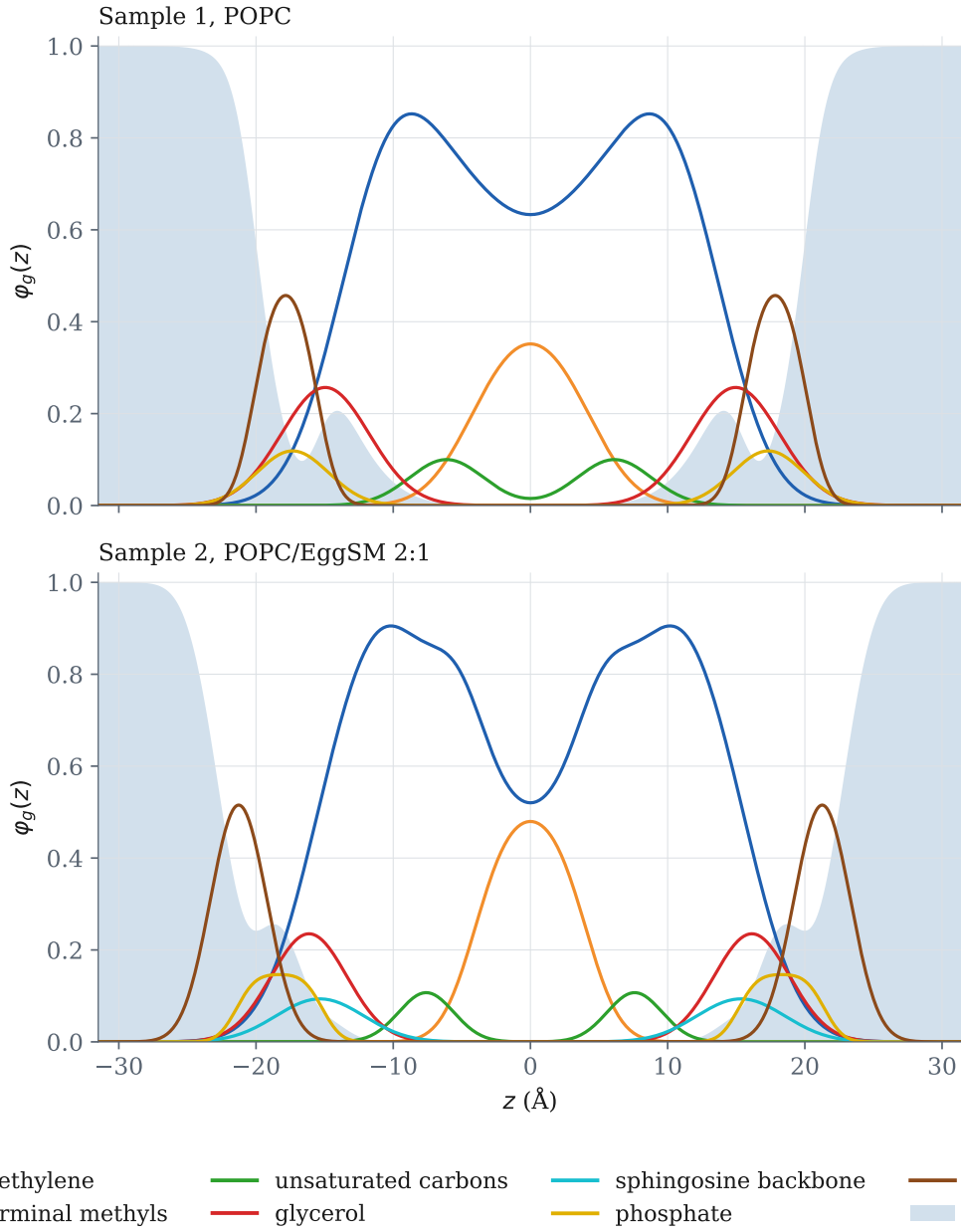


Figure 3.4: Volume-fraction profiles $\phi_g(z)$ of the final configuration, for the representative run of the group near the threshold; the shaded area is the water, $1 - \sum_g \phi_g(z)$. Samples 1 and 2.

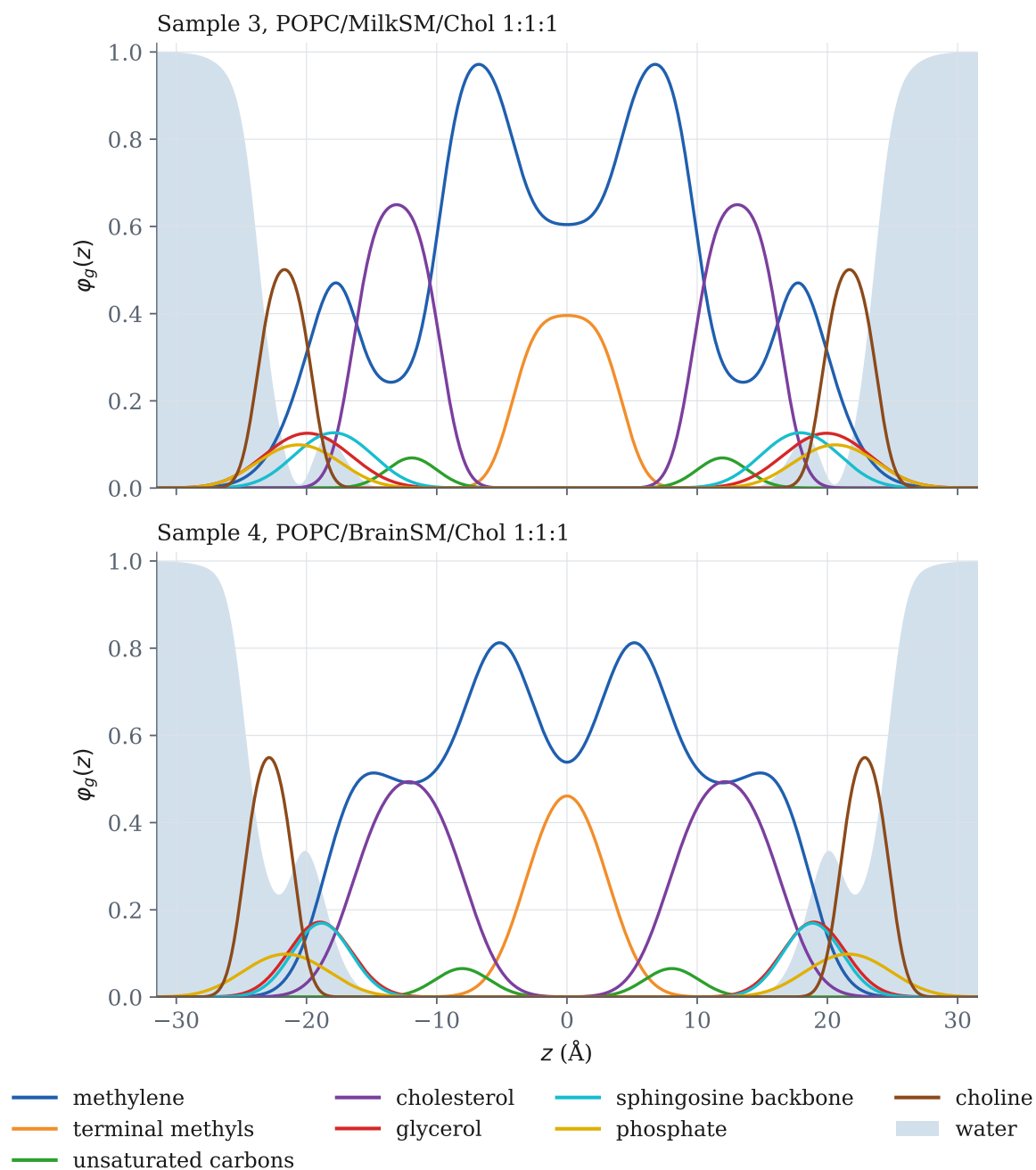


Figure 3.4: Volume-fraction profiles (continued). Samples 3 and 4.

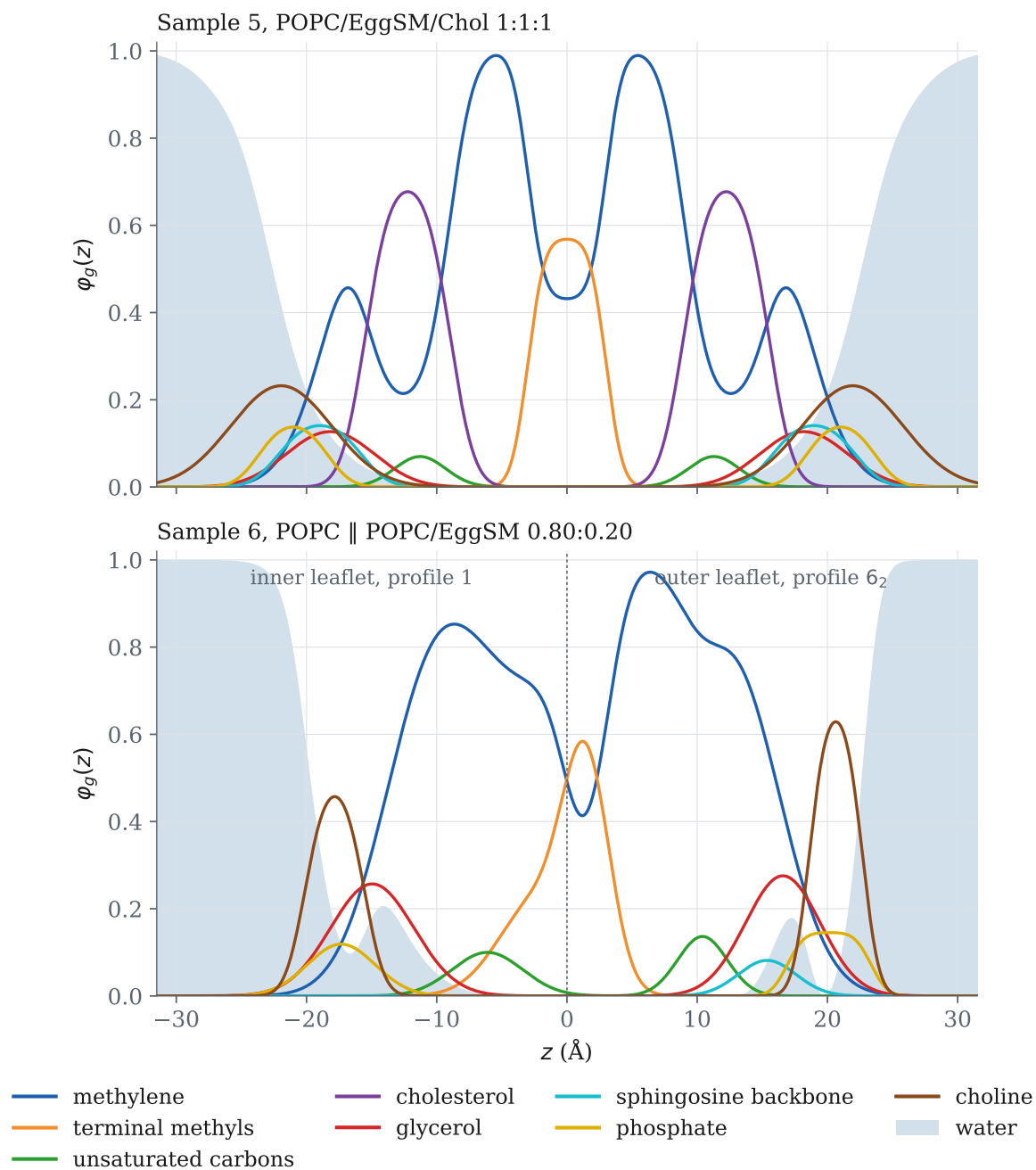


Figure 3.4: Volume-fraction profiles (continued). Samples 5 and 6; for sample 6, the outer lamella.

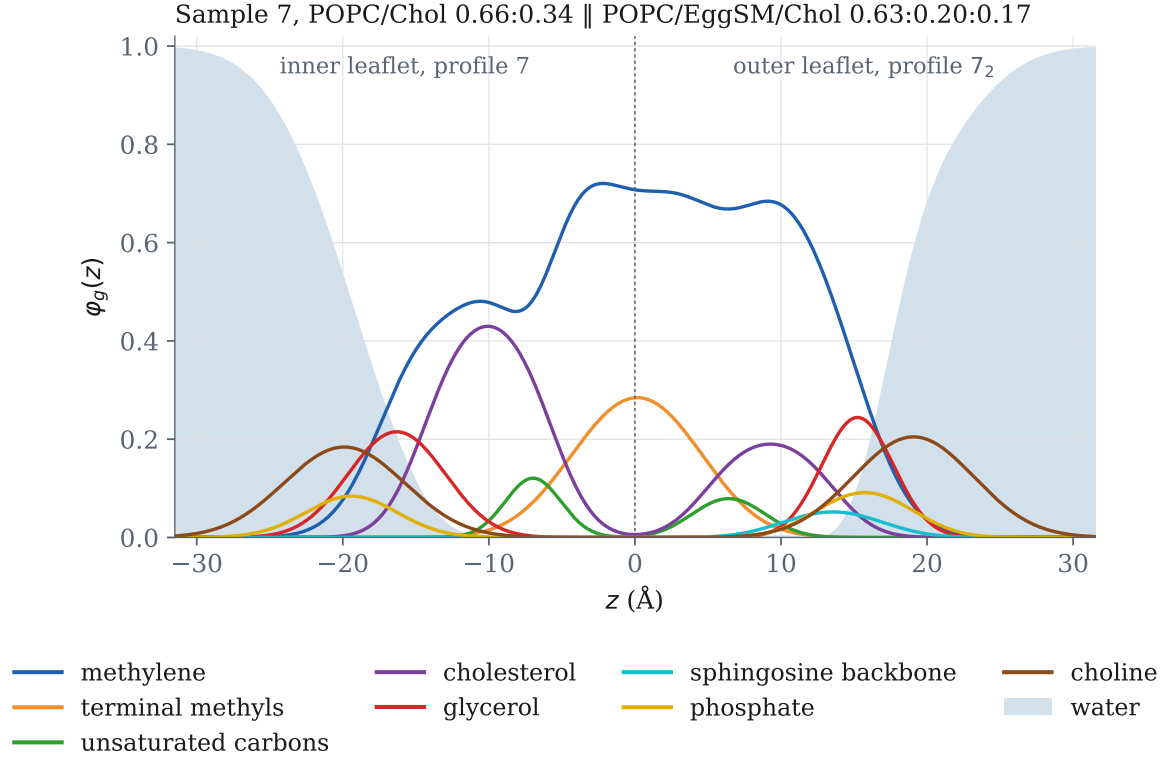


Figure 3.4: Volume-fraction profiles (continued). Sample 7, outer lamella.

lamella is shown, with the inner leaflet at $z < 0$; in sample 6 the inner leaflet is profile 1, as assumed in Sec. 2.5.

The position and the effective width of each distribution can be compared between samples. The detailed shape within that width depends on the split into w and σ , which the curves do not determine (Sec. 3.3.2). Table 3.2 gives, for each profile, the area, D_h , the effective widths of phosphate and choline and Δ_{PC}^{out} (Sec. 2.8.5). The marks identify the values set by a bound or held at the threshold of a restraint, discussed in Sec. 3.3.3; all the parameters are listed in Appendix A.

3.3 Parameters determined by the data

This section answers the first question of Sec. 1.6, which of the structural quantities returned by the model are determined by the SAXS curves and which by the constraints. The local analysis of Sec. 2.11 was carried out at the representative run of each group of solutions, and it is compared with the spread over the 26 runs of the group near the threshold.

3.3.1 Numerical rank and curvature

The singular values of the data Jacobian decrease over eight orders of magnitude without a gap, no value exceeding the next by more than a factor of three. Nothing separates the directions that the curves determine from those they do not, so the numerical rank depends on the cutoff, from 47 to 100 of 198 between 10^{-3} and 10^{-8} . For the same reason the local curvature gives no usable uncertainty. The uncertainty that the curvature assigns to $\Delta_{\text{PC}}^{\text{out}}$ of profile 3 grows from 119 Å to $2.5 \cdot 10^7$ Å over the same range of cutoffs (Sec. 2.12). Even the smallest value is more than twice the thickness of the bilayer, about 48 Å in profile 3, for a quantity whose value is -0.29 Å. The only uncertainty reported in this work is therefore the spread across runs, and the quantities are examined one by one in the rest of the section. Seventeen columns, whose derivatives depend on the finite-difference step in the restraint terms, are excluded from the matrices. Among them are the positions of the phosphate in profiles 1 and 3 and of the choline in profile 1, whose behavior is read from the runs (Sec. 3.2.2).

3.3.2 Half-width and smoothing

Half-width and smoothing of a group act on the curves almost only through their combination σ_{eff} . In the group near the threshold, 40 of the 45 pairs of w_g and σ_g with both columns in the matrices have nearly parallel columns, with a cosine above 0.99. The response along the direction of constant σ_{eff} (2.28) has a median of 1.2 % of that along the direction that changes it. Across the 26 runs, w and σ of phosphate and choline therefore move along the curve of constant σ_{eff} (Figure 3.5), with w varying by 26–49 % and σ_{eff} by 5–7 %. The data do not determine the split of σ_{eff} into w and σ , and σ_{eff} alone is taken as the structural quantity (Sec. 2.11).

3.3.3 Quantities set by the bounds and by the restraints

Much of the structure is set by the bounds and the restraints rather than by the curves. In at least half of the runs, 56 parameters are at a bound and 33 quantities at the threshold of a soft restraint (Table 3.3), and they take the value of the range or of the restraint. Among them is the condensation of samples 4 and 5, whose areas stop at the lower bound of 45 Å² (Sec. 2.8). The effective width of cholesterol is set in all profiles, and its position in four of the five. The areas of the asymmetric outer leaflets are held by the relations (2.19), and the volumes of cholesterol and of the two headgroups stay at the edges of their windows. Twelve bounds belong to the structure factor, to f_{ML} and to the background, whose narrow ranges were inherited (Sec. 2.6). The values of all these quantities are conditional on the ranges and on the restraints.

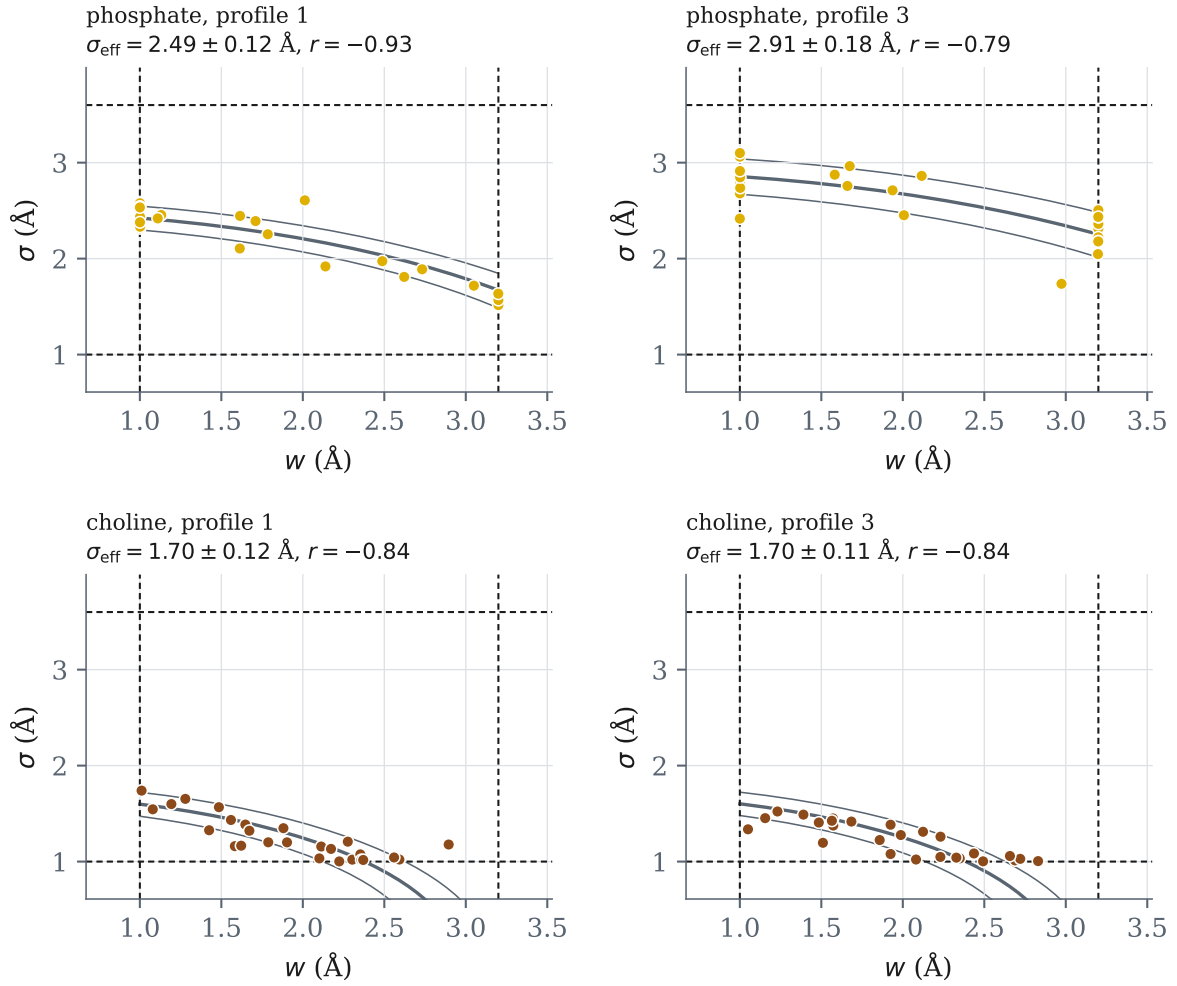


Figure 3.5: Half-width w and smoothing σ of phosphate and choline in profiles 1 and 3 for the 26 runs of the group near the threshold. The thick line is the curve of constant σ_{eff} (2.24) at the mean over the runs, the thin lines are at the mean $\pm$ one standard deviation, and the dashed lines are the bounds of the ranges. Each title gives σ_{eff} and the correlation coefficient r between w and σ .

3.3.4 Determined and undetermined quantities

The SAXS curves determine the effective width of the phosphate in profile 1, and in profile 3 except in five of the 26 runs, where it reaches the extreme of the spread restraint. In profile 1 no restraint on the widths holds it, and it is the same in the two groups of solutions, which differ in the width of the choline (Sec. 3.2.2).

The curves determine neither the split of σ_{eff} into w and σ nor, in profile

Class	Number	Quantities
At a bound in at least half of the runs		
w and σ both at a bound	18	terminal methyls 5, 7, 7 ₂ ; cholesterol 4, 7, 7 ₂ ; choline 5, 7; phosphate 6 ₂
w or σ at a bound	7	w of phosphate 1, 3; w of terminal methyls 1, 6 ₂ ; σ of terminal methyls 4; σ of cholesterol 3; w of sphingosine backbone 5
Positions	8	unsaturated carbons 1, 6 ₂ , 7, 7 ₂ ; cholesterol 7, 7 ₂ ; sphingosine backbone 3, 6 ₂
Areas	2	$A^{(4)}$, $A^{(5)}$ at 45 Å ²
σ_h	5	profiles 1, 3, 6 ₂ , 7, 7 ₂
Volumes, σ_{mid}	4	unsaturated carbons, sphingosine backbone, bulk water; σ_{mid} of sample 6
$S(q)$, f_{ML} , background	12	f_{ML} of curves 6, 7; d of curve 7; σ_d/d of curves 1–5; κ of curve 7; α_p of curves 3–5
At the threshold of a restraint in at least half of the runs		
Width window or spread	19	σ_{eff} of glycerol 1, 3, 4, 7; sphingosine backbone 4, 6 ₂ , 7 ₂ ; phosphate 5, 6 ₂ , 7, 7 ₂ ; unsaturated carbons 1, 6 ₂ , 7; cholesterol 3, 4, 5, 7, 7 ₂
Area relations (2.19)	3	$A^{(2)}$ and $A^{(6_2)}$; $A^{(7_2)}$
Order (2.20)	4	choline and phosphate in profiles 1, 6 ₂ , 7; phosphate and glycerol in 7 ₂
Envelope (2.26)	2	$\Delta_{\text{PC}}^{\text{out}}$ of profiles 1, 3
Cholesterol edge (2.23)	2	profiles 3, 5
Volume windows	3	v_{Chol} ; PC and SM headgroups

Table 3.3: Quantities at a bound, within 0.1 % of the range, or at the threshold of a restraint in at least half of the 26 runs of the group near the threshold. Number counts parameters for the bounds, so a group with w and σ both at a bound counts twice, and quantities for the restraints. Numbers after a group refer to the profiles or to the curves.

1, the effective width of the choline, which differs between the two groups of solutions at the same χ_{data}^2 (Sec. 3.2.2). In the group near the threshold that width follows from the order and envelope restraints. The quantities of Sec. 3.3.3 are set by the bounds or by the restraints, and the headgroup volumes by the balance between curve 7 and their restraint (Sec. 3.1.2).

The envelope restraint removes the inversions deeper than 0.5 Å at no measurable cost in χ_{data}^2 (Secs. 3.2.1 and 3.2.2). The value of -0.29 Å in profiles 1 and 3 is its threshold, not a measured separation, and the curves determine

neither the inversion nor the separation between the two outer edges.

3.4 Convergence of the simulations

The simulations of Sec. 2.13 are the reference for the second and third questions of Sec. 1.6, and their averages are meaningful only on stationary trajectories. The eight trajectories cover 300 ns each (Table B.1). Over the production windows the area per lipid is stationary, with a linear drift below 1 % over the last 200 ns and the means of the two halves within two standard errors. The surface tension is zero within 1.3 standard errors. The exception is the 1:1:1 bilayer, whose area still varies slowly but stays within $0.3 \text{ \AA}^2$ after 100 ns. Volumes and distributions are averaged over windows that start at or after the beginning of the production window.

The asymmetric bilayers were built so that each leaflet keeps the area per lipid of its own symmetric bilayer (Sec. 2.13.3). Their lateral area matches the prediction within 0.01 % for sample 6 and 0.4 % for sample 7, so the construction compresses or stretches each leaflet by at most that fraction.

In the four systems that contain cholesterol no molecule changed leaflet in 300 ns, even counting every crossing of the bilayer center. With no event in 31–61 μs of cumulative sampling per system, the 95 % Poisson bound places the mean residence time of a cholesterol molecule in its leaflet above 10–20 μs , so each leaflet keeps the composition with which it was built.

3.5 Group volumes from the simulations

MDtoSDP was applied to the symmetric bilayers with the compositions of samples 1, 2 and 5 and to the two asymmetric bilayers. The group volumes are the first product of the MDtoSDP chain (Figure 2.6), and every distribution of Sec. 3.6 depends on them.

3.5.1 Chemically heterogeneous groups

The model assigns one electron density to each group, but some groups contain chemically different parts in different regions of the bilayer (Figure 3.6). The clearest case is cholesterol. Its ring system carries 151 of the 216 electrons of the molecule at 8–17 Å from the center, with an electron density of 0.40–0.43 e/Å³, while its branched chain fills the central 8 Å with methylenes at 0.295 and methyls at 0.17 e/Å³. The methine carbons also lie in three different places, in the acyl chains at about 10 Å from the center, in the sphingosine at 18 Å and in the cholesterol chain near the center.

A single specific volume cannot describe such a group. Two routes lead from the simulation to the groups of the model (Sec. 2.15.2). In the direct route the volumes are solved on the groups of the model, with cholesterol as one group. In the second they are solved on a fine partition into chemically homogeneous groups, and the groups of the model are obtained by summing their volume fractions.

Without cholesterol the two routes give the same volumes within 4 % (Table 3.4). With cholesterol the direct route gives too little volume to the center of the bilayer, where the light chain lies, because it uses one specific volume

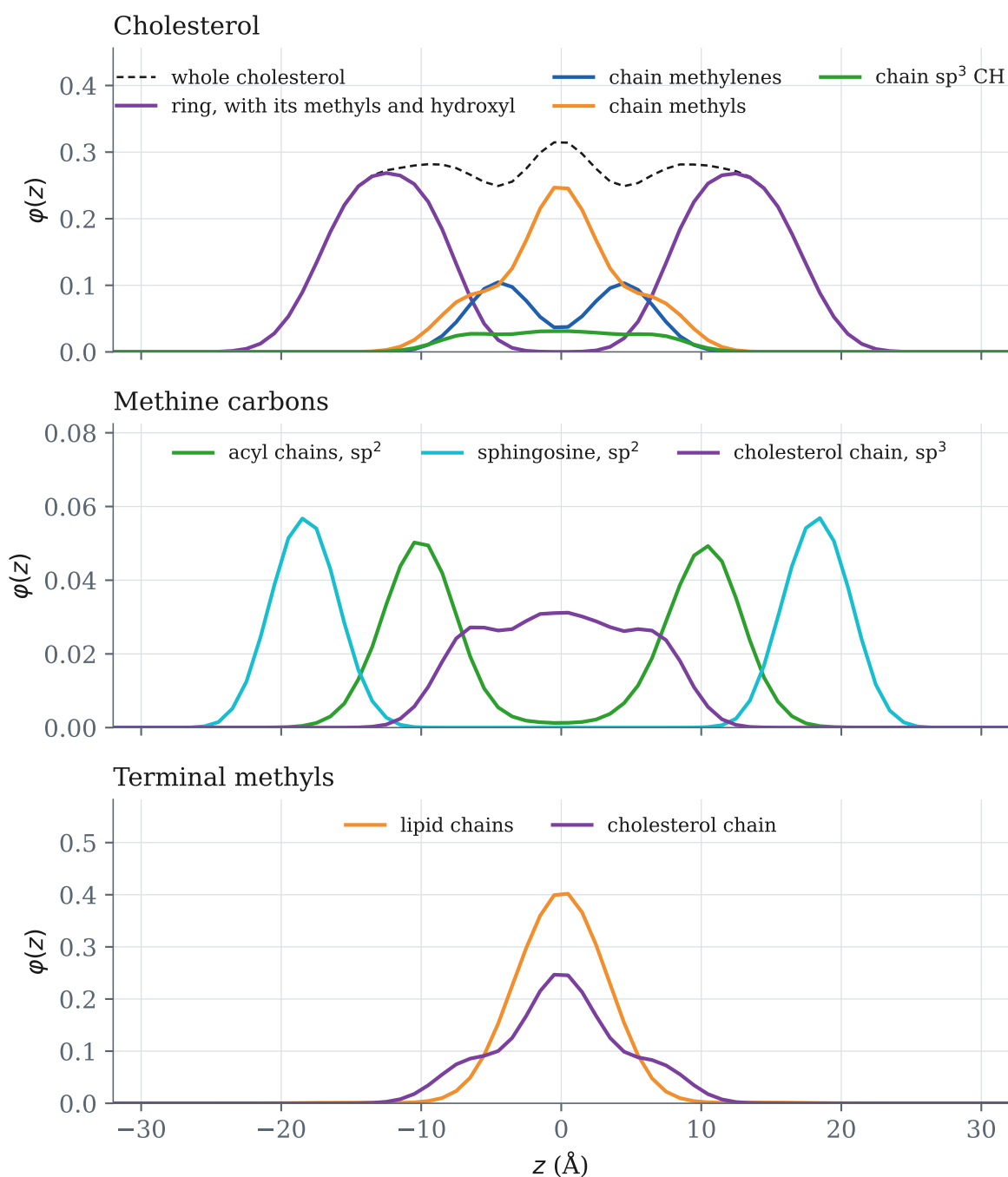


Figure 3.6: Volume-fraction distributions of the parts of three chemically heterogeneous groups in the simulated POPC/EggSM/Chol 1:1:1 bilayer: the ring and the chain of cholesterol, with their sum; the methine carbons of the acyl chains, of the sphingosine and of the cholesterol chain; the terminal methyls of the lipid chains and of the cholesterol chain.

for the whole molecule. In the asymmetric bilayer of sample 7 the missing volume goes to the terminal methyls, which grow from 54.8 to 69.5 Å³. In the 1:1:1 bilayer, where the volume of the methyls is imposed, it stays as a hole,

filled only to 0.85 at the center.

Each route gives an electron density of the simulated bilayer, recombined from its distributions and from the electron content of its groups, and from it a SAXS curve (Sec. 2.15.4). The reference is the curve of the fine partition, in which each piece keeps its own electron density and which agrees with the density computed directly from the trajectory within R_F (2.35) of 0.02 above $0.08 \text{ \AA}^{-1}$ (Appendix B.3). The curves of the two routes are compared with it through the shift of the first minimum and through the same index, used for the representation in Sec. 2.15.3, here between the amplitudes of each route and of the fine partition. The summation keeps the volumes of the fine partition but gives cholesterol a single electron density, which moves the first minimum of the curve to lower q , from 0.235 to $0.216 \text{ \AA}^{-1}$ in the 1:1:1 bilayer and from 0.244 to $0.230 \text{ \AA}^{-1}$ in sample 7, by 8 and 6 % of its position. The direct route comes closer to the curve of the fine partition, with R_F of 0.18 and 0.06 against 0.22 and 0.15, but with the deformed volumes above.

System	Largest volume difference (%)	R_F		First minimum (%)	
		direct	summation	direct	summation
POPC/EggSM 2:1	3.6, phosphate	0.001	0.006	0	0
Asymmetric, sample 6	0.7, phosphate	0.000	0.001	0	0
POPC/EggSM/Chol 1:1:1	17.3, glycerol	0.18	0.22	−5	−8
Asymmetric, sample 7	27.0, methyls	0.06	0.15	−1	−6

Table 3.4: The two routes to the groups of the model, volumes solved directly on the groups of the model (direct) and summation from the fine partition. Largest relative difference of a group volume between the two routes. R_F (2.35) and relative shift of the position of the first minimum of the curve of each route with respect to that of the fine partition, both computed from the simulated electron density (Sec. 2.15.4), on a grid of step $0.0028 \text{ \AA}^{-1}$ between 0.05 and $0.6 \text{ \AA}^{-1}$.

The summation was adopted. It keeps the volume distributions and gives each group the chemical content of the corresponding group of the fit, so that

positions, widths and volumes can be compared with the fit (Sec. 3.7). Its difference from the fine partition in the curve is the imprint of the single electron density of cholesterol, an approximation of the model and not of the simulation, which Secs. 3.6.2 and 3.8.1 quantify.

3.5.2 Volumes of the groups of the model

Table 3.5 lists the volumes of the groups of the model. The solution is well conditioned once the geometric terms are added, the volumes of the molecules add up to the volume of the box within 0.3 %, and the filling deviates from one by at most 1.4 % root mean square (Appendix B.2). Phosphate and choline are strongly anticorrelated in the solution, with r between -0.91 and -0.95 , so their sum is better determined than either and varies less among the systems. The partition between them depends on the definition of volume at the interface. The radical tessellation, which assigns the free volume of the interface to the neighboring atoms, gives the phosphate and the choline of POPC 55.7 and 148.1 Å³, against 64.7 and 109.9 Å³ from the filling. The dry molecular volume of POPC, 1242 Å³ from the sum of the group volumes, lies within the window of Table 2.5, centered on the densitometric value [45], and the PC headgroup has 306–319 Å³ in the simulations, against about 331 Å³ in the literature [24] and the window of the fit (Sec. 3.7.1).

Group	POPC/EggSM		POPC/EggSM/Chol 1:1:1	Asymmetric	
	POPC	2:1		sample 6	sample 7
Methylene	26.9	27.1	27.1 [†]	27.2	26.8
Terminal methyls	58.4	57.8	52.6	57.6	54.8
Unsaturated carbons	26.5	23.9	21.5 [†]	23.8	23.5
Glycerol	143.9	132.5	139.4	139.8	143.1
Sphingosine backbone	–	177.3	160.6	175.7 [†]	178.2 [†]
Phosphate	64.7	59.3	63.3	63.2	60.2
Choline	109.9	114.4	111.9	112.1	114.2
Cholesterol	–	–	624.1 [†]	–	653.7 [†]
Phosphate and choline	174.6	173.7	175.1	175.3	174.4
PC headgroup	318.5	306.2	314.6	315.1	317.5

Table 3.5: Volumes ($\text{\AA}^3$) of the groups of the model in the simulated systems. The sphingosine backbone includes its two unsaturated carbons, and the cholesterol the whole molecule. The PC headgroup is the sum of glycerol, phosphate and choline. [†] Volume imposed at the geometric value, for the whole group or for one of its members. The statistical errors from the blocks are at most $0.7 \text{ \AA}^3$, or 1.3 %.

3.6 Simulated distributions and their representation

This section answers the first part of the second question of Sec. 1.6, whether the erf slabs reproduce the simulated distributions of the groups and how large an imprint their shape residual leaves on the SAXS curve. The distributions are those of the groups of the model, obtained by summation from the fine partition (Sec. 3.5.1).

3.6.1 Representation with erf slabs

Figure 3.7 shows the distributions of the asymmetric bilayer of sample 7 with their representation. Table 3.6 gives the shape residuals in all five systems, below 1 % of the peak for every group except cholesterol. The distribution of cholesterol sums those of the ring and of the chain (Figure 3.6), which a single

slab does not reproduce. The positions and the effective widths (2.24) of the slabs agree with the mean and the standard deviation of the simulated distributions within 0.1 Å for the polar groups, 0.2 Å for the unsaturated carbons and 0.35 Å for cholesterol.

Group	POPC/EggSM		POPC/EggSM/Chol 1:1:1	Asymmetric	
	POPC	2:1		sample 6	sample 7
Methylene	0.64	0.47	0.89	0.38	0.66
Terminal methyls	0.53	0.47	0.17	0.69	0.54
Unsaturated carbons	0.71	0.32	0.93	0.40	0.36
Cholesterol	–	–	2.17	–	1.14
Glycerol	0.16	0.26	0.22	0.28	0.25
Sphingosine backbone	–	0.29	0.15	0.23	0.25
Phosphate	0.21	0.47	0.28	0.42	0.41
Choline	0.17	0.34	0.24	0.33	0.27
R_F on $ F(q) $	0.023	0.026	0.071	0.039	0.027
Largest shift of a minimum (%)	0.0	0.0	1.3	0.7	1.0

Table 3.6: Root-mean-square residual of the representation of each group over the whole box, in percent of the peak of the simulated distribution; the methylene is represented by two slabs, not by the envelope of the model. Below, R_F (2.35) between the amplitudes computed from the distributions and from their representation, and largest relative shift of a minimum of $|F(q)|^2$ present in both curves, in the q -range 0.05–0.6 Å^{−1} on a grid of step 0.0028 Å^{−1}.

In the symmetric bilayers the two leaflets, which have the same distributions, are represented independently. For the polar groups and the unsaturated carbons, whose two slabs do not overlap, the half-widths of a group differ between the leaflets by up to 2.0 Å and the effective widths by at most 0.09 Å. The split of the effective width into half-width and smoothing is thus not determined even on the simulated distributions, as in the fit of the experimental curves (Sec. 3.3.2) and in the closure test (Sec. 3.8.2).

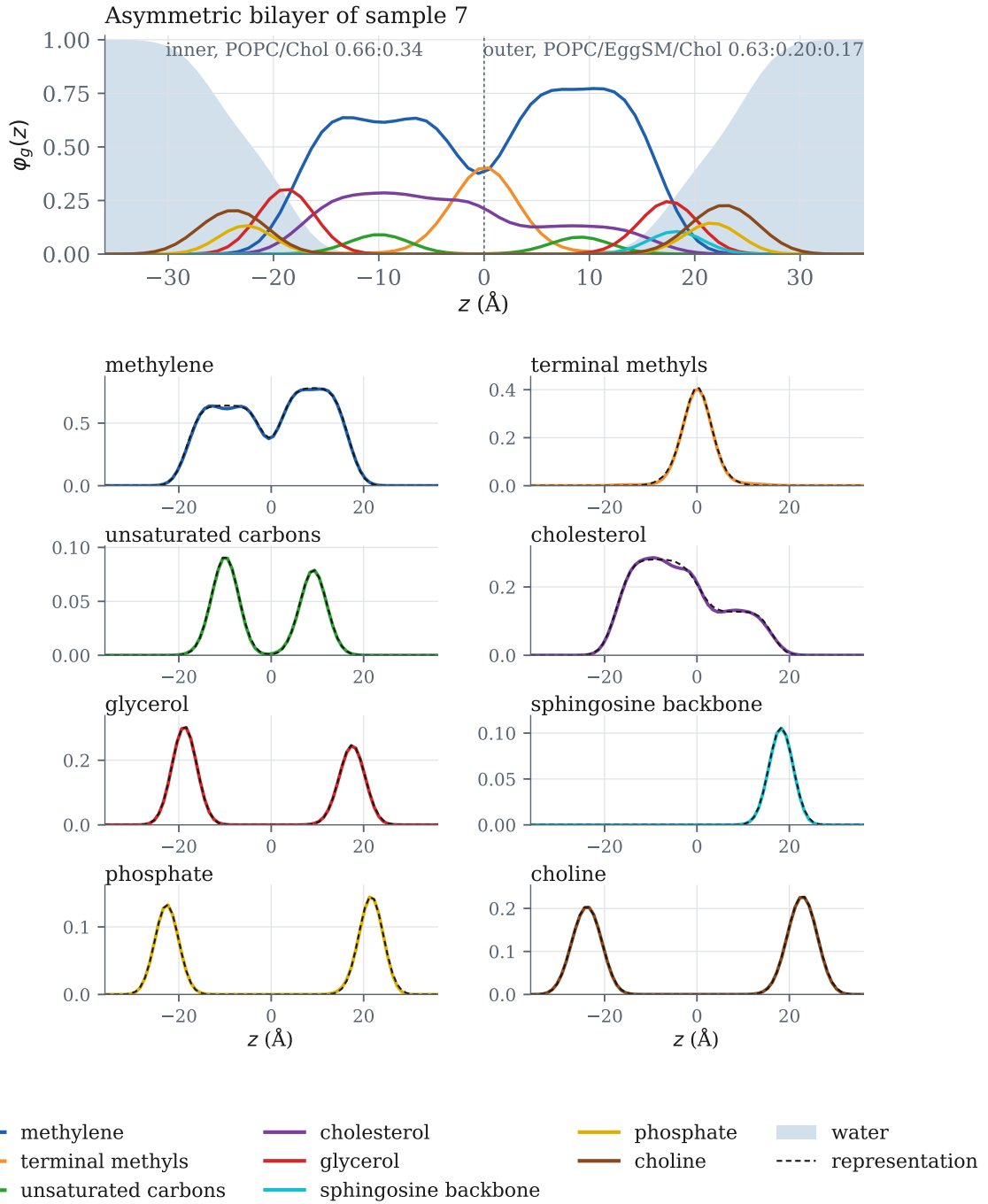


Figure 3.7: Volume-fraction distributions $\phi_g(z)$ of the groups of the model in the simulated asymmetric bilayer of sample 7, inner leaflet at $z < 0$. Above, all groups, with the water $1 - \sum_g \phi_g(z)$ shaded. Below, each group (solid) with its representation by erf slabs (dashed).

3.6.2 Imprint on the amplitudes

The curve of the representation is computed from the simulated electron density in which each distribution is replaced by its slabs at the electron density of the group, with the rest of the density unchanged (Sec. 2.15.3). Its amplitudes differ from those of the distributions by R_F of at most 0.071. The minima of $|F(q)|^2$ shift by at most one step of the q grid within the range of the measured curves, which end at 0.44–0.49 Å⁻¹, and by two steps beyond it (Table 3.6 and Appendix B.3). The single electron density of cholesterol leaves a larger imprint, R_F of 0.22 and 0.15 in the 1:1:1 bilayer and in sample 7 with respect to the fine partition (Table 3.4).

The erf slab thus reproduces the simulated distributions except cholesterol. In the two systems with cholesterol, the part it does not reproduce leaves on the amplitudes a smaller imprint than the single electron density of cholesterol, which the closure test identifies as the main error of the functional form (Sec. 3.8).

3.7 Comparison between fit and molecular dynamics

This section answers the third question of Sec. 1.6, whether the structure obtained from the SAXS curves agrees with simulated membranes of the same nominal composition. Profiles 1, 2 and 5 are compared with the symmetric simulated bilayers, profile 6₂ with the outer leaflet of the asymmetric bilayer of sample 6, and profiles 7 and 7₂ with the two leaflets of that of sample 7. The fit values are means over the 26 runs of the group near the threshold, and

the simulated ones are those of the representation (Sec. 3.6.1). A quantity that the fit holds at a bound or at the threshold of a restraint (Sec. 3.3.3) is set by them. For such a quantity the comparison tells whether the simulation lies in the allowed region, not whether it agrees with the curves. As a scale, the same POPC leaflet has effective widths 0.25–0.30 Å apart in the symmetric bilayer and in the asymmetric bilayer of sample 6.

3.7.1 Profile-by-profile comparison

Table 3.7 compares the quantities that describe the polar region. The only one determined by the curves is the effective width of the phosphate in profile 1 (Sec. 3.3.4). It is 2.49 Å in the fit against 2.69 and 2.99 Å in the two simulated POPC leaflets, the second the inner leaflet of sample 6, a difference of the size of the scale above. The fit and the simulations contain the same amount of chain, since the hydrophobic volume per molecule, dominated by the methylene, agrees within 0.7 % in all six profiles. With the same volume, the larger areas of profiles 1 and 7 in the fit, by 6 and 10 %, give thinner hydrophobic thicknesses. In all six profiles the phosphate of the fit lies closer to the hydrocarbon region than in the simulations, by 0.8–3.8 Å, and in four of them its position is held by the order restraint.

The simulations satisfy with margin every geometric restraint that the fit holds at its threshold (Table 3.3). The simulated choline lies 1.2 Å outside the phosphate, against the 0.5 Å of the order restraint, and its outer edge 1.6–1.8 Å beyond that of the phosphate, against the threshold of –0.3 Å of the envelope restraint. The area relations and the cholesterol edge are also met. These restraints therefore exclude configurations that the simulations do not

Profile	A ($\text{\AA}^2$)		D_h		$z_P - D_h$		$\sigma_{\text{eff, phosphate}}$		$\Delta_{\text{PC}}^{\text{out}}$	
	fit	MD	fit	MD	fit	MD	fit	MD	fit	MD
1	68.7	64.9	13.4	14.3	4.3 [‡]	5.1	2.49	2.69	-0.29 [‡]	1.78
2	56.9 [‡]	60.4	15.6	14.9	2.5	5.2	2.22	3.02	3.01	1.66
5	45.0*	43.0	17.4	18.2	3.7	5.1	2.07 [‡]	2.35	3.29	1.63
6 ₂	56.0 [‡]	62.1	16.1	14.4	4.1 [‡]	5.2	2.07 ^{†‡}	3.08	0.02	1.66
7	51.5	46.6	15.9	17.7	4.0 [‡]	4.9	3.08 [‡]	2.57	1.47	1.77
7 ₂	59.1 [‡]	51.9	14.4	16.5	1.4 [‡]	5.1	2.75 [‡]	2.65	4.61	1.70

Table 3.7: Area, hydrophobic thickness, distance of the phosphate from the hydrophobic thickness, effective width of the phosphate and separation between the outer edges of choline and phosphate, in the fit (mean over the 26 runs of the group near the threshold) and in the simulations; lengths in $\text{\AA}$. Simulated areas from Table B.1, divided by the lipids of each leaflet in the asymmetric bilayers; the representation uses the area obtained by integrating the distributions, which differs by at most 1.5 %. The simulated hydrophobic thickness is the hydrophobic volume divided by that area. * At a bound. [†] w and σ each at a bound. [‡] At the threshold of a restraint; for z_P , of the order restraint with the choline (1, 6₂, 7) or with the glycerol (7₂). All marks refer to at least half of the runs, as in Table 3.2.

adopt either.

The windows and bounds of Table 3.8, instead, exclude the simulated values, and they fall into three groups. The cholesterol slab of the fit is held narrow and away from the center by its bounds. The simulated cholesterol, whose distribution sums the ring and the chain (Sec. 3.6.1), is 1.6 to 2.3 times as wide and closer to the center. The condensation of sample 5 by cholesterol is stronger in the simulation than the lower bound of the area at which the fit stops. The headgroup volumes of the fit lie at the edges of their windows, with the PC headgroup above its window and the sphingosine backbone at its lower bound, whereas the simulations give a smaller PC headgroup and a larger backbone.

Quantity	Profiles	Allowed region	Fit	MD
Cholesterol slab				
σ_{eff} (Å)	5	2.0–3.8	2.41	5.58
σ_{eff} (Å)	7, 7 ₂	$w \leq 4.3, \sigma \leq 2.3$	3.38	5.51, 5.47
f_{Chol}	7, 7 ₂	0.65–0.88	0.65	0.50, 0.46
Condensation				
$A^{(5)}$ (Å ²)	5	≥ 45	45.0	43.0
Volumes (Å ³)				
PC headgroup	all	321–341	343.2	306.2–318.5
SM headgroup	all	264–284	285.1	291.5–310.3
v_{amide}	all	82–112	82.0	116.4–135.9
v_{CH}	all	20–23	22.9	23.5–26.5

Table 3.8: Windows and bounds that exclude the simulated values, with the fit at the edge of the window or at the bound in at least half of the runs. The position is $f_g = |z_g|/D_{h,\alpha}$ (Sec. 2.8). The SM headgroup is $v_{\text{amide}} + v_{\text{P}} + v_{\text{choline}}$, with v_{amide} the volume of the sphingosine backbone without its two unsaturated carbons. The volume of the unsaturated carbons lies inside its range only in the 1:1:1 bilayer, where it is imposed.

3.7.2 The asymmetric samples

In the simulation the POPC leaflet of the asymmetric bilayer of sample 6 places the groups within 0.25 Å of the symmetric POPC bilayer, as the fit assumes by using profile 1 for it (Sec. 2.5). The asymmetry of that bilayer is barely visible in the amplitude. The indicator of Sec. 2.15.4, the maximum of $|\text{Im}F(q)|$ divided by that of $|F(q)|$, is 1.33 %, about as large as in the symmetric bilayers, 0.30–1.30 %, while for sample 7 it is 5.09 %.

In sample 7 the fit and the simulation agree on the direction of the asymmetry (Table 3.9). In both the outer leaflet has the larger area per lipid, and its phosphate and choline lie closer to the bilayer center. In the fit, however, the area difference is held by its relation and the outer phosphate by the order restraint, so that the size of the asymmetry is set by the restraints. In sample 6 the simulated leaflets are almost alike, while the fit has an area difference of 18 % and a phosphate displaced by 2.55 Å. Here too the area is at its relation, and the

choline follows the phosphate through the order restraint. A simulated bilayer like that of sample 6 leaves almost no trace of its asymmetry in the curve, and the asymmetry of the fit comes from its restraints. The closure test shows that a single curve does not determine these differences in position (Sec. 3.8.2).

Sample	δA ($\text{\AA}^2$)		δz_P ($\text{\AA}$)		$\delta z_{\text{choline}}$ ($\text{\AA}$)	
	fit	MD	fit	MD	fit	MD
6	$-12.70^\ddagger$	-2.81	$+2.55^\ddagger$	+0.06	$+2.56^\ddagger$	+0.03
7	$+7.65^\ddagger$	+5.27	$-4.10^\ddagger$	-1.01	$-1.15^\ddagger$	-1.06

Table 3.9: Differences between the outer and the inner leaflet of the two asymmetric samples in area and in the distances of phosphate and choline from the bilayer center, in the fit (profiles 6₂ and 1, 7₂ and 7) and in the simulated asymmetric bilayers. The simulated δA follows from the numbers of lipids chosen in the construction of the simulated bilayers (Sec. 2.13.3). ‡ Involves quantities at the threshold of a restraint, the area relations (2.19) for δA and the order restraint (2.20) for the positions, with the choline and the phosphate held 0.5 $\text{\AA}$ apart in profiles 1, 6₂ and 7 and the phosphate held 0.5 $\text{\AA}$ from the glycerol in profile 7₂.

3.7.3 Agreement and differences

The fit and the simulations agree on the amount of chain and on the direction of the asymmetry of sample 7. The one quantity determined by the curves differs by as much as two simulated environments of the same leaflet differ from each other. They differ in the areas of profiles 1 and 7 and in the position of the phosphate, which in the fit lies closer to the hydrocarbon region. The geometric restraints of the fit are compatible with the simulations, whereas the bounds on the cholesterol slab, the lower bound on $A^{(5)}$ and the windows on the headgroup volumes exclude them. In sample 6 the fit has an asymmetry, held by its restraints, that the simulation does not show.

3.8 Closure test on synthetic curves

This section reports the closure test of Sec. 2.14, the second part of the second question of Sec. 1.6. Sec. 3.8.1 measures the error of the functional form with the model at the true parameters, and Sec. 3.8.2 what a single synthetic curve determines with all parameters free. The results shown are those at the lowest noise, 0.1 % on the first point, the case most favorable to the recovery. The other noise levels and bounds are in Appendix C.

3.8.1 Error of the functional form

For sample 7 with cholesterol as one group, the model at the true parameters does not reproduce the synthetic curve, with χ^2_{data} of 706, whereas with cholesterol as ring and chain it reaches 3.20 (Figure 3.8 and Table 3.10). Freeing the position and the widths of cholesterol, together with the widths of the terminal methyls, lowers it to 161 but not to the level of the noise, and in every converged run the cholesterol positions end at their bound. For sample 6, which has no cholesterol, the terminal methyls are the only groups free in this test. Their widths absorb most of the residual at the true parameters, since the representation does not define their split between the two leaflets (Sec. 2.14).

Bilayer	True parameters	Methyls and cholesterol free	All free
Sample 6	16.2	1.52	1.15
Sample 7, cholesterol as one group	706	161	1.14
Sample 7, cholesterol as ring and chain	3.20	1.46	1.14

Table 3.10: χ^2_{data} of the closure test with 0.1 % noise. At the true parameters; with only w_g and σ_g of the terminal methyls, the position, w_g and σ_g of cholesterol (or of ring and chain) and σ_{mid} free within ± 50 % (for sample 6, the terminal methyls and σ_{mid}); with all parameters free within ± 20 %. For the fits, lowest value among the converged runs. Other noise levels in Table C.1.

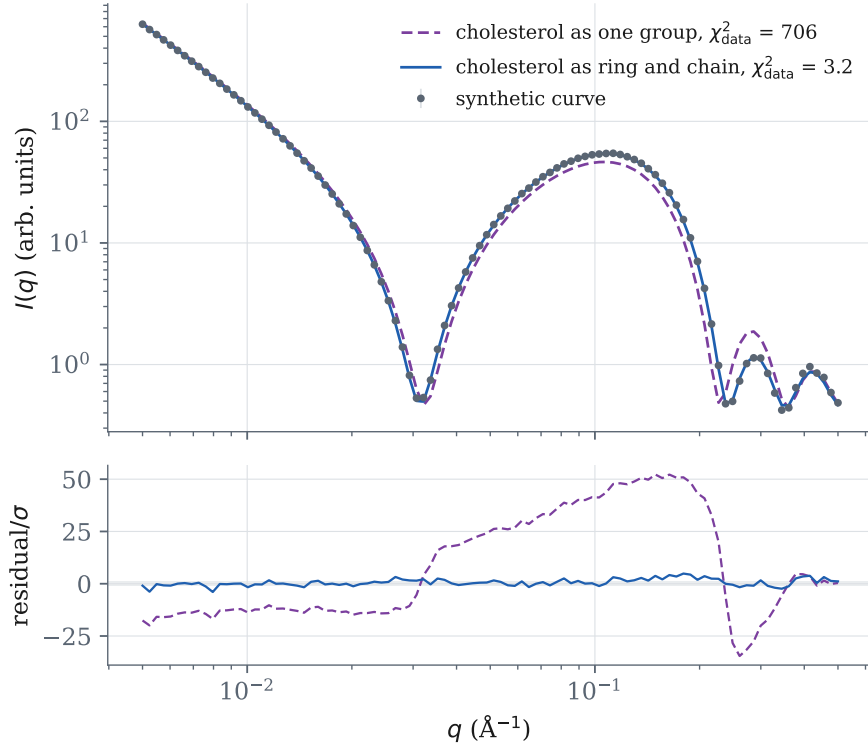


Figure 3.8: Synthetic curve of the asymmetric bilayer of sample 7 with 0.1 % noise, and the model at the true parameters with cholesterol as one group and as ring and chain. Below, residuals in units of the uncertainty.

For sample 7 the single electron density of cholesterol is therefore the main error in the functional form of the ASDP model, and dividing cholesterol into ring and chain removes it.

3.8.2 Recovery of the parameters

With all parameters free, the best fits reach the level of the noise in all three cases (Table 3.10). The true parameters are those of the representation of the simulated distributions, not the minimum of χ^2_{data} on the curve. Left free, the fit moves away from them and compensates the difference between the representation and the simulated distributions by deforming the profile. The curve of sample 7 with cholesterol as one group is thus reproduced by parameters that differ from the true ones, and a χ^2_{data} at the level of the noise does not reveal

the error of the functional form. In the best fits of sample 6 and of sample 7 with ring and chain (Figure 3.9), the hydrophobic thicknesses and the areas are within about 6 % of their true values, the positions of the polar groups within 26 % and the volumes within 16 %. Among the converged runs the thicknesses and the areas vary by up to 22 %, the positions by up to 30 %, and the volumes cover almost the whole interval of ± 20 %. The other noise levels give spreads of the same order. In these runs 28 and 33 % of the parameter values end at a bound, and 40–56 % of the half-widths and smoothings (Appendix C.2).

When the rest of the profile is held at its true values, as in the second test of Sec. 3.8.1 with bounds at ± 20 %, half-width and smoothing of the inner cholesterol ring are degenerate, as on the experimental curves (Sec. 3.3.2). The smoothing reaches its lower bound in every run and the half-width grows, while the effective width (2.24) stays 5–8 % below the true value. For the outer ring half-width and smoothing end at opposite bounds, and the chain is not determined.

With all parameters free, instead, the effective widths are not recovered. Their median deviation over groups and runs is 14 % for sample 6 and 17 % for sample 7 with ring and chain, with bounds at ± 20 %, and half-width and smoothing grow or shrink together across the runs.

The area difference between the leaflets keeps its sign in every converged run for sample 7 and is not recovered even in sign for sample 6 (Table 3.11). The differences in the positions of phosphate and choline span 3–9 Å among the runs.

The closure test thus separates the two causes. The error of the functional form lies in the single electron density of cholesterol, and dividing cholesterol

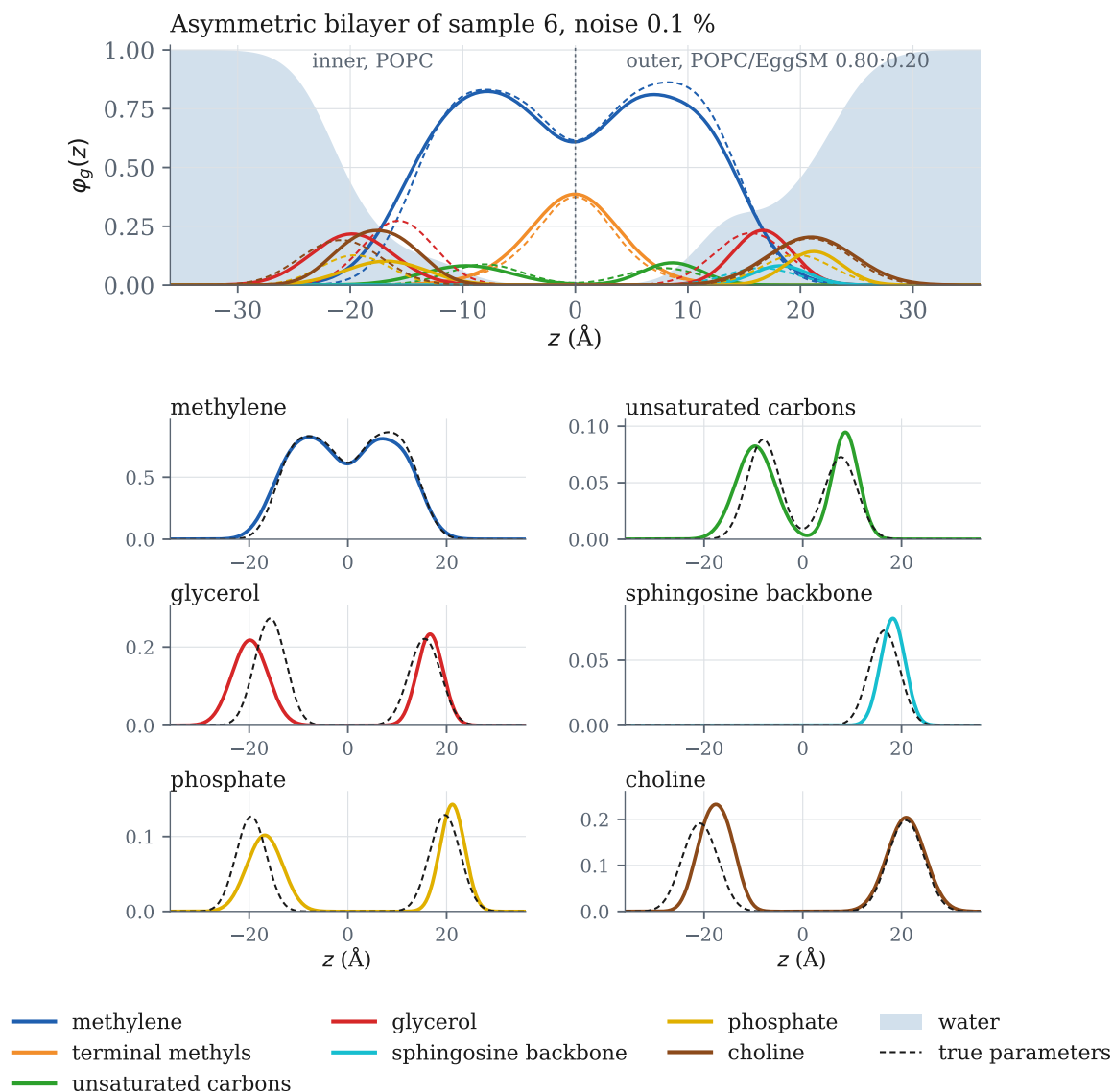


Figure 3.9: Volume-fraction profiles of the best fit of the closure test (solid) and at the true parameters (dashed), with 0.1 % noise and bounds at ± 20 %. Above, all groups, with the water of the best fit shaded; below, each group except the terminal methyls. Asymmetric bilayer of sample 6, inner leaflet at $z < 0$.

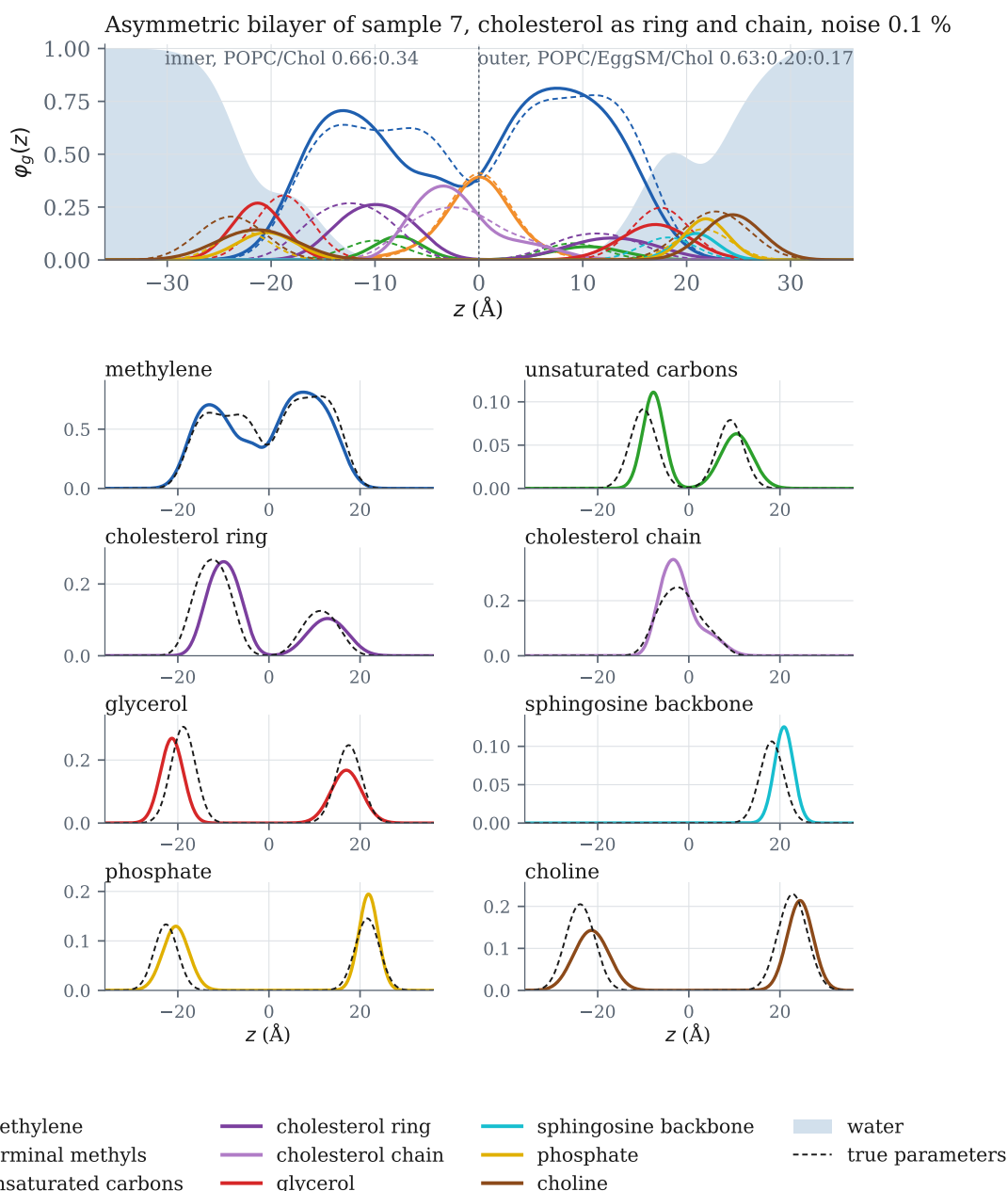


Figure 3.9: Best fits of the closure test (continued). Asymmetric bilayer of sample 7, cholesterol as ring and chain.

into ring and chain removes it. The limit of the recovery lies in what a single curve determines. Without restraints the curve is reproduced at the level of the noise by parameter sets in which the positions of the polar groups differ by up to 30 %. Widths and volumes cover most of the interval allowed by the bounds. The same holds with cholesterol as one group (Appendix C.3). The curve determines the sign of the area difference of sample 7 and, when the

Bilayer		True	Best fit (converged runs)
Sample 6	δA ($\text{\AA}^2$)	-2.6	+2.1 (-8.4, +2.1)
	δz_P ($\text{\AA}$)	+0.1	+4.3 (-2.5, +4.3)
	$\delta z_{\text{choline}}$ ($\text{\AA}$)	0.0	+3.3 (+0.9, +4.3)
Sample 7, ring and chain	δA ($\text{\AA}^2$)	+5.0	+6.0 (+3.1, +11.9)
	δz_P ($\text{\AA}$)	-1.0	+1.4 (-3.5, +1.4)
	$\delta z_{\text{choline}}$ ($\text{\AA}$)	-1.1	+3.1 (-6.3, +3.1)

Table 3.11: Differences between the outer and the inner leaflet in area and in the distances of phosphate and choline from the bilayer center, at the true parameters and in the best fit, with the smallest and largest value among the converged runs; 0.1 % noise, all parameters free within ± 20 %. The true δA uses the areas of the representation (Sec. 2.15.3) and differs by 0.2–0.3 $\text{\AA}^2$ from the simulated value of Table 3.9. Other noise levels in Appendix C.2.

rest of the profile is known, the effective width of the inner cholesterol ring. It does not determine the effective widths when all parameters are free, the split between half-width and smoothing, or the differences between the leaflets in the positions of phosphate and choline.

Chapter 4

Conclusions

This thesis presents the first application and test of the ASDP model for the analysis of small-angle scattering data, providing a detailed chemical-group description of lipid membrane asymmetry along the bilayer normal (Sec. 1.3). The work covers seven samples of different composition, from a single phospholipid to two asymmetric membranes. Their SAXS curves were analyzed in a single global fit (Sec. 2.5). Some parameters, such as the volumes of the chemical groups, were shared among the curves, so as to reduce the number of free parameters.

Molecular dynamics simulations of five of the same membranes provided an independent description of their structure. The MDtoSDP program, written for this work, applied to the simulations the same treatment as the ASDP model, with the same chemical groups and the same smooth slabs. To check what the fit can recover, a closure test fitted synthetic curves whose structure was known in advance.

The model reproduces the seven curves well (Sec. 3.2.1). However, a good fit does not imply that every model parameter is robustly determined. Many parameters are not determined uniquely by the SAXS data. The clearest case is the shape of each group: its half-width and smoothing act on the curve almost exclusively through their combination, namely through the effective width (Sec. 3.3.2). Consequently, only the effective width can be meaningfully interpreted. The simulations and the closure test exhibit the same behavior.

For this reason, restraints are an essential component of the analysis. They exclude chemically implausible configurations, such as a choline group lying below the phosphate, without worsening the agreement with the data (Sec. 3.2.2). Given the high number of parameters, a fit guided by restraints does not inherently represent a weakness, especially when these restraints stem from independent measurements on the same system. However, this approach alters the interpretation of the fitted values. A parameter that settles at a predefined bound or at the threshold of a restraint derives its value from them rather than from the SAXS curves. The resulting fitted parameters therefore fall into two categories: those dictated by the bounds and restraints, and those genuinely determined by the curves (Secs. 3.3.3 and 3.3.4). Only the second category carries the structural information that can actually be extracted from the SAXS data.

The simulations also address a different question: whether chemical groups can be represented in this manner at all. For every group except one, the smooth-slab description of the volume fraction adopted by the ASDP model matches the corresponding MD-simulated distribution closely, both in sym-

metric and asymmetric membranes. The residual remains below one percent of the peak height, while the position and the effective width of the slab agree with the mean and the spread of the simulated distribution within 0.2 Å (Sec. 3.6). The exception is cholesterol. Its rigid ring possesses a significantly higher electron density than its tail, meaning that a single group with a uniform electron density cannot properly capture its volume-fraction distribution (Sec. 3.5.1). In the closure test, the model treating cholesterol as a single group fails to reproduce the SAXS curve generated from the MD simulations, whereas treating the ring and tail as separate groups succeeds (Sec. 3.8.1).

The closure test also yields one of the key methodological findings of this work. With all parameters free, the fit successfully reproduced the synthetic SAXS curves within their noise level, even when employing the incorrect description of cholesterol. However, the recovered volume fraction distributions differed widely from the true ones (Sec. 3.8.2). Consequently, achieving excellent agreement with a SAXS curve does not guarantee that the underlying volume-fraction distributions are correct.

Molecular dynamics should be viewed in the same spirit. It is not the ultimate truth against which the fit of SAXS curves is judged, since simulations also rest on their own approximations. Rather, it serves as an independent criterion because it does not rely on the SAXS curves. In this work, MD simulations served three main purposes. First, they validated the restraints: the simulated membranes respect all geometric constraints, confirming that these latter exclude only configurations that simulations themselves do not adopt. Second, they revealed incompatibilities: the simulated values for cholesterol,

the area per lipid molecule of a given sample, and the headgroup volumes fall outside the specific ranges imposed in the SAXS fit, suggesting that these bounds should be revised (Sec. 3.7.1). Third, they provided direct access to the shape of the volume fraction distributions, thereby revealing the inadequacy of treating cholesterol as a single uniform group.

The two asymmetric samples illustrate what the ASDP model can and cannot extract from SAXS data. In sample 7, the asymmetry of the simulated membrane produces a recognizable signal in the SAXS curve, roughly four times larger than in sample 6. The fit and the MD simulation agree on the qualitative trend of the asymmetry: in both cases, the outer leaflet exhibits a larger area per lipid and its polar groups lie closer to the center (Sec. 3.7.2). This consistent trend represents the most reliable result concerning asymmetry, whereas its precise numerical magnitude is dictated by the restraints. In sample 6, by contrast, a membrane like the MD simulated one produces a signal as small as that of a symmetric membrane. Consequently, the asymmetric profile derived from the SAXS fit stems primarily from the restraints. The closure test supports this finding, since from a single curve it recovers only whether the area difference is positive or negative for sample 7, rather than its exact value.

Despite these insights, several limitations remain regarding both the experimental protocol and the ASDP model. Experimentally, the exact compositions of the two leaflets were assumed from the exchange protocol rather than being directly derived from the SAXS curves (Sec. 2.2). Such compositions should ideally be confirmed by independent measurements, as a SAXS curve does not always provide sufficient sensitivity to resolve them. The simulated membrane

of sample 6 is a clear case in point: although its two leaflets differ in composition, its electron density profile along the normal, which is the quantity directly probed by X-rays, exhibits minimal asymmetry (Sec. 3.7.2).

All seven samples also contained a fraction of multilamellar vesicles. Their SAXS signal overlaps with that of the single bilayer, requiring additional parameters for its description. Furthermore, at the lowest q values, residual buffer contributions and larger aggregates introduce extra components, which complicate the data analysis and increase overall uncertainty (Sec. 2.4). Future experimental work should therefore aim exclusively at lipid dispersions of pure unilamellar vesicles, free from multilamellar structures and aggregates, to ensure a clean and isolated bilayer signal. At the same time, employing higher lipid concentrations would raise the sample scattering above the buffer background. A stronger signal would also reduce the statistical uncertainty of the data points at high q , which was notably large in the present work. Consequently, the fits required weighting via an assumed error law rather than experimental uncertainties, meaning that the current conclusions inherently rely on this weighting scheme (Sec. 2.7).

Regarding the ASDP model itself, the primary limitation stems from the large number of parameters. A finer subdivision into chemical groups yields greater structural detail, but also introduces more parameters than the experimental SAS data can uniquely determine. The grouping strategy must therefore carefully balance the level of detail required by the quantities of interest against the constraints imposed by parameter dimensionality. Furthermore, chemical groups are optimally chosen such that their electron densities (or

their scattering length densities in the case of small-angle neutron scattering, SANS) differ significantly from each other and from the surrounding solvent. A group whose scattering length density closely matches that of its environment remains difficult to resolve.

These limitations naturally outline the prospective directions for future work. The most immediate advancement involves SANS. By utilizing deuterated lipids, such as deuterated sphingomyelins, or by employing buffers with varying deuteration levels (mixture of D_2O and H_2O), the same membrane system can be investigated across multiple contrast conditions. Performing a simultaneous global fit of both SAXS and SANS curves would drastically constrain the space of viable structures compared to analyzing either data set in isolation, thereby significantly alleviating parameter degeneracy.

The ASDP framework itself can also be refined and extended along several lines. The most straightforward modification concerns the grouping strategy, which could be tailored more precisely to the specific lipid species present in the mixture. Cholesterol, in particular, should ideally be subdivided at least into its rigid ring system and flexible hydrocarbon tail. A more profound structural advancement involves refining how individual leaflet boundaries are treated. Because the terminal methyl groups of the opposing leaflets overlap at the bilayer midplane, only their sum is rigorously defined by SAS data (Sec. 3.8.1). Consequently, they could be treated explicitly as a single central domain shared by both leaflets. Furthermore, the functional forms describing the spatial distributions could be adapted directly from atomistic MD simulations to better capture non-Gaussian shapes where appropriate. Finally, the

methodology can be extended to systems of small unilamellar vesicles (SUVs). In such highly curved architectures, the membrane geometry deviates significantly from a flat slab, allowing the united chemical group distributions to directly quantify how curvature induced stress impacts nanoscale membrane structure.

The developed software tool MDtoSDP could evolve along similar lines. It will be expanded to compute SANS curves alongside SAXS profiles, as well as X-ray and neutron reflectivity curves. Dedicated plug-ins will enable testing alternative functional forms for the chemical groups. A further methodological improvement will involve the way scattering curves are computed: rather than deriving them from one-dimensional density profiles, they will be generated directly from the three-dimensional atomic coordinates of the simulation. This is crucial because large simulated membranes naturally undergo local thermal fluctuations and bending, which otherwise lead to artificially broadened distributions in a one-dimensional reduction.

In summary, the ASDP model provides a reliable tool for the interpretation of SAXS curves of both symmetric and asymmetric membranes, extracting the asymmetric volume fraction distributions of their constituent chemical groups. Nevertheless, because scattering data alone cannot uniquely constrain the entire parameter space of the ASDP model, physical restraints and independent criteria remain essential components of the methodology. Where membrane asymmetry generates a distinct scattering signature, the model successfully resolves these structural differences; where the signal is subdued, the resulting profile is primarily governed by the applied restraints. Ultimately, the ASDP

model establishes itself as an effective and rigorous approach for the structural characterization of asymmetric membranes, provided that SAS measurements are combined with independent experimental or computational validation.

Appendix A. Parameters of the final configuration

The three tables of this appendix list all the parameters of the final configuration, as mean $\pm$ standard deviation over the 26 runs of the group near the threshold (Sec. 3.2.2). Positions are distances from the midplane, and D_h follows from the area and the hydrophobic volume. The half-width and the smoothing of a group are listed separately, although the curves do not determine them separately (Sec. 3.3.2). For the terminal methyls, w and σ_{eff} describe the methyl region of the whole bilayer. The marks are those of Table 3.2. Positions are also marked ‡ when the choline or the phosphate is at the threshold of the order restraint (2.20) or the cholesterol edge at the limit of its window (2.23), and the volume of cholesterol at the lower edge of its window (Table 2.5). Spreads below 0.05 appear as 0.0 when one decimal is given.

Quantity	1	2	3	4
A ($\text{\AA}^2$)	68.7 ± 0.7	$56.9 \pm 0.4^\ddagger$	46.4 ± 0.6	$45.0 \pm 0.1^*$
D_h ($\text{\AA}$)	13.4 ± 0.1	15.6 ± 0.1	18.3 ± 0.2	17.8 ± 0.0
σ_h ($\text{\AA}$)	$3.2 \pm 0.0^*$	3.2 ± 0.0	$3.2 \pm 0.0^*$	2.5 ± 0.2
Glycerol				
z ($\text{\AA}$)	14.4 ± 0.4	16.6 ± 0.3	19.5 ± 0.6	18.8 ± 0.2
w ($\text{\AA}$)	1.8 ± 0.8	2.7 ± 0.8	2.0 ± 0.8	1.7 ± 0.8
σ ($\text{\AA}$)	3.0 ± 0.2	2.3 ± 0.3	2.9 ± 0.3	2.1 ± 0.3
σ_{eff} ($\text{\AA}$)	$3.20 \pm 0.01^\ddagger$	2.82 ± 0.17	$3.13 \pm 0.17^\ddagger$	$2.41 \pm 0.00^\ddagger$
Sphingosine backbone				
z ($\text{\AA}$)	—	15.4 ± 0.5	$17.6 \pm 0.5^*$	18.9 ± 0.2
w ($\text{\AA}$)	—	2.8 ± 0.7	1.5 ± 0.5	1.7 ± 0.6
σ ($\text{\AA}$)	—	2.6 ± 0.2	2.4 ± 0.3	2.0 ± 0.3
σ_{eff} ($\text{\AA}$)	—	3.09 ± 0.11	2.59 ± 0.29	$2.26 \pm 0.11^\ddagger$
Phosphate				
z ($\text{\AA}$)	17.7 ± 0.2	18.1 ± 0.1	21.0 ± 0.4	21.9 ± 0.2
w ($\text{\AA}$)	$1.9 \pm 0.9^*$	2.7 ± 0.7	$2.0 \pm 1.0^*$	2.8 ± 0.6
σ ($\text{\AA}$)	2.2 ± 0.4	1.5 ± 0.4	2.6 ± 0.3	2.5 ± 0.2
σ_{eff} ($\text{\AA}$)	2.49 ± 0.12	2.22 ± 0.08	2.91 ± 0.18	3.01 ± 0.10
$z + \sigma_{\text{eff}}$ ($\text{\AA}$)	20.2 ± 0.1	20.3 ± 0.1	23.9 ± 0.2	24.9 ± 0.1
Choline				
z ($\text{\AA}$)	$18.2 \pm 0.2^\ddagger$	21.1 ± 0.4	22.0 ± 0.3	22.6 ± 0.2
w ($\text{\AA}$)	1.9 ± 0.5	2.2 ± 0.9	2.0 ± 0.5	1.9 ± 0.8
σ ($\text{\AA}$)	1.3 ± 0.2	1.7 ± 0.4	1.2 ± 0.2	2.6 ± 0.7
σ_{eff} ($\text{\AA}$)	1.70 ± 0.12	2.19 ± 0.10	1.70 ± 0.11	2.93 ± 0.50
$z + \sigma_{\text{eff}}$ ($\text{\AA}$)	19.9 ± 0.1	23.3 ± 0.3	23.7 ± 0.2	25.5 ± 0.5
$\Delta_{\text{PC}}^{\text{out}}$ ($\text{\AA}$)	$-0.29 \pm 0.02^\ddagger$	3.01 ± 0.33	$-0.29 \pm 0.01^\ddagger$	0.61 ± 0.41
Unsaturated carbons				
z ($\text{\AA}$)	$6.0 \pm 0.1^*$	8.6 ± 1.0	11.9 ± 0.1	8.1 ± 0.1
w ($\text{\AA}$)	1.8 ± 0.5	1.6 ± 0.5	1.1 ± 0.1	1.4 ± 0.4
σ ($\text{\AA}$)	2.5 ± 0.1	1.9 ± 0.1	1.8 ± 0.0	1.9 ± 0.1
σ_{eff} ($\text{\AA}$)	$2.70 \pm 0.03^\ddagger$	2.13 ± 0.17	1.94 ± 0.04	2.06 ± 0.13
Terminal methyls				
w ($\text{\AA}$)	$4.2 \pm 0.0^*$	3.9 ± 0.4	4.2 ± 0.0	3.0 ± 0.1
σ ($\text{\AA}$)	2.8 ± 0.0	2.0 ± 0.3	1.4 ± 0.2	$1.3 \pm 0.5^*$
σ_{eff} ($\text{\AA}$)	3.69 ± 0.00	3.00 ± 0.13	2.81 ± 0.11	2.21 ± 0.36
Cholesterol				
z ($\text{\AA}$)	—	—	$13.0 \pm 0.2^\ddagger$	13.4 ± 0.8
w ($\text{\AA}$)	—	—	3.3 ± 0.0	$4.3 \pm 0.0^*$
σ ($\text{\AA}$)	—	—	$1.4 \pm 0.0^*$	$2.3 \pm 0.0^*$
σ_{eff} ($\text{\AA}$)	—	—	$2.38 \pm 0.00^\ddagger$	$3.38 \pm 0.00^\ddagger$

Table A.1: Parameters of profiles 1–4 in the final configuration.

Quantity	5	6 ₂	7	7 ₂
A (Å ²)	45.0 ± 0.0*	56.0 ± 0.3 [‡]	51.5 ± 0.5	59.1 ± 0.3 [‡]
D_h (Å)	17.4 ± 0.0	16.1 ± 0.1	15.9 ± 0.2	14.4 ± 0.1
σ_h (Å)	3.2 ± 0.1	3.2 ± 0.0*	3.2 ± 0.0*	3.2 ± 0.0*
Glycerol				
z (Å)	18.7 ± 0.3	16.7 ± 0.1	16.7 ± 0.2	15.2 ± 0.1
w (Å)	3.2 ± 0.0	1.3 ± 0.3	2.1 ± 0.8	1.4 ± 0.6
σ (Å)	2.0 ± 0.3	2.6 ± 0.2	2.9 ± 0.2	2.5 ± 0.2
σ_{eff} (Å)	2.75 ± 0.21	2.75 ± 0.18	3.20 ± 0.00 [‡]	2.63 ± 0.26
Sphingosine backbone				
z (Å)	18.6 ± 0.4	15.4 ± 0.2*	—	14.0 ± 0.3
w (Å)	3.1 ± 0.3*	1.3 ± 0.5	—	2.2 ± 0.9
σ (Å)	2.1 ± 0.3	2.1 ± 0.2	—	2.9 ± 0.2
σ_{eff} (Å)	2.78 ± 0.23	2.27 ± 0.13 [‡]	—	3.26 ± 0.11 [‡]
Phosphate				
z (Å)	21.1 ± 0.1	20.3 ± 0.1	19.8 ± 0.2	15.7 ± 0.1 [‡]
w (Å)	2.4 ± 0.8	3.0 ± 0.6*	2.3 ± 0.8	1.8 ± 0.7
σ (Å)	1.4 ± 0.4	1.1 ± 0.3*	2.7 ± 0.2	2.5 ± 0.5
σ_{eff} (Å)	2.07 ± 0.08 [‡]	2.07 ± 0.08 ^{‡‡}	3.08 ± 0.08 [‡]	2.75 ± 0.52 [‡]
$z + \sigma_{\text{eff}}$ (Å)	23.2 ± 0.1	22.3 ± 0.1	22.9 ± 0.2	18.5 ± 0.6
Choline				
z (Å)	22.4 ± 0.1	20.8 ± 0.1 [‡]	20.3 ± 0.2 [‡]	19.2 ± 0.1
w (Å)	3.2 ± 0.1*	1.9 ± 0.3	3.2 ± 0.0*	3.1 ± 0.1
σ (Å)	3.6 ± 0.1*	1.1 ± 0.2	3.6 ± 0.0*	3.5 ± 0.1
σ_{eff} (Å)	4.03 ± 0.06 [†]	1.58 ± 0.11	4.05 ± 0.00 [†]	3.92 ± 0.12
$z + \sigma_{\text{eff}}$ (Å)	26.5 ± 0.2	22.4 ± 0.1	24.4 ± 0.2	23.1 ± 0.1
$\Delta_{\text{PC}}^{\text{out}}$ (Å)	3.29 ± 0.16	0.02 ± 0.14	1.47 ± 0.08	4.61 ± 0.63
Unsaturated carbons				
z (Å)	11.2 ± 0.2	10.5 ± 0.1*	7.1 ± 0.1*	6.5 ± 0.0*
w (Å)	1.2 ± 0.2	1.1 ± 0.1	1.1 ± 0.1	1.9 ± 0.6
σ (Å)	1.9 ± 0.1	1.8 ± 0.0	1.8 ± 0.0	2.2 ± 0.3
σ_{eff} (Å)	1.98 ± 0.10	1.91 ± 0.03 [‡]	1.92 ± 0.03 [‡]	2.47 ± 0.20
Terminal methyls				
w (Å)	3.0 ± 0.2*	3.0 ± 0.0*	4.2 ± 0.0*	4.2 ± 0.0*
σ (Å)	1.0 ± 0.0*	1.6 ± 0.3	3.4 ± 0.0*	3.4 ± 0.0*
σ_{eff} (Å)	2.02 ± 0.08 [†]	2.38 ± 0.22	4.18 ± 0.00 [†]	4.18 ± 0.00 [†]
Cholesterol				
z (Å)	12.1 ± 0.1 [‡]	—	10.3 ± 0.1*	9.3 ± 0.0*
w (Å)	3.4 ± 0.1	—	4.3 ± 0.0*	4.3 ± 0.0*
σ (Å)	1.4 ± 0.0	—	2.3 ± 0.0*	2.3 ± 0.0*
σ_{eff} (Å)	2.41 ± 0.05 [‡]	—	3.38 ± 0.00 ^{†‡}	3.38 ± 0.00 ^{†‡}

Table A.2: Parameters of profiles 5, 6₂, 7 and 7₂ in the final configuration. Profiles 6₂ and 7₂ are the outer leaflets of samples 6 and 7, and profile 7 is the inner leaflet of sample 7.

Curve	f_{ML}	d (Å)	N_{lam}	σ_d/d (%)	$\log_{10} \kappa$	α_p	σ_{mid} (Å)
1	0.508 ± 0.003	65.12 ± 0.02	2.76 ± 0.02	$17.44 \pm 0.00^*$	-9.15 ± 0.04	3.81 ± 0.01	–
2	0.503 ± 0.002	65.35 ± 0.00	4.02 ± 0.03	$12.52 \pm 0.00^*$	-10.24 ± 0.09	3.96 ± 0.04	–
3	0.838 ± 0.000	70.28 ± 0.00	2.99 ± 0.02	$10.19 \pm 0.00^*$	-10.72 ± 0.01	$3.96 \pm 0.00^*$	–
4	0.674 ± 0.000	70.38 ± 0.00	4.20 ± 0.03	$7.70 \pm 0.00^*$	-7.87 ± 0.03	$2.83 \pm 0.00^*$	–
5	0.794 ± 0.000	67.99 ± 0.00	4.76 ± 0.03	$9.37 \pm 0.00^*$	-8.01 ± 0.04	$2.91 \pm 0.03^*$	–
6	$0.207 \pm 0.000^*$	65.35 ± 0.03	2.34 ± 0.02	0.24 ± 0.00	-14.26 ± 0.28	1.99 ± 0.04	$2.50 \pm 0.00^*$
7	$0.210 \pm 0.000^*$	$67.57 \pm 0.00^*$	3.34 ± 0.02	1.04 ± 0.00	$-6.91 \pm 0.00^*$	3.33 ± 0.00	3.25 ± 0.11

Shared parameter	Value
v_{CH_2} , methylene	27.23 ± 0.08
v_{CH_3} , terminal methyl	56.92 ± 1.06
v_{CH} , unsaturated carbon	$22.91 \pm 0.33^*$
v_{CG} , glycerol	140.04 ± 0.47
v_{amide} , sphingosine backbone without its two unsaturated carbons	$82.00 \pm 0.01^*$
v_{P} , phosphate	55.87 ± 1.61
v_{choline} , choline	147.26 ± 1.67
v_{Chol} , cholesterol	$610.98 \pm 0.00^\ddagger$
v_{w} , bulk water	$29.50 \pm 0.00^*$
relative density of the hydration water, $v_{\text{w}}/v_{\text{N}}$	1.022 ± 0.002

Table A.3: Parameters of each curve, with σ_{mid} of the two asymmetric samples, and group volumes (Å³) shared among the profiles. The hydration water has volume v_{N} , equal to the bulk value v_{w} divided by its relative density.

Appendix B. Checks of the simulations

This appendix collects the checks of the simulations and of MDtoSDP on which Secs. 3.4–3.6 rely.

B.1 Convergence

System	Start (ns)	τ_{int} (ns)	Independent samples	A ($\text{\AA}^2$)	γ (mN/m)	Drift (%)	Halves (s.e.)	Averaging (ns)
POPC	5	4.4	33	64.86 ± 0.17	-0.3 ± 0.7	0.76	0.4	200–300
POPC/EggSM 2:1	0	3.4	44	60.43 ± 0.09	-1.2 ± 1.0	0.39	0.7	100–300
POPC/EggSM/Chol 1:1:1	82	4.8	23	43.04 ± 0.06	-1.3 ± 1.9	1.16	2.3	100–300
POPC/EggSM 0.80:0.20	0	4.6	33	62.16 ± 0.15	1.1 ± 1.0	0.98	0.2	–
POPC/Chol 0.66:0.34	6	4.1	36	46.43 ± 0.07	-1.8 ± 2.1	0.74	0.3	–
POPC/EggSM/Chol 0.63:0.20:0.17	34	3.0	44	51.68 ± 0.07	-0.4 ± 1.8	0.57	0.0	–
Asymmetric, sample 6	0	3.5	43	63.48 ± 0.10	0.5 ± 1.7	0.29	0.5	0–300
Asymmetric, sample 7	115	2.8	33	49.12 ± 0.08	-1.1 ± 2.4	0.54	0.6	115–300

Table B.1: Convergence of the eight trajectories. Start of the production window, integrated autocorrelation time and number of independent samples of the area per lipid, whose mean A over the production window is given with its standard error. Surface tension γ over the whole trajectory, with the standard error from four blocks. Magnitude of the linear drift of the area relative to its mean, and distance between the means of the two halves in units of their standard error, both over 100–300 ns. Window over which group volumes and distributions are averaged, for the five systems analyzed with MDtoSDP. For the asymmetric bilayers A is the lateral area divided by the mean number of molecules per leaflet. Cholesterol counts as a molecule in all systems.

B.2 Solution for the group volumes

With all radii set equal, the tessellation of MDtoSDP reproduces the volumes computed with freud within 0.01–0.02 % for the molecules and 0.19 % for the ions on the development system of Sec. 2.15.2. On the fine partition the geometric terms, the recall and the volumes imposed at their geometric value, lower the condition number from 66–421 to 20–174 (Table B.2). The final volumes lie within 1.7 tolerances of their geometric values, except the terminal methyls of POPC and of the 2:1 bilayer and the sphingosine glycerol of the asymmetric bilayer of sample 7, at 3.0, 2.7 and 2.0 tolerances.

System	Free volumes	Condition number		Effective degrees of freedom	rms of $\phi_{\text{tot}} - 1$ (%)			Volume balance (%)
		without geometric terms	with geometric terms		without geometric terms	with geometric terms	geometric volumes	
POPC	6	66	63	5.84	0.19	0.19	2.79	0.05
POPC/EggSM 2:1	8	421	174	5.64	0.19	0.19	2.77	0.10
POPC/EggSM/Chol 1:1:1	7	129	20	4.82	0.63	1.38	3.58	0.27
Asymmetric, sample 6	7	364	68	5.64	0.26	0.26	2.57	0.08
Asymmetric, sample 7	8	210	35	6.07	0.43	0.70	3.07	0.10

Table B.2: Solution of (2.33) on the fine partition. Condition number and root-mean-square deviation of ϕ_{tot} from one without and with the geometric terms (recall and imposed values); effective degrees of freedom determined by the filling condition; deviation with the volumes of the radical tessellation alone; relative difference between the sum of the molecular volumes and the volume of the box.

B.3 Electron density and amplitudes

The electron density computed directly from the trajectory and the one recombined from the distributions of the fine partition differ in the water by 1.06–1.14 %, the contribution of the ions excluded from the partition. Their largest difference along the normal lies at the bilayer center, where ϕ_{tot} deviates most from one. It is 1.0–1.1 % of the peak electron density in the systems without cholesterol, 1.7 % in the asymmetric bilayer of sample 7 and 2.6 % in the 1:1:1 bilayer. Computed between the square

roots of the two intensities, R_F (2.35) is 0.018–0.022 in the q -range 0.08–0.5 $\text{\AA}^{-1}$, 0.10–0.18 in 0.03–0.08 $\text{\AA}^{-1}$ and 0.02–0.29 in 0.005–0.03 $\text{\AA}^{-1}$, where the intensity depends on the contrast integrated over the bilayer.

On a profile of erf slabs the quadrature (2.36) agrees with the closed form (2.6) within a few parts in 10^{16} of the maximum of $|F(q)|$. Figure B.1 compares the amplitudes computed from the distributions of the groups of the model and from their representation, which overlap except near the minima.

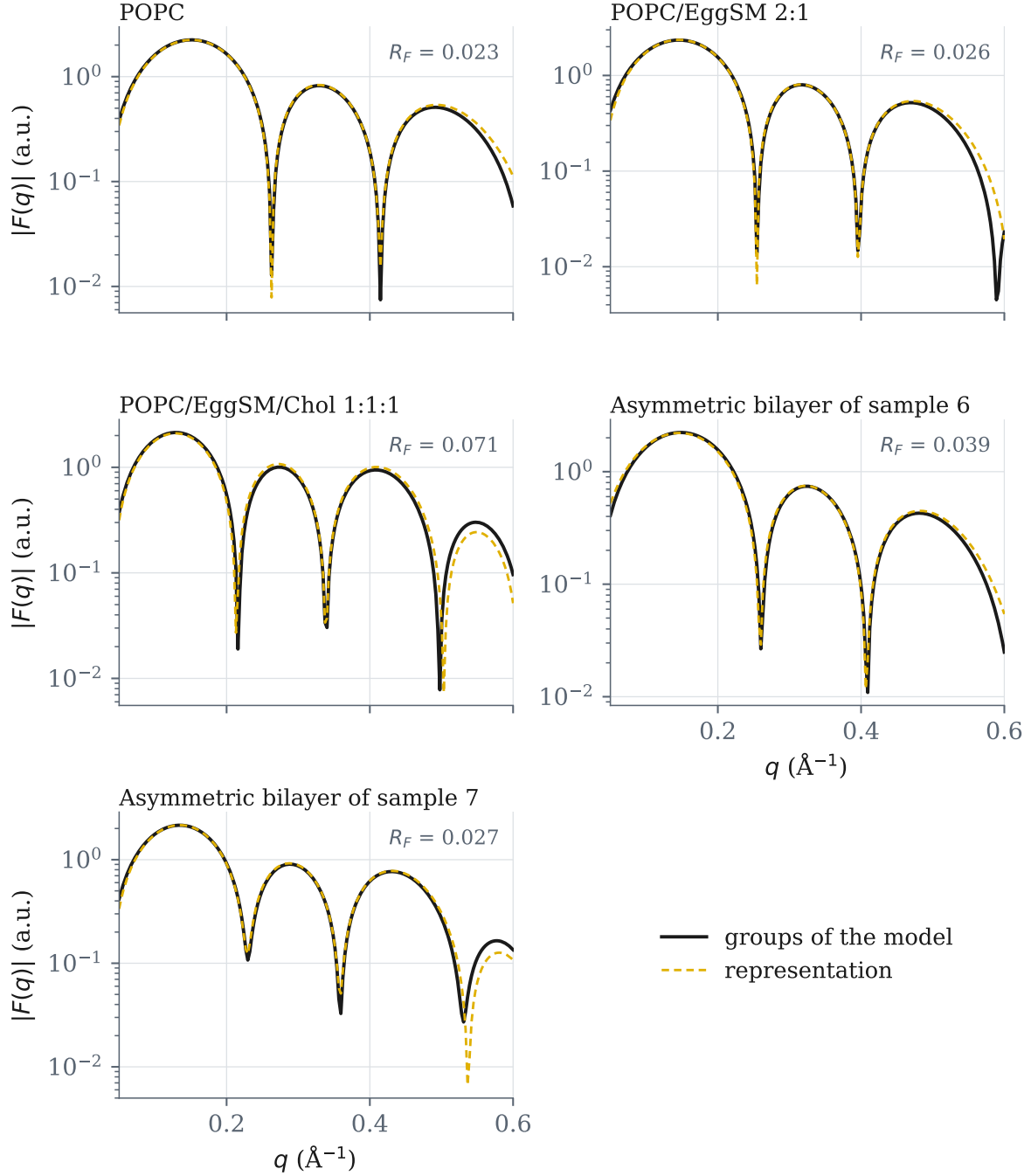


Figure B.1: Modulus of the amplitude $|F(q)|$ computed from the simulated distributions of the groups of the model (solid) and from their representation (dashed), on a common scale for each system; R_F is given in each plot.

Appendix C. Closure test

This appendix completes Sec. 3.8 with the convergence of the minimizer, the other noise levels and bounds, and sample 7 with cholesterol as one group.

C.1 Convergence

With all parameters free within $\pm 20\%$, 1 % noise and $T_i = 10^{-3}$, the initial temperature of the experimental fits, 19 and 22 of 30 runs converge for samples 6 and 7. The others end in a minimum where almost all volumes and areas sit together at their upper bound. There the hydrophobic thickness is correct but the electron densities of the groups are not, and χ^2_{data} is 552 and 413. With $T_i = 0.1$, 20 of 30 runs converge for both samples. With $T_i = 10$ and 1000 all 30 converge, and the median χ^2_{data} of the converged runs is that at 10^{-3} within 0.4 %.

C.2 Noise levels and bounds

Table C.1 extends Table 3.10 to the three noise levels. As the noise decreases, χ^2_{data} at the true parameters grows much faster for sample 7 with cholesterol as one group than with ring and chain, because the error of the functional form stays while the noise shrinks.

Bilayer	Parameters	Noise on the first point		
		1 %	0.3 %	0.1 %
Sample 6	true	1.22	2.58	16.2
	methyls free	1.16	1.20	1.52
	all free	1.13	1.16	1.15
Sample 7, cholesterol as one group	true	7.80	81.7	706
	methyls and cholesterol free	2.50	19.9	161
	all free	1.05	1.14	1.14
Sample 7, cholesterol as ring and chain	true	1.10	1.55	3.20
	methyls, ring and chain free	1.06	1.22	1.46
	all free	1.05	1.13	1.14

Table C.1: χ^2_{data} of the closure test at the three noise levels, as in Table 3.10.

With all parameters free, the recovery does not improve as the noise decreases. At 1 % noise the median range of D_h and A among the converged runs is 13–19 % for sample 6 and sample 7 with ring and chain, against 14–18 % at 0.1 %. That of the positions of the polar groups is 22–25 %, against 16–19 %. At every noise level the area difference of sample 7 keeps its sign in all runs, that of sample 6 changes sign among the runs, and the differences in the positions of phosphate and choline spread over 4–11 Å. With the rest of the profile at its true values, the effective width of the inner cholesterol ring stays within 2.5 and 2.3 % of the true value at 1 and 0.3 % noise.

Narrower bounds bring more parameters to a bound. The fraction of parameter values at a bound grows from 28–48 % at $\pm 20\%$ to 35–53 % at $\pm 10\%$ and 40–72 % at $\pm 5\%$. At $\pm 20\%$ volumes and areas are at a bound in at most 20 and 10 % of the values, half-widths and smoothings in 40–93 %.

C.3 Cholesterol as one group

With cholesterol as one group, the best fit at 0.1 % noise reproduces the curve at the level of the noise (Table 3.10). In the inner leaflet its cholesterol slab is narrower and denser than the simulated one, and its glycerol is displaced outward (Figure C.1). Its parameters deviate from the true ones by as much as with ring and chain.

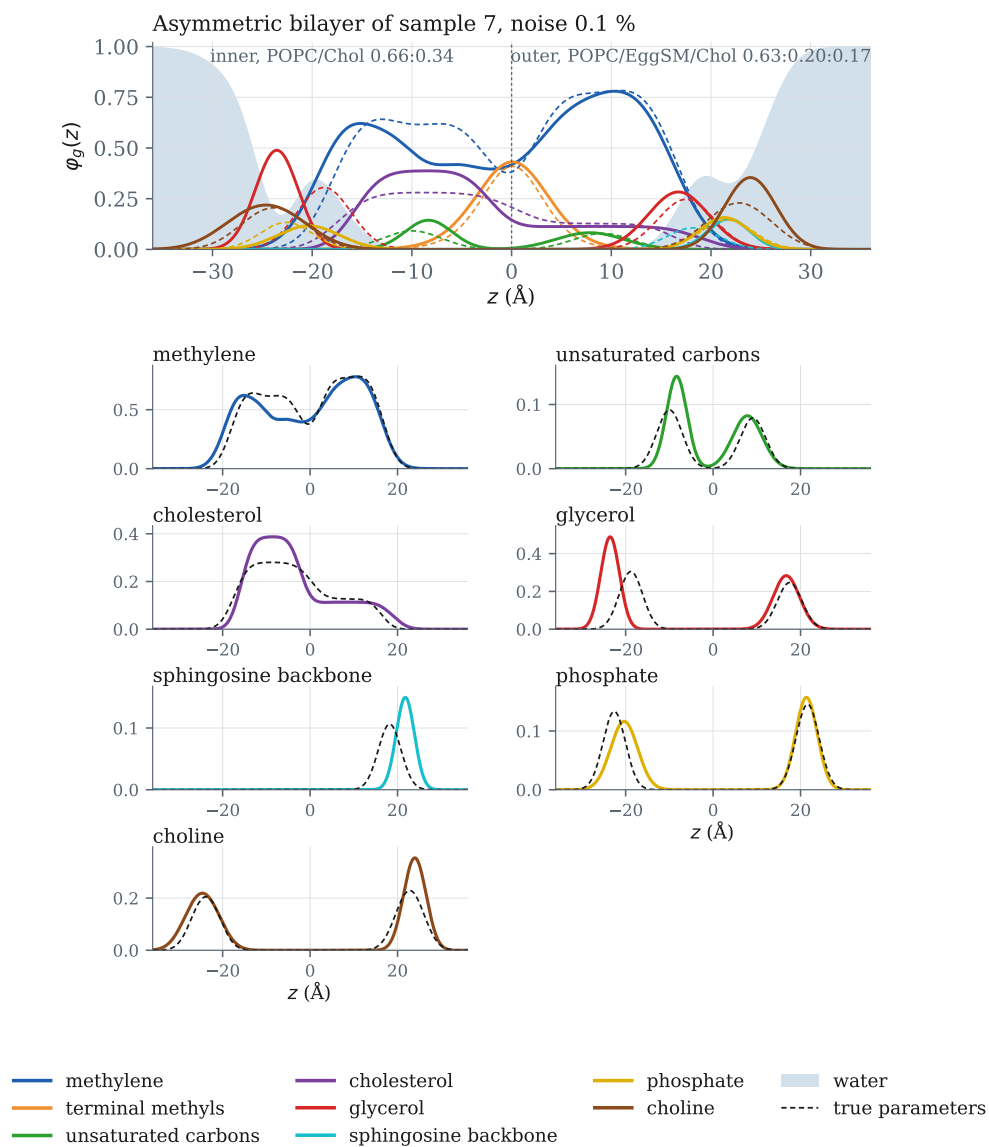


Figure C.1: Volume-fraction profiles of the best fit of the closure test (solid) and at the true parameters (dashed) for the asymmetric bilayer of sample 7 with cholesterol as one group, 0.1 % noise and bounds at ± 20 %, as in Figure 3.9.

List of symbols

Symbol	Meaning	Defined in
<i>Scattering</i>		
λ	wavelength of the radiation	Sec. 1.2
$\mathbf{k}, \mathbf{k}_0$	wave vectors of the scattered and of the incident wave	Eq. (1.1)
$\mathbf{q}, q$	scattering vector and its modulus	Eqs. (1.1), (1.2)
2θ	scattering angle	Eq. (1.2)
$q_{\min}, q_{\max}$	limits of the q -range of a curve	Table 2.3
δ	nominal spatial resolution, $2\pi/q_{\max}$	Eq. (2.1)
ρ_e	electron density	Eq. (1.5)
ρ_0	density of the solvent (electron density, or SLD in the model)	Eq. (1.5)
$F_p(\mathbf{q})$	amplitude scattered by one particle	Eq. (1.5)
$d\Sigma/d\Omega$	differential scattering cross section per unit volume	Eq. (1.6)
n_p	number of particles per unit volume	Eq. (1.6)
$\langle \cdot \rangle_\omega$	average over orientations	Eq. (1.6)
R	radius of the disk-shaped membrane	Eq. (1.7)
ψ	angle between $\mathbf{q}$ and the membrane normal	Eq. (1.7)
J_1	Bessel function of the first kind and order one	Eq. (1.7)
$F(q)$	amplitude of the bilayer (form factor)	Eqs. (1.7), (2.11)
A_p	area of the particle in its plane	Eq. (1.8)
ϕ_L	volume fraction of lipid in the sample	Eq. (1.9)
$d_B, d_{B,\alpha}$	Luzzati thickness of the bilayer and of monolayer α	Eqs. (1.9), (2.12)
R_g	radius of gyration of a vesicle	Sec. 2.4
<i>The ASDP model</i>		
z	coordinate along the normal, origin at the midplane	Sec. 2.6
α	monolayer index, $\alpha = 1$ for $z < 0$, $\alpha = 2$ for $z > 0$	Sec. 2.6
j, N_S	index and number of molecular species	Sec. 2.6
$x_{j,\alpha}$	fraction of the molecules of species j in monolayer α	Sec. 2.6
C_j	overall molar concentration of species j	Sec. 2.6

Symbol	Meaning	Defined in
$\zeta_{j,\alpha}$	stoichiometric coefficient in the molecular unit	Eq. (2.2)
A_α	area per molecular unit of monolayer α	Sec. 2.6
$A^{(i)}$	area per molecular unit of transverse profile i	Eq. (2.19)
D_α, V_α	thickness and volume of the hydrated molecular unit	Sec. 2.6
$V_{h,\alpha}, V_{p,\alpha}, V_{w,\alpha}$	volumes of the hydrocarbon, polar and water regions	Sec. 2.6
$V_{\text{dry},\alpha}$	dry volume of the molecular unit	Eq. (2.12)
g	index of a chemical group	Sec. 2.6
N_h, N_p, N	numbers of apolar and polar groups; index of the hydration water	Sec. 2.6
$v_{g,\alpha}$	volume of group g	Sec. 2.6
$b_{g,\alpha}$	scattering length of group g	Sec. 2.6
$\rho_{g,\alpha}$	scattering length density (SLD) of group g	Sec. 2.6
$n_{g,\alpha}, m_{g,j}$	number of groups g in the molecular unit and in molecule j	Sec. 2.6
$\phi_{g,\alpha}(z)$	volume-fraction distribution of group g	Eqs. (2.3), (2.7)
E_2	double error function	Eq. (2.5)
$z_{g,\alpha}, w_{g,\alpha}, \sigma_{g,\alpha}$	position, half-width and smoothing of group g	Eq. (2.5)
$\phi_{h,\alpha}(z)$	hydrophobic envelope	Eq. (2.8)
$D_{h,\alpha}, \sigma_{h,\alpha}$	hydrophobic thickness and its smoothing	Eq. (2.8)
σ_{mid}	interpenetration of the chains at the midplane	Eq. (2.8)
$H_\alpha(z)$	indicator of monolayer α	Eq. (2.10)
f_g	position of group g as a fraction of $D_{h,\alpha}$	Sec. 2.8
$I(q)$	computed intensity of one contribution	Eq. (2.12)
K	constant that fixes the units of the computed intensity	Eq. (2.12)
$S(q)$	lamellar structure factor	Eq. (2.13)
d, σ_d	lamellar repeat distance and its fluctuation	Eq. (2.13)
$N_{\text{lam}}, g_{N_{\text{lam}},n}$	number of lamellae; weight of pairs n repeats apart	Eq. (2.13)
η, η_q	Caillé parameter and its q -dependent form	Eq. (2.13)
γ_E	Euler–Mascheroni constant	Eq. (2.13)
f_{ML}	weight of the multilamellar contribution	Eq. (2.14)
κ, α_p	scale and exponent of the low- q power law	Eq. (2.14)
σ_{eff}	effective width of a group	Eq. (2.24)
$\Delta\sigma_{\text{eff},g}$	spread of σ_{eff} of group g across the profiles	Sec. 2.8.4
$v_{\text{PC}}^{\text{head}}, v_{\text{SM}}^{\text{head}}$	volumes of the phosphocholine headgroups	Sec. 2.8
$\Delta_{\text{PC}}^{\text{out}}$	distance between the outer edges of choline and phosphate	Eq. (2.25)
$\delta A, \delta z_P, \delta z_{\text{choline}}$	differences between the outer and the inner leaflet in area and in the distances of phosphate and choline from the center	Table 3.9
<i>Fit and identifiability</i>		
σ_i, I_i	uncertainty and intensity of point i	Eq. (2.15)
k_e, I_1	relative error and intensity of the reference point	Eq. (2.15)
β	exponent of the error law	Eq. (2.15)

Symbol	Meaning	Defined in
χ_{data}^2	data term of the objective function	Eq. (2.16)
$e, n_{q,e}$	index of an experiment and its number of grid points	Eq. (2.16)
W, W_+	flat-bottom window and one-sided threshold	Eq. (2.17)
L, U, s_0	bounds of the allowed region and scale of a restraint	Eq. (2.17)
Φ	objective function	Eq. (2.18)
λ_k, P_k	weight and penalty of restraint family k	Eq. (2.18)
$c_{k,j}, n_k$	cost of a residual of family k ; number of its residuals	Eq. (2.18)
T_j, T_i, T_f	annealing temperatures: level j , initial, final	Eq. (2.27)
N_T, N_c	number of temperature levels and of cycles per level	Eq. (2.27)
$\mathbf{J}$	Jacobian of the residuals with respect to the parameters	Sec. 2.11
$\mathbf{U}, \mathbf{V}, s_k$	factors and singular values of the decomposition of $\mathbf{J}$	Sec. 2.11
r	correlation coefficient between two parameters across the runs	Figure 3.5
<i>Molecular dynamics and MDtoSDP</i>		
$U(\mathbf{r}^N)$	potential energy of the force field; the terms are defined below it	Eq. (1.10)
$N_{\text{in}}, N_{\text{out}}$	numbers of lipids of the inner and outer leaflet	Eq. (2.29)
$A_{\text{in}}, A_{\text{out}}$	equilibrium areas per lipid of the two leaflet compositions	Eq. (2.29)
γ	mechanical surface tension	Sec. 2.13.5
$L_z, P_{xx}, P_{yy}, P_{zz}$	box length along the normal; diagonal of the pressure tensor	Sec. 2.13.5
τ_{int}	integrated autocorrelation time	Sec. 2.13.5
$\rho_g^{\text{m}}(z)$	mass-density profile of group g	Sec. 2.15.2
u_g, μ_g	specific volume and mass of the unit of group g	Sec. 2.15.2
$\tilde{\phi}_g(z), \phi_{\text{tot}}(z)$	unnormalized volume fraction and filling factor	Eq. (2.30)
z_k, n_z	grid points along the normal and their number	Eq. (2.31)
$\mathbf{X}$	matrix of the filling system	Eq. (2.31)
π_i, r_i	power distance and van der Waals radius of atom i	Eq. (2.32)
Φ_V	function minimized for the volumes	Eq. (2.33)
$\ell, V_\ell, M_{g,\ell}$	molecule index, its volume, mass of group g in it	Eq. (2.33)
$\mathcal{T}, \tau_g, \omega_g$	groups with geometric recall, tolerance, weight	Eq. (2.33)
u_g^{geom}	specific volume from the radical tessellation	Eq. (2.33)
σ_ϕ	root-mean-square deviation of ϕ_{tot} from one	Eq. (2.33)
p	proportion that places a slab between its extremes	Eq. (2.34)
R_F, c	discrepancy factor between amplitudes and its scale	Eq. (2.35)
$\sigma(q), a$	uncertainty of a synthetic point and conversion scale	Sec. 2.15.4

Bibliography

- [1] Gerrit van Meer, Dennis R. Voelker, and Gerald W. Feigenson. Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology*, 9(2):112–124, 2008. doi: 10.1038/nrm2330.
- [2] Drew Marquardt, Barbara Geier, and Georg Pabst. Asymmetric lipid membranes: towards more realistic model systems. *Membranes*, 5(2): 180–196, 2015. doi: 10.3390/membranes5020180.
- [3] Jacob N. Israelachvili. *Intermolecular and Surface Forces*. Elsevier, Amsterdam, 3 edition, 2011. doi: 10.1016/C2009-0-21560-1.
- [4] Jacob N. Israelachvili, D. John Mitchell, and Barry W. Ninham. Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers. *Journal of the Chemical Society, Faraday Transactions 2*, 72:1525–1568, 1976. doi: 10.1039/f29767201525.
- [5] Rodrigo F.M. de Almeida, Aleksandre Fedorov, and Manuel Prieto. Sphingomyelin/phosphatidylcholine/cholesterol phase diagram: boundaries and composition of lipid rafts. *Biophysical Journal*, 85(4):2406–2416, 2003. doi: 10.1016/S0006-3495(03)74664-5.
- [6] Ilya Levental and Ed Lyman. Regulation of membrane protein structure and function by their lipid nano-environment. *Nature Reviews Molecular Cell Biology*, 24(2):107–122, 2023. doi: 10.1038/s41580-022-00524-4.
- [7] Takeshi Harayama and Howard Riezman. Understanding the diversity of membrane lipid composition. *Nature Reviews Molecular Cell Biology*, 19(5):281–296, 2018. doi: 10.1038/nrm.2017.138.
- [8] Domenico Lombardo and Mikhail A. Kiselev. Methods of liposomes preparation: formation and control factors of versatile nanocarriers for biomedical and nanomedicine application. *Pharmaceutics*, 14(3):543, 2022. doi: 10.3390/pharmaceutics14030543.
- [9] Stéphanie Andrade, Maria J. Ramalho, Joana A. Loureiro, and Maria Carmo Pereira. Liposomes as biomembrane models: biophysical

- techniques for drug-membrane interaction studies. *Journal of Molecular Liquids*, 334:116141, 2021. doi: 10.1016/j.molliq.2021.116141.
- [10] Jay R. Werber, Chinedum O. Osuji, and Menachem Elimelech. Materials for next-generation desalination and water purification membranes. *Nature Reviews Materials*, 1(5):16018, 2016. doi: 10.1038/natrevmats.2016.18.
 - [11] Milka Doktorova, Ilya Levental, and Frederick A. Heberle. Seeing the membrane from both sides now: lipid asymmetry and its strange consequences. *Cold Spring Harbor Perspectives in Biology*, 15(8):a041393, 2023. doi: 10.1101/cshperspect.a041393.
 - [12] Joseph H. Lorent, Kandice R. Levental, Lukmanee Ganesan, Gordon Rivera-Longsworth, Erdinc Sezgin, Milka Doktorova, Edward Lyman, and Ilya Levental. Plasma membranes are asymmetric in lipid unsaturation, packing and protein shape. *Nature Chemical Biology*, 16(6): 644–652, 2020. doi: 10.1038/s41589-020-0529-6.
 - [13] Yee-Hung M. Chan and Steven G. Boxer. Model membrane systems and their applications. *Current Opinion in Chemical Biology*, 11(6):581–587, 2007. doi: 10.1016/j.cbpa.2007.09.020.
 - [14] Alessandra Luchini and Giuseppe Vitiello. Mimicking the mammalian plasma membrane: an overview of lipid membrane models for biophysical studies. *Biomimetics*, 6(1):3, 2021. doi: 10.3390/biomimetics6010003.
 - [15] Rumiana Dimova. Giant vesicles and their use in assays for assessing membrane phase state, curvature, mechanics, and electrical properties. *Annual Review of Biophysics*, 48:93–119, 2019. doi: 10.1146/annurev-biophys-052118-115342.
 - [16] Milka Doktorova, Frederick A. Heberle, Barbara Eicher, Robert F. Standardt, John Katsaras, Erwin London, Georg Pabst, and Drew Marquardt. Preparation of asymmetric phospholipid vesicles for use as cell membrane models. *Nature Protocols*, 13(9):2086–2101, 2018. doi: 10.1038/s41596-018-0033-6.
 - [17] Robin S. Petruzielo, Frederick A. Heberle, Paul Drazba, John Katsaras, and Gerald W. Feigenson. Phase behavior and domain size in sphingomyelin-containing lipid bilayers. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1828(4):1302–1313, 2013. doi: 10.1016/j.bbamem.2013.01.007.

- [18] Theyencheri Narayanan, Michael Sztucki, Thomas Zinn, Jérôme Kieffer, Alejandro Homs-Puron, Jacques Gorini, Pierre Van Vaerenbergh, and Peter Boesecke. Performance of the time-resolved ultra-small-angle X-ray scattering beamline with the Extremely Brilliant Source. *Journal of Applied Crystallography*, 55(1):98–111, 2022. doi: 10.1107/S1600576721012693.
- [19] L. A. Feigin and D. I. Svergun. *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*. Springer US, Boston, MA, 1987. doi: 10.1007/978-1-4757-6624-0.
- [20] William Lawrence Bragg. The diffraction of short electromagnetic waves by a crystal. *Proceedings of the Cambridge Philosophical Society*, 17: 43–57, 1913.
- [21] Doris Orthaber, Alexander Bergmann, and Otto Glatter. SAXS experiments on absolute scale with kratky systems using water as a secondary standard. *Journal of Applied Crystallography*, 33(2):218–225, 2000. doi: 10.1107/S0021889899015216.
- [22] John F. Nagle and Stephanie Tristram-Nagle. Structure of lipid bilayers. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes*, 1469(3):159–195, 2000. doi: 10.1016/S0304-4157(00)00016-2.
- [23] John F. Nagle, Richard M. Venable, Ezekiel Maroclo-Kemmerling, Stephanie Tristram-Nagle, Paul E. Harper, and Richard W. Pastor. Revisiting volumes of lipid components in bilayers. *The Journal of Physical Chemistry B*, 123(12):2697–2709, 2019. doi: 10.1021/acs.jpcb.8b12010.
- [24] Norbert Kučerka, Frederick A. Heberle, Jianjun Pan, and John Katsaras. Structural significance of lipid diversity as studied by small angle neutron and X-ray scattering. *Membranes*, 5(3):454–472, 2015. doi: 10.3390/membranes5030454.
- [25] Barbara Eicher, Frederick A. Heberle, Drew Marquardt, Gerald N. Reicherger, John Katsaras, and Georg Pabst. Joint small-angle X-ray and neutron scattering data analysis of asymmetric lipid vesicles. *Journal of Applied Crystallography*, 50(2):419–429, 2017. doi: 10.1107/S1600576717000656.
- [26] Frederick A. Heberle, Drew Marquardt, Milka Doktorova, Barbara Geier, Robert F. Standaert, Peter Heftberger, Benjamin Kollmitzer, Jonathan D. Nickels, Robert A. Dick, Gerald W. Feigenson, John Katsaras, Erwin

- London, and Georg Pabst. Subnanometer structure of an asymmetric model membrane: interleaflet coupling influences domain properties. *Langmuir*, 32(20):5195–5200, 2016. doi: 10.1021/acs.langmuir.5b04562.
- [27] Raffaella De Rosa, Francesco Spinozzi, and Rosangela Itri. Hydroperoxide and carboxyl groups preferential location in oxidized biomembranes experimentally determined by small angle X-ray scattering: implications in membrane structure. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1860(11):2299–2307, 2018. doi: 10.1016/j.bbamem.2018.05.011.
- [28] Francesco Spinozzi, Claudio Ferrero, Maria Grazia Ortore, Alejandro De Maria Antolinos, and Paolo Mariani. GENFIT: software for the analysis of small-angle X-ray and neutron scattering data of macromolecules in solution. *Journal of Applied Crystallography*, 47(3):1132–1139, 2014. doi: 10.1107/S1600576714005147.
- [29] Ruitian Zhang, Robert M. Suter, and John F. Nagle. Theory of the structure factor of lipid bilayers. *Physical Review E*, 50(6):5047–5060, 1994. doi: 10.1103/PhysRevE.50.5047.
- [30] Thomas Fröhwrth, Gerhard Fritz, Norbert Freiburger, and Otto Glatter. Structure and order in lamellar phases determined by small-angle scattering. *Journal of Applied Crystallography*, 37(5):703–710, 2004. doi: 10.1107/S0021889804012956.
- [31] Jeffery B. Klauda, Richard M. Venable, J. Alfredo Freites, Joseph W. O’Connor, 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. *The Journal of Physical Chemistry B*, 114(23):7830–7843, 2010. doi: 10.1021/jp101759q.
- [32] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, 1998. doi: 10.1021/jp973084f.

- [33] Norbert Kučerka, John Katsaras, and John F. Nagle. Comparing membrane simulations to scattering experiments: introducing the SIMtoEXP software. *Journal of Membrane Biology*, 235(1):43–50, 2010. doi: 10.1007/s00232-010-9254-5.
- [34] Norbert Kučerka, John F. Nagle, Jonathan N. Sachs, Scott E. Feller, Jeremy Pencer, Andrew Jackson, and John Katsaras. Lipid bilayer structure determined by the simultaneous analysis of neutron and X-ray scattering data. *Biophysical Journal*, 95(5):2356–2367, 2008. doi: 10.1529/biophysj.108.132662.
- [35] Milka Doktorova and Harel Weinstein. Accurate in silico modeling of asymmetric bilayers based on biophysical principles. *Biophysical Journal*, 115(9):1638–1643, 2018. doi: 10.1016/j.bpj.2018.09.008.
- [36] Soohyung Park, Wonpil Im, and Richard W. Pastor. Developing initial conditions for simulations of asymmetric membranes: a practical recommendation. *Biophysical Journal*, 120(22):5041–5059, 2021. doi: 10.1016/j.bpj.2021.10.009.
- [37] M.J. Hope, M.B. Bally, G. Webb, and P.R. Cullis. Production of large unilamellar vesicles by a rapid extrusion procedure. characterization of size distribution, trapped volume and ability to maintain a membrane potential. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 812(1): 55–65, 1985. doi: 10.1016/0005-2736(85)90521-8.
- [38] Robert C. MacDonald, Ruby I. MacDonald, Bert Ph.M. Menco, Keizo Takeshita, Nanda K. Subbarao, and Lan-rong Hu. Small-volume extrusion apparatus for preparation of large, unilamellar vesicles. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1061(2):297–303, 1991. doi: 10.1016/0005-2736(91)90295-J.
- [39] Hui-Ting Cheng, Megha, and Erwin London. Preparation and properties of asymmetric vesicles that mimic cell membranes. *Journal of Biological Chemistry*, 284(10):6079–6092, 2009. doi: 10.1074/jbc.M806077200.
- [40] Yuli Zhu, Lionel Porcar, Thirupathi Ravula, Krishna C. Batchu, Tera L. Lavoie, Ying Liu, and Ursula Perez-Salas. Unexpected asymmetric distribution of cholesterol and phospholipids in equilibrium model membranes. *Biophysical Journal*, 123(22):3923–3934, 2024. doi: 10.1016/j.bpj.2024.10.004.
- [41] Gašper Šolinc, Marija Srnko, Franci Merzel, et al. Cryo-EM structures of a protein pore reveal a cluster of cholesterol molecules and diverse

- roles of membrane lipids. *Nature Communications*, 16:2972, 2025. doi: 10.1038/s41467-025-58334-z.
- [42] Veronica Chappa, Yuliya Smirnova, Karlo Komorowski, Marcus Müller, and Tim Salditt. The effect of polydispersity, shape fluctuations and curvature on small unilamellar vesicle small-angle X-ray scattering curves. *Journal of Applied Crystallography*, 54:557–568, 2021. doi: 10.1107/S1600576721001461.
 - [43] Jesse Bennett Hopkins, Richard E. Gillilan, and Soren Skou. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *Journal of Applied Crystallography*, 50(5):1545–1553, 2017. doi: 10.1107/S1600576717011438.
 - [44] Michael Sztucki. SAXSutilities2: a graphical user interface for processing and analysis of small-angle X-ray scattering data. software, 2021.
 - [45] Alexander I. Greenwood, Stephanie Tristram-Nagle, and John F. Nagle. Partial molecular volumes of lipids and cholesterol. *Chemistry and Physics of Lipids*, 143(1–2):1–10, 2006. doi: 10.1016/j.chemphyslip.2006.04.002.
 - [46] Zoran Arsov, Emilio J. González-Ramírez, Felix M. Goñi, Stephanie Tristram-Nagle, and John F. Nagle. Phase behavior of palmitoyl and egg sphingomyelin. *Chemistry and Physics of Lipids*, 213:102–110, 2018. doi: 10.1016/j.chemphyslip.2018.03.003.
 - [47] Moritz P. K. Frewein, Milka Doktorova, Frederick A. Heberle, Haden L. Scott, Enrico F. Semeraro, Lionel Porcar, and Georg Pabst. Structure and interdigitation of chain-asymmetric phosphatidylcholines and milk sphingomyelin in the fluid phase. *Symmetry*, 13(8):1441, 2021. doi: 10.3390/sym13081441.
 - [48] Milka Doktorova, Norbert Kučerka, Jacob J. Kinnun, Jianjun Pan, Drew Marquardt, Haden L. Scott, Richard M. Venable, Richard W. Pastor, Stephen R. Wassall, John Katsaras, and Frederick A. Heberle. Molecular structure of sphingomyelin in fluid phase bilayers determined by the joint analysis of small-angle neutron and X-ray scattering data. *The Journal of Physical Chemistry B*, 124(25):5186–5200, 2020. doi: 10.1021/acs.jpcb.0c03389.
 - [49] Drew Marquardt, Norbert Kučerka, Stephen R. Wassall, Thad A. Harroun, and John Katsaras. Cholesterol’s location in lipid bilayers. *Chemistry and Physics of Lipids*, 199:17–25, 2016. doi: 10.1016/j.chemphyslip.2016.04.001.

- [50] Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11):1859–1865, 2008. doi: 10.1002/jcc.20945.
- [51] Sunhwan Jo, Joseph B. Lim, Jeffery B. Klauda, and Wonpil Im. CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophysical Journal*, 97(1):50–58, 2009. doi: 10.1016/j.bpj.2009.04.013.
- [52] Emilia L. Wu, Xi Cheng, Sunhwan Jo, Huan Rui, Kevin C. Song, Eder M. Dávila-Contreras, Yifei Qi, Jumin Lee, Viviana Monje-Galvan, Richard M. Venable, Jeffery B. Klauda, and Wonpil Im. CHARMM-GUI membrane builder toward realistic biological membrane simulations. *Journal of Computational Chemistry*, 35(27):1997–2004, 2014. doi: 10.1002/jcc.23702.
- [53] Jumin Lee, Xi Cheng, Jason M. Swails, Min Sun Yeom, Peter K. Eastman, Justin A. Lemkul, Shuai Wei, Joshua Buckner, Jong Cheol Jeong, Yifei Qi, Sunhwan Jo, Vijay S. Pande, David A. Case, Charles L. Brooks, Alexander D. MacKerell, Jeffery B. Klauda, and Wonpil 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. doi: 10.1021/acs.jctc.5b00935.
- [54] Richard M. Venable, Alexander J. Sodt, Brent Rogaski, Huan Rui, Elizabeth Hatcher, Alexander D. MacKerell, Richard W. Pastor, and Jeffery B. Klauda. CHARMM all-atom additive force field for sphingomyelin: elucidation of hydrogen bonding and of positive curvature. *Biophysical Journal*, 107(1):134–145, 2014. doi: 10.1016/j.bpj.2014.05.034.
- [55] Joseph B. Lim, Brent Rogaski, and Jeffery B. Klauda. Update of the cholesterol force field parameters in CHARMM. *The Journal of Physical Chemistry B*, 116(1):203–210, 2012. doi: 10.1021/jp207925m.
- [56] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L. de Groot, Helmut Grubmüller, and Alexander D. MacKerell. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14(1):71–73, 2017. doi: 10.1038/nmeth.4067.
- [57] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffery D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential

- functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983. doi: 10.1063/1.445869.
- [58] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2:19–25, 2015. doi: 10.1016/j.softx.2015.06.001.
 - [59] Szilárd Páll, Artem Zhmurov, Paul Bauer, Mark Abraham, Magnus Lundborg, Alan Gray, Berk Hess, and Erik Lindahl. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *The Journal of Chemical Physics*, 153(13):134110, 2020. doi: 10.1063/5.0018516.
 - [60] Scott E. Feller and Richard W. Pastor. Constant surface tension simulations of lipid bilayers: the sensitivity of surface areas and compressibilities. *The Journal of Chemical Physics*, 111(3):1281–1287, 1999. doi: 10.1063/1.479313.
 - [61] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007. doi: 10.1063/1.2408420.
 - [62] Mattia Bernetti and Giovanni Bussi. Pressure control using stochastic cell rescaling. *The Journal of Chemical Physics*, 153(11):114107, 2020. doi: 10.1063/5.0020514.
 - [63] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993. doi: 10.1063/1.464397.
 - [64] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995. doi: 10.1063/1.470117.
 - [65] Peter J. Steinbach and Bernard R. Brooks. New spherical-cutoff methods for long-range forces in macromolecular simulation. *Journal of Computational Chemistry*, 15(7):667–683, 1994. doi: 10.1002/jcc.540150702.
 - [66] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. LINCS: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997. doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

- [67] Berk Hess. P-LINCS: a parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation*, 4(1):116–122, 2008. doi: 10.1021/ct700200b.
- [68] Shuichi Miyamoto and Peter A. Kollman. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, 1992. doi: 10.1002/jcc.540130805.
- [69] John D. Chodera. A simple method for automated equilibration detection in molecular simulations. *Journal of Chemical Theory and Computation*, 12(4):1799–1805, 2016. doi: 10.1021/acs.jctc.5b00784.
- [70] Horia I. Petrache, Scott E. Feller, and John F. Nagle. Determination of component volumes of lipid bilayers from simulations. *Biophysical Journal*, 72(5):2237–2242, 1997. doi: 10.1016/S0006-3495(97)78867-2.
- [71] Franz Aurenhammer. Power diagrams: properties, algorithms and applications. *SIAM Journal on Computing*, 16(1):78–96, 1987. doi: 10.1137/0216006.
- [72] A. Bondi. van der Waals volumes and radii. *The Journal of Physical Chemistry*, 68(3):441–451, 1964. doi: 10.1021/j100785a001.
- [73] Vyas Ramasubramani, Bradley D. Dice, Eric S. Harper, Matthew P. Spellings, Joshua A. Anderson, and Sharon C. Glotzer. freud: a software suite for high throughput analysis of particle simulation data. *Computer Physics Communications*, 254:107275, 2020. doi: 10.1016/j.cpc.2020.107275.
- [74] R. Scott Rowland and Robin Taylor. Intermolecular nonbonded contact distances in organic crystal structures: comparison with distances expected from van der Waals radii. *The Journal of Physical Chemistry*, 100(18):7384–7391, 1996. doi: 10.1021/jp953141+.
- [75] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul

- van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3): 261–272, 2020. doi: 10.1038/s41592-019-0686-2.
- [76] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020. doi: 10.1038/s41586-020-2649-2.
- [77] Bradley Efron. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979. doi: 10.1214/aos/1176344552.
- [78] Edwin B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158): 209–212, 1927. doi: 10.1080/01621459.1927.10502953.