Computational and Structural Biotechnology Journal 20 (2022) 2798–2806

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Unravelling hierarchical levels of structure in lipid membranes

Alexandre Blanco-González [a,b,c], Ángel Piñeiro [a,*], Rebeca García-Fandiño [b,*]

[a] Departamento de Física Aplicada, Facultade de Física, Universidade de Santiago de Compostela, E-15782 Santiago de Compostela, Spain
[b] Departamento de Química Orgánica, Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares (CiQUS), Universidade de Santiago de Compostela, Campus Vida s/n, E-15782 Santiago de Compostela, Spain
[c] MD.USE Innovations SL, Edificio Emprendia, 15782 Santiago de Compostela, Spain

ARTICLE INFO

Article history:
Received 13 March 2022
Received in revised form 21 May 2022
Accepted 21 May 2022
Available online 25 May 2022

Keywords:
Lipid bilayer
Cell membrane
Hierarchical levels of structure
Molecular dynamics simulations

ABSTRACT

In analogy with the hierarchical levels typically used to describe the structure of nucleic acids or proteins and keeping in mind that lipid bilayers are not just mere envelopers for biological material but directly responsible for many important functions of life, it is discussed here how membrane models can also be interpreted in terms of different hierarchies in their structure. Namely, lipid composition, interaction between leaflets, existence and interaction of domains arising from the coordinate behavior of lipids and their properties, plus the manifest and specific perturbation of the lipid organization around macromolecules embedded in a membrane are hereby used to define the primary, secondary, tertiary and quaternary structures, respectively. Molecular Dynamics simulations are used to illustrate this proposal. Alternative levels of organization and methods to define domains can be proposed but the final aim is to highlight the paradigm arising from this description which is expected to have significant consequences on deciphering the underlying factors governing membranes and their interactions with other molecules.

1. Introduction

What Rosalind Franklin, James Watson and Francis Crick experienced when they first got a glimpse of the DNA 3D structure [1,2], or when John Kendrew captured what was coming out of myoglobin RX into his sausage model [3,4], was surely comparable to what a 20 diopters-myopic person feels when putting on glasses for the first time. Those events changed the logic of biology because the way in which we currently view DNA or proteins affects how we think about their function. This has a significant impact on hypotheses development, experimental design, and data interpretation.

The three-dimensional structure of large biological molecules such as proteins and nucleic acids is complex and confusing to the naked eye. It is not easy to describe them accurately just using mundane words. Because of this, it is convenient to introduce new concepts. A series of hierarchical levels -typically referred as primary, secondary, tertiary, and quaternary structures- are commonly employed to describe biological macromolecules. These different "organization levels" exhibit regularity and predictability for natural phenomena represented in diverse ways, thus facilitating scientific inquiry and hierarchical causal explanations of different biological processes.

In proteins and nucleic acids, *the primary structure* is associated to the covalent chemical structure. It consists of linear arrays of covalently bound building blocks with a common scaffold and a variety of chemical substitutions. The number of different building blocks is restricted to five in typical nucleic acids (including RNA), and to twenty in proteins. Thus, the primary structure can be specified by the linear order, or sequence, of these building blocks. The *secondary structure* refers to regular patterns of interaction between adjacent or facing residues. The most illustrative examples are alpha-helices or beta-sheets observed in many proteins and double helices found in virtually all nucleic acids. The *tertiary structure* refers to the location of the atoms in three-dimensional space, taking into consideration geometrical and steric constraints. It usually provides the starting point for studies that attempt to correlate structure and function. In proteins it comes from the

Abbreviations: P.O.P.C., 1-Palmytoyl-2-Oleoyl-sn-glycero-3-Phosphocholine; P.O.P.E., 1-Palmytoyl-2-Oleoyl-sn-glycero-3-PhosphoEthanolamine; P.O.P.G., 1-Palmytoyl-2-Oleoyl-sn-glycero-3-Phospho-Glycerol; D.M.P.C., 1,2-DiMyristoyl-sn-glycero-3-PhosphoCholine; D.O.P.C., 1,2-DiOleoyl-sn-glycer0-e-PhosphoCholine; L.R.S., Local Reference System; S.V.D., Singular Value Decomposition.

* Corresponding authors.
E-mail addresses: Angel.Pineiro@usc.es (Á. Piñeiro), rebeca.garcia.fandino@usc.es (R. García-Fandiño).

https://doi.org/10.1016/j.csbj.2022.05.042
2001-0370/

folding of the secondary structure into distinct arrangements known as domains, whereas the tertiary arrangement of DNA's double helix in space includes B-DNA, A-DNA, and Z-DNA. *Quaternary structure* describes the assembly of individual molecular units into more complex arrays. The simplest example of quaternary structure is a protein that consists of multiple subunits, which are not connected to each other by covalent bonds and may be identical or different. In nucleic acids, it refers to interactions with other molecules, for example in the form of chromatin which leads to its interactions with the histone small proteins.

This hierarchical representation in four structural layers of organization is extremely robust and well established empirically for proteins and nucleic acids but it quickly loses its significance when applied throughout the whole spectrum of biological entities. Different macromolecules and supramolecular structures, such as polysaccharides and lipid membranes, exhibit clear three-dimensional shapes and are built out of smaller subunits. However, different levels of structure have not been defined in these cases. Polysaccharides are quite similar to proteins in that they are chains of repeated building blocks (monosaccharides) that acquire well-defined three-dimensional topologies. The case of cell membranes requires a more detailed explanation. According to the fluid mosaic model proposed by Singer and Nicolson in 1972[5], the cellular plasma membrane is a two-dimensional liquid, in which lipid and protein molecules are mixed like a mosaic and undergo thermal diffusion. This model is still widely accepted by the research community. Under this approach all the molecules are expected to be homogeneously distributed throughout the plasma membrane, undergoing simple Brownian diffusion. However, an increasing number of evidence showing that this is not the case are being found[6]. Like proteins, nucleic acids and even sugars, lipid membranes can be also described using a hierarchical organization in different levels of structure. The point of view arising from such description is expected to help to understand the underlying factors governing membranes and their interactions with other molecules. This is especially important since lipid membranes do constitute one of the most important parts of any living cell, acting as the main barrier between the cytoplasm and its surrounding environment[7]. Their function is not limited to a mere division between the inside and the outside of cells, but it is extended to a wide range of processes, such as serving as anchor points and modulators for signaling[8,9] and transport[9,10] proteins, or playing a key role in cell motility[9,11]. Lipid membranes are also known to be subject to substantial attacks and changes upon disease related processes. For instance, some families of pathogens such as enveloped viruses -HIV and SARS-CoV-2, among many others- are known to enter a cell by fusing their membranes with the host's[12,13]. Once internalized and just before the viral reproduction cycle is completed, they leave the cell by selectively ripping apart[14] the infected cell's membrane, making it their own and leaving the host's cell heavily modified in its lipidic composition. This modification is hypothesized to sometimes be able to trigger an autoimmune response in the diseased host leading to chronic inflammation[15], which might have fatal consequences. On the other hand, cancerous processes are also known to display significant alterations on the membranes of the affected tissues [16,17]. While healthy mammalian somatic cells present an asymmetry in their composition depending on the leaflet, leaving most of the negatively charged lipids towards the cytoplasm; cancer cells tend to lose this property providing them with negative charges on their exterior, a characteristic that share with bacterial cells[18,19]. All of this illustrates the important role that lipid membranes play as targets both to invasive pathogens and to potential therapies for diseases derived from such processes. It is also reasonable to think that they play a role in some chronic autoimmune responses[20,21]. Therefore, a proper understanding of the underlying factors governing membranes and their interactions with different species is mandatory to obtain rational criteria from which then derive effective and efficient solutions to the problems that they are involved in. Characterizing living membranes is not easy, as they are densely packed fluids with literally thousands of different components, and their compositions vary greatly depending on the type or source of the cell. Even with much simpler models, empirical study can be cumbersome and there are great difficulties in reaching the resolution needed to get to the atomic or quasi-atomic levels at which some important processes happen. Therefore, computational Molecular Dynamics (MD) simulations are an excellent choice to develop and analyze models of such systems[22].

In the present work, a set of hierarchical levels of structure are proposed for lipid membranes, in line with those widely recognized for other biomolecular structures such as proteins and DNA. Original MD simulations results for large POPC, POPE-POPG (1:1 ratio) and DMPC lipid bilayers at coarse grained (CG) resolution; and of several macromolecules embedded in an atomistic DOPC membrane from a previous work[23] are employed to illustrate the proposal. This perspective for the description of lipid membranes is inspired by what Franklin, Watson and Crick did with DNA and Kendrew did with proteins. Alternative levels of structure could of course be proposed to describe biological membranes. The aim of this work is to discuss an approach that is expected to contribute to understand better the behavior of membrane models as well as of actual biological envelopes based on lipids.

2. Methods

2.1. Simulation parameters

Three flat-square membranes (POPC, POPE-POPG and DMPC) consisting of 5000 Martini 2.2 lipids[24] per leaflet were built using the CHARMM-GUI Martini Maker[25–27]. The resulting bilayers, spanning about 57 nm in the X and Y dimensions, were solvated using a ratio of 23 non-polarizable Martini water particles per single lipid molecule, thus leading to a 57x57x8 nm^3 box. Monovalent ions were added to the systems in just the needed amount to make the total charge neutral. The coarse grained resolution was employed due to the large size of the simulated systems, that cannot be afforded at atomic resolution. Periodic boundary conditions were applied to the three spatial dimensions and the energy of the systems was minimized for 5000 steps using the GROMACS 2020.4 molecular dynamics engine[28,29] with a steepest descent algorithm, a tolerance of 10 $kJ{\cdot}mol^{-1}{\cdot}nm^{-1}$ and a step size of 0.01 nm. The simulation boxes were then equilibrated for 100 ns of molecular dynamics simulation in five consecutive stages, where the time step was sequentially increased from 2 fs to 25 fs using the leap-frog integrator. During this process the temperatures were maintained at 310 K with a velocity rescale algorithm[30] and a coupling constant of 1 ps, while the pressures were maintained at 1 atm with a Berendsen semi-isotropic barostat,[31] a coupling constant of 5 ps and a compressibility modulus of $3{\cdot}10^{-4}$ bar^{-1}. The electrostatic interactions were calculated with the reaction-field method with a cutoff radius of 1.1 nm and a dielectric constant of 15 beyond the cutoff. The van der Waals interactions were calculated with a cutoff of 1.1 nm.

After equilibration, unrestrained 10-μs-long MD production trajectories using a leap-frog integrator with a timestep of 25 fs were obtained. Removal of center of mass motion was performed each 1000 steps, and position coordinates were written with the same period (25 ps). Neighbor lists were obtained with the Verlet [32,33] method and updated every 20 steps. Temperatures were

maintained again at 310 K with a velocity-rescale algorithm and a coupling constant of 1 ps; while pressures were kept constant at 1 atm with a semi-isotropic Parrinello-Rahman barostat,[34] a coupling parameter of 12 ps and a compressibility modulus of $3{\cdot}10^{-4}$ bar^{-1}. Electrostatic interactions were calculated with the reaction-field scheme, a cutoff radius of 1.1 nm and a dielectric constant of 15 beyond the given radius. van der Waals interactions were calculated with a cutoff of 1.1 nm.

2.2. Analysis

All the analyses were carried with homemade scripts written in Python. The recurrently used libraries were MDAnalysis[35,36] to directly read trajectory files in GROMACS format, Numba[37] to facilitate high-performance parallelization schemes, Statsmodels [38] to calculate autocorrelation functions, NumPy[39] for general data handling and mathematical operations, and Matplotlib[40] for data representation.

2.2.1. Determination of local properties

The analysis performed in this work was done within the Monge gauge framework, as described by M. Deserno[41]. For that, we relied on the discretization of the lipid membrane in a square grid spanning the whole XY plane, with an average of $\boldsymbol{N_L}$ lipid heads (PO4 groups) per cell. The coordinates of the grid cells were the same for the two leaflets and were kept constant over the entire trajectory. A battery of different properties was then determined for each cell and independently analyzed for each leaflet.

In order to describe the dynamic behavior of the membrane, as well as local domains and the connections between them, a local reference system (LRS) per grid cell as a function of time was determined. The orthonormal vectors describing these LRS were defined as: (*i*) the normal to each cell, $\boldsymbol{Z_i'(t)}$; (*ii*) the normalized cross product between $\boldsymbol{Z_i'(t)}$and the gradient of the curvature., $\boldsymbol{X_i'(t)}$ -*i.e.* the direction of Isocurvature; and (*iii*) the cross product between the previous two vectors, $\boldsymbol{Y_i'(t)}$, -the direction of Anisocurvature. The protocol used to define these vectors is described in detail in the Supplementary Material (SM). By using these LRS it is possible to determine structural properties such as membrane thickness or tail orientation while accounting for local deformations of the lipid bilayer. Such deformations might be ignored in small membranes, but they are significant in patches of several tens of nm long per edge, as it will be shown later.

The set of properties analyzed in the present paper can be computed with different size resolution. This is given by the number of lipids per grid cell ($\boldsymbol{N_L}$). Different values of this parameter were tested in our code and after several tests we decided to present the results obtained with $\boldsymbol{N_L}$ = 10. which allows observing significant changes in the quantitative properties computed for the lipid bilayer, with good spatial resolution and optimizing the computational resources employed.

Using these parameters, the topography of the membrane was determined as the average Z coordinate of the PO4 groups per grid cell. The local membrane curvature was calculated as described in the Supporting Materials. The thickness of the membrane was determined using the LRS. The thickness is given by the average difference between the position of PO4 groups along the local $\boldsymbol{Z_i'(t)}$ axes. The orientation of the tails was represented by a Lipid Director vector, as described by M. Ergüder and M. Deserno[42] and explained in the Supporting Materials. The Director tilt was calculated with respect to the L.R.S.

2.2.2. Global analysis of the lipid bilayer

Once the previously described properties are computed for each grid cell, time dependent local domains could be directly observed by plotting them in suitable representations (see results section and SM). A Singular Value Decomposition (S.V.D.) analysis is then performed to identify a set of independent properties associated to the membrane structure as well as to quantify the coupling between the different studied properties (see SM for a detailed description). Each of the three components of the previously defined vectorial magnitudes were independently treated as scalar properties for this analysis.

We ensured that the individual trajectory frames used for the analyses were statistically independent by calculating the autocorrelation functions (A.C.Fs.) for the variance of the curvature over the surface of the membranes for each frame, along the whole trajectories. Values close to 0 in the A.C.Fs. indicate that the property for which the function was calculated gets decoupled from itself after the corresponding value of time, thus indicating statistical independence. For all the simulations a sensible choice was $\tau = 20ns$ (Figures S10, S11 and S12), as the A.C.Fs. have almost completely decayed by that time. Hence, just frames separated by 20 ns were selected for the S.V.D. analysis.

3. Results & discussion

The different hierarchical levels of structure proposed to reinterpret membrane models are described here in detail. As explained in the introduction, MD simulations of large POPC, POPE-POPG and DMPC membranes at CG resolution, as well as atomistic simulations of different macromolecules embedded in a DOPC membrane taken from a previous work are used to illustrate the new concepts. The large membranes were chosen to show the impact of the system size in some of the analyzed properties, to prove that even a membrane with a single component exhibits the proposed levels of structure and to test the effect lipid headgroups, tail unsaturation and tail length have in such structural descriptions. In the main text just the results for the large POPC bilayer will be shown, for the sake of brevity. However, the results for the other two cases can be seen in the Supplementary Materials and are qualitatively equivalent to those discussed in the Main Text. The atomistic simulations of more complex systems were employed to show the highest level of structure, as explained later.

3.1. The ***primary structure*** *of the lipid bilayer: A "sequence" of thousands of non-covalent basic units*

In our analogy with the hierarchical representation of four structural layers of organization in proteins and nucleic acids, the **primary structure** (Fig. 1A) of lipid membranes is identified with the composition of each leaflet, that is different in asymmetric bilayers. Variations in headgroups and aliphatic chains allows the existence of more than 1,000 different lipid species just in a single eukaryotic cell[43]. As explained in the introduction section, a number of infections and diseases have a serious impact on the lipid composition of the host cells. This lipid composition can be employed as a target for the actuation of the immune system and for artificial therapeutic treatments. Actually, all living organisms have a series of short, cationic and amphipathic peptides known as Host Defense Peptides (HDPs) which are able to recognize and disrupt pathological and pathogenic membranes based on their composition[44,45]. These HDPs are an important part of the innate immune system. This illustrates the huge importance of the lipid composition of biological membranes, and it justifies that it is identified with the primary structure of the lipid bilayer. However, there are clear discrepancies in this analogy with respect to the primary structure of nucleic acids and proteins. The most obvious difference is that the lipids in a membrane are not covalently bound, in contrast to the nucleotides in DNA or RNA and

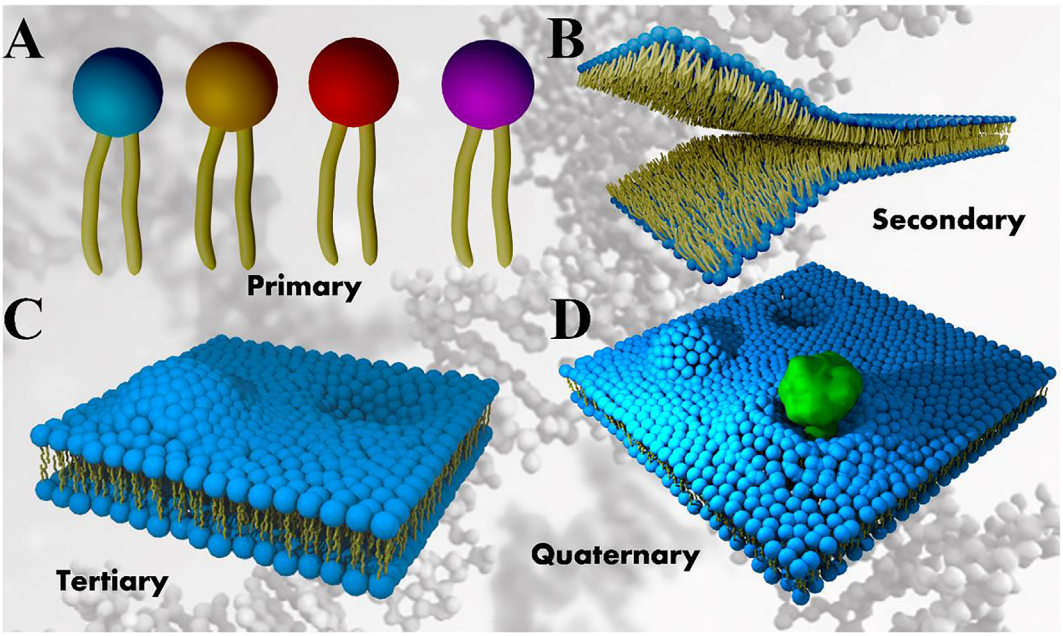

Fig. 1. Schematic representation of primary (A), secondary (B), tertiary (C) and quaternary (D) levels of structure of a lipid bilayer. Each color in A represents a different lipid type. See text for details.

to amino acids in proteins. Another evident difference is that lipid membranes are locally flat, even though they can form closed quasi-spherical compartments with significant defects or protuberances. Because of the previous two features, membranes have lipids diffusing throughout them, incorporate other molecules into their structure, fuse with other membranes and exhibit a high ability to adapt to environmental changes by altering their composition. Lipid membrane composition alterations take place as a response to almost any environmental change: pH, temperature, diet, etc.[46] This would be equivalent to a protein or a nucleic acid altering its sequence as a response to an external perturbation, which is not possible because their residues are joined by covalent bonds. The high plasticity and flexibility of biological membranes allow them to maintain their bioactivity despite the effect of external perturbations. Thus, within our proposal, the primary sequence of membrane is simply the proportions of different lipids in each leaflet.

*3.2. The **secondary structure** of the lipid bilayer: A Velcro strap more resistant than helices or beta sheets*

The secondary structure of DNA arises from the interaction between two polynucleotide chains forming a double helix while in proteins it is defined by local and specific interactions between amino acids leading to alpha-helices or beta-sheets, among other patterns. The case of lipid bilayers is similar to DNA in the sense that they consist of two monolayers facing each other as a Velcro strap (Fig. 1B), whereas DNA has two complementary chains interacting as a zipper. This analogy is not only structural but also with respect to the interactions. The local interactions (by H-bonds) between two nucleotides in a DNA fragment is significantly stronger than the hydrophobic (mainly van der Waals type) interaction between two lipids of different leaflets in a membrane patch, however these latter interactions are highly cooperative - exactly as in a Velcro strap - and altogether makes the interaction between monolayers extremely strong.

This secondary structure is critical for the stability of lipid bilayers as lipid monolayers do not exist in solution (although they are stable in a hydrophobic/hydrophilic interphase). Again, there are differences between the two systems for this level of structure, *e.g.,* the interaction between leaflets taking place in two dimensions instead of the one-dimensional interaction between polynucleotide chains, and the fluidity of lipids in a membrane compared to the fixed position of nucleotides in each DNA strand. As a result of this level of structure, the two leaflets in a lipid bilayer are highly correlated to each other. This is clearly illustrated in the topography of both monolayers as well as in the field lines arising from the Normals, Isocurvature and Anisocurvature (see Methods) vector fields (Fig. 2, Fig. S6, Fig S9).

*3.3. Determination of local quantitative descriptors for the lipid bilayer. The **tertiary structure** of the lipid bilayer.*

Higher levels of structure can be inferred from the coordinate behaviour of lipids, leading to local domains that interact with each other (Fig. 1C). It is worth to note that these domains do not correspond to different thermodynamic phases in equilibrium with each other but they are spatial regions defined by strong correlations in a given property of the lipids: topography, thickness, tail orientation, etc. Therefore, the definition of the specific domains depends on the analyzed property. For instance, the topography of a membrane patch may exhibit protuberances and valleys (Fig. 2, S6 and S9), the same structure could also have regions of different thickness (Fig. S23, S25 and S.27), and the lipid tails could be orientated in a coordinate way forming patterns (Fig. S13, S14 and S15). Each of these properties would lead to different criteria to define domains. The analysis proposed in the methodology section identifies different kind of lipid domains as well as of interactions or couplings between them. The results corresponding to the large bilayers simulated using coarse grained resolution are presented and discussed in what follows.

3.3.1. Identification of domains

As shown above (Fig. 2), the diameter and amplitude of the membrane protuberances conditions the vector fields arising from the LRS. Both the topography and these vector fields form well-defined domains. The distribution of the curvature exhibits a maximum at 0 $Å^{-1}$, as expected as the membranes are overall flat and

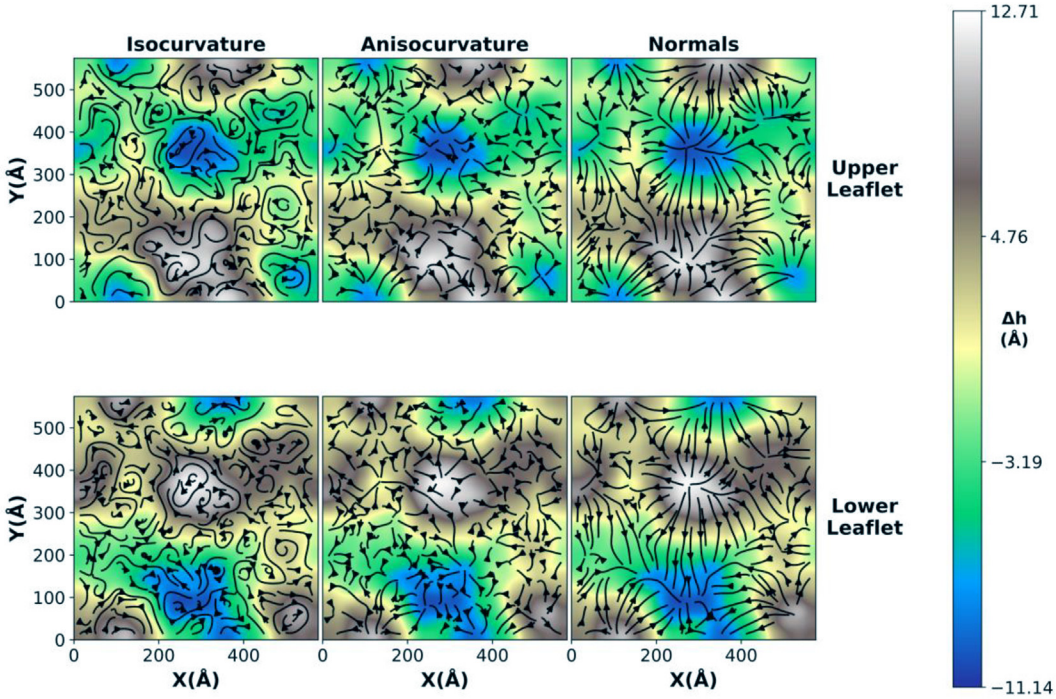

Fig. 2. Topographic heatmap and LRS vector fields for both leaflets (see labels on plots) of the POPC membrane. Image obtained from the average over the last ns of the trajectory. The topography maps exhibit clear hills and wells whose projection in the XY plane spans a diameter of ~ 15 nm with an amplitude of ~ 2 nm in Z. The negative correlation between the two leaflets is evident. The field lines obtained from the Isocurvature, Anisocurvature and Normals vector fields are also clearly correlated with the topography of the membrane. The correlation/connection betweenn different regions and the anti-complementarity between the two leaflets are also obvious in these vector field maps. The Normals show field lines falling down the top of the hills and rising up the bottom of the wells, as expected. The Isocurvature exhibits strong curls and long, continuous fields lines contouring regions of same elevation. The Anisocurvature displays shorter and straighter field lines, with no curls at all, and with vector line sources in regions where the Isocurvature presents dextro-rotating curls and wells in the regions where the Isocurvature presents levo-rotating curls.

are indistinguishable for both leaflets (Fig. S2, S5 and S8). The associated heatmaps (Fig. S1, S4 and S7) show that the more extreme values for the curvature correspond to parts where the topography of the membrane diverts further from the ideal planarity.

The vector fields arising from the Lipid Director were determined as described in section 2.2.1 and in the Supplementary Materials. (Fig. S2). The field lines emanate from the lowest points and converge in the highest regions (Fig. S13, S14 and S15) a pattern that was already expected since the curvature itself conditions the orientation of the lipid tails in the observed directions. The vector field of the Lipid Director is antiparallel to that obtained from the Normals (***Z'***). It is also useful to measure the angle between the lipid tail vectors and the *Z'* axis of the LRS (Fig. S16, S18 and S20). The corresponding heatmaps (Fig. S17, S19 and S21) are less clear than, for instance, the curvature, likely because the lipid bilayer is in fluid phase; but slighter lower angles (*i.e.*, with lower projection over the *Z'* axis of the LRS) can be observed for the regions where the topographic slope is maximum.

The membrane thickness calculated considering the LRS exhibits a narrow normal-like distribution centered at different values depending on the specific membrane, as expected. (Fig. S22, S24 and S26). The associated heatmaps show domains of different thickness, in a more or less obvious way depending on the case (Fig. S23, S25 and S27). The correlation between the thickness and the topography is not evident, although the order of the lipid tails is expected to simultaneously depend on both properties and so they are expected to have some connection.

3.3.2. Filtering non-redundant properties

In general, it is clear that all the analyzed properties arise from the interaction between groups of lipids and that they define structural regions - or domains - in the lipid bilayer that span several cells of the discrete grid defined for the analysis. These domains, as well as the interaction between them, will be identified in the present work with the **tertiary structure of the lipid bilayer**, in analogy with that in nucleic acids and proteins arising from the relative spatial location of nucleic acids and amino acids.

However, as stated above, a single definition of the domains is not clear because those defined by the topography may be different from those defined by the thickness or by any other property. It is convenient to identify which properties are coupled to each other, thus providing identical or cooperating domains, and which properties are independent from the rest thus providing non redundant information. Next, a minimal set of quantitative descriptors obtained from the properties calculated in the previous section will be identified. This set of variables should be able to describe the structure of the lipid bilayer without redundancy. An SVD factorization of the data matrix containing the variables' z-score normalized values, over all the grid cells and for all the selected independent frames of our trajectory, will be employed for this aim (see methods section for details). As both leaflets behave equivalently, the SVD analysis was separately carried out for each leaflet, giving equivalent results (Figs S28 - S29, S31 - S32 and S34 – S35).

The explained variance per eigenvector is a step function with several groups of properties contributing equally to the total variance (Figs S28 - S36, panel A). This is a consequence of the close connection between variables – such as the topography of the membrane and the curvature, or the Normals, Isocurvature and Anisocurvature vector fields - and their cooperative behavior forming the membrane domains, which in these cases would not be defined by just one property but by the combination of several ones. Complementarily, just 11 variables are required to explain the 95% of the total variance for all the studied membranes

(Figs S28 - S29, S31 - S32 and S34 – S35; panel A). Furthermore, the last 3–4 eigenvectors are almost completely redundant, as they do not add anything to the Cumulative Variance. The V^T matrices containing the eigenvector components in the space generated by the properties used for the analysis, indicates too that several variables are coupled, since they can be explained by almost identical eigenvectors.

This inter-variable coupling was quantified through the dot product between the column vectors consisting of the normalized relevance matrix (Figs S28 - S29, S31 - S32 and S34 – S35; panel C). The results of these dot products were then represented as a coupling matrix (Figs S28 - S29, S31 - S32 and S34 – S35; panel D). In these matrices, it is clearly seen that the X and Y components of the Isocurvature, Anisocurvature and Normals vector fields are fully coupled to each other. Moreover, the X and Y components of the surface normal are strongly coupled to the X and Y components of the Lipid Director. These correlations were already expected after inspecting and comparing Fig. S3, S6 and S9 with Fig. S13, S14 and S15 - as the antiparallel disposition of the Director and Normals vector fields is rather evident to the eye - but finding them as a result can be considered as a validation of the method.

The previous information suggests that the minimum number of properties used to efficiently characterize the lipid membrane can be reduced. Such set of properties could be considered as the origin for an orthogonal basis of quantitative descriptors for the lipid bilayer, this is, their eigenvectors. It is interesting to observe the absolute lack of correlation between different properties, indicating that they are independent and complementary to each other. For instance, the topography of the membrane does not display any correlation with the X or Y component of either the Iso or Anisocurvature but does faintly show it with the Z component of the latter vector field, and quite more clearly with the membranes' curvature. It is clearly seen in Fig. 2 that the X and Y components of the normal vector are negligible in the maxima and minima of the membrane protuberances, thus a clear correlation between the Z component of said vector and the relative Z position of the lipid heads was to be expected. The reason for the observed lack of correlation is that it is only present at such maxima and minima but, not in general, such as at intermediate positions which weigh much more in the ensemble. For instance, the normal vectors at flat regions are also parallel to the Z axis of the global reference system, although the relative Z position of the corresponding head groups is null in these cases.

The redundant properties of the previous analyses were detected based on their shape similarities in the eigenvector space. This is, starting from the total SVD, pairs of properties that share an identical or quasi-identical - as an example, $\mathbf{X'_X}$ and $\mathbf{Y'_Y}$ in Fig. S28 – profile along a column in the Relevance matrix, are deemed equivalent, and one of the two is then filtered out. After removing such variables, the SVD analysis was repeated with the following descriptors: the Z component of the Isocurvature, the X and Y component of the Anisocurvature, the complete Normals vector field, the Curvature scalar field, the membrane's Thickness, the Lipid Director Tilt, and the full Lipid Director vector field. It is remarkable that by following this filtering-out procedure, we take out four variables (Height Function, X and Y Isocurvature components and

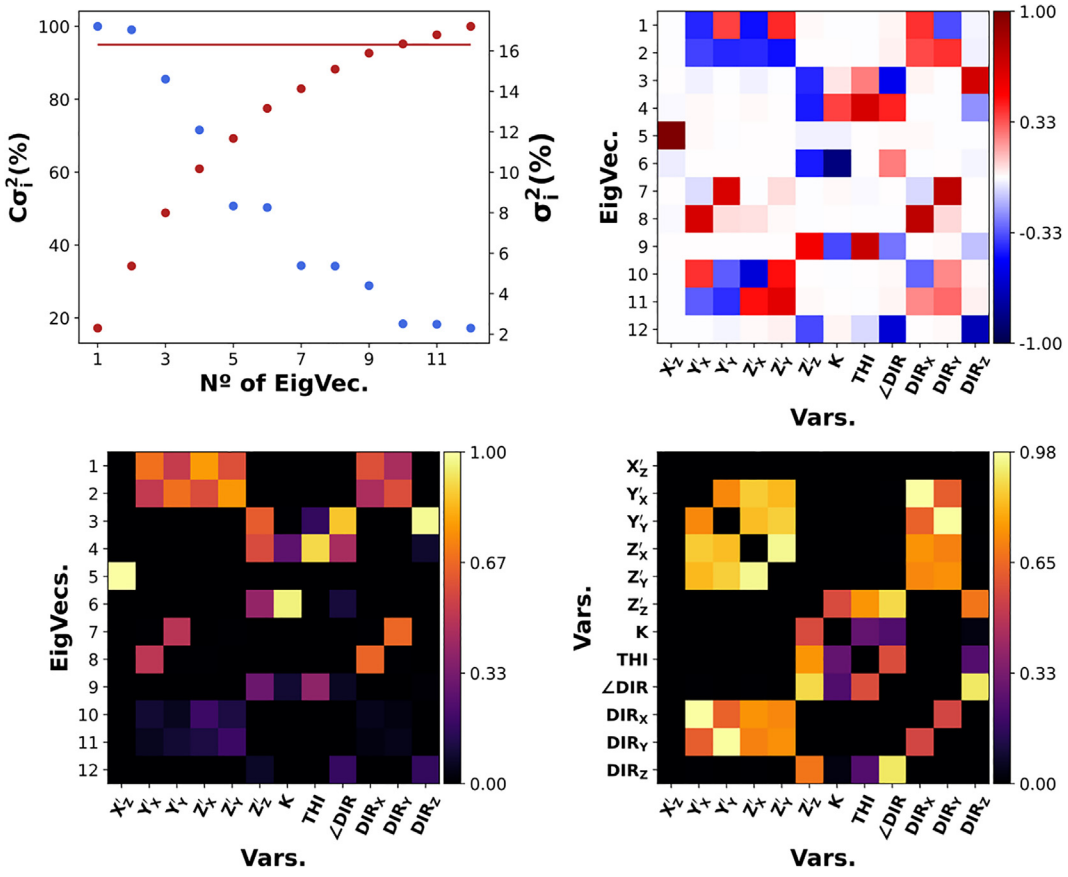

Fig. 3. Results for the SVD analysis performed on the nonredundant set of variables. Cumulative (red) and explained (blue) variance, 95% explained marked with a red line (A); V^T matrix (B); relevance matrix (C); coupling matrix (D). All the L.R.S. components are coupled to each other, thus one eigenvector is associated to the Z component of the Isocurvature. The L.R.S. components are simultaneously coupled to the X and Y components of the Lipid Tail Director, indicating that the orientation of the lipid tails is governed by the directions of curvature of the membrane. The Z component of the Normals vector field, the Curvature scalar field, the thickness and the angle of the lipid director are coupled between them, pointing to a close interplay between these magnitudes. The latter properties are also coupled to the Z component of the Lipid Tail Director vector field. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Z Anisocurvature component), in good agreement with the fact that the full SVD analysis produces 4 redundant eigenvectors. Note that by not explicitly considering these variables we are not actually eliminating them, but assuming that they can be well represented by some other property. It is also worth noting that for all the studied membranes the candidate to filtering out properties were always the same, indicating that the underlying mechanisms governing the interaction between membrane domains – the proposed Tertiary structure - are independent of membrane composition at this level.

No redundant eigenvectors were observed in the cumulative variance plot (Fig. 3A, S30, S33 and S36), as expected when all the properties considered for the analysis are independent from each other. The final V^T, normalized relevance and coupling matrices (Fig. 3B-D) shows that there are strong couplings between different variables. Some similarities are seen again in the Relevance matrices, but because of the previously mentioned lack of redundant eigenvectors, no more variables can be removed this time.

From the initial set of 31 properties, considering both leaflets, 12 quantitative descriptors resulted to be not superfluous. In analogy with the tertiary structure of proteins, the different domains interacting with each other -described by the set of non-redundant properties as shown in Fig. 3D- and their couplings can be associated to a tertiary structure in lipid bilayers. In the present work a simple homogenous POPC, a binary POPE-POPG and a homogeneous DMPC bilayer are analyzed. The correlation between the different analyzed properties are comparable for all the simulations. This means that the observed relationships between variables are an underlying mechanism of how membranes organize themselves in equilibrium, beyond their composition – or Primary Structure. This will not necessarily happen if different properties are analyzed. Going further, the projection of the different eigenvector over the MD trajectory (Fig. S37, S38 and S39) could provide an alternative group of complementary domains defined by the collective behavior of groups of lipids considering specific linear combinations of the quantitative descriptors employed in our analysis.

3.4. The quaternary structure of the lipid bilayer

The levels of structure described above are expected to be present in all membranes, regardless of their composition, however the highest-order hierarchy remains to be established. In nucleotides, the quaternary structure is related to the interaction with different biomolecules which fold the double stranded DNA in different patterns. Following the same philosophy, it is sensible to test the response of membranes to the presence of macromolecules in their environment (Fig. 1D).

Taking advantage of previous work[23] we test and compare the behaviour of an atomistic DOPC membrane model when in the presence of a carbon nanotube (CNT) and a simple β-barrel model (Fig. 4 and Fig. S40). Just by visual inspection, the influence

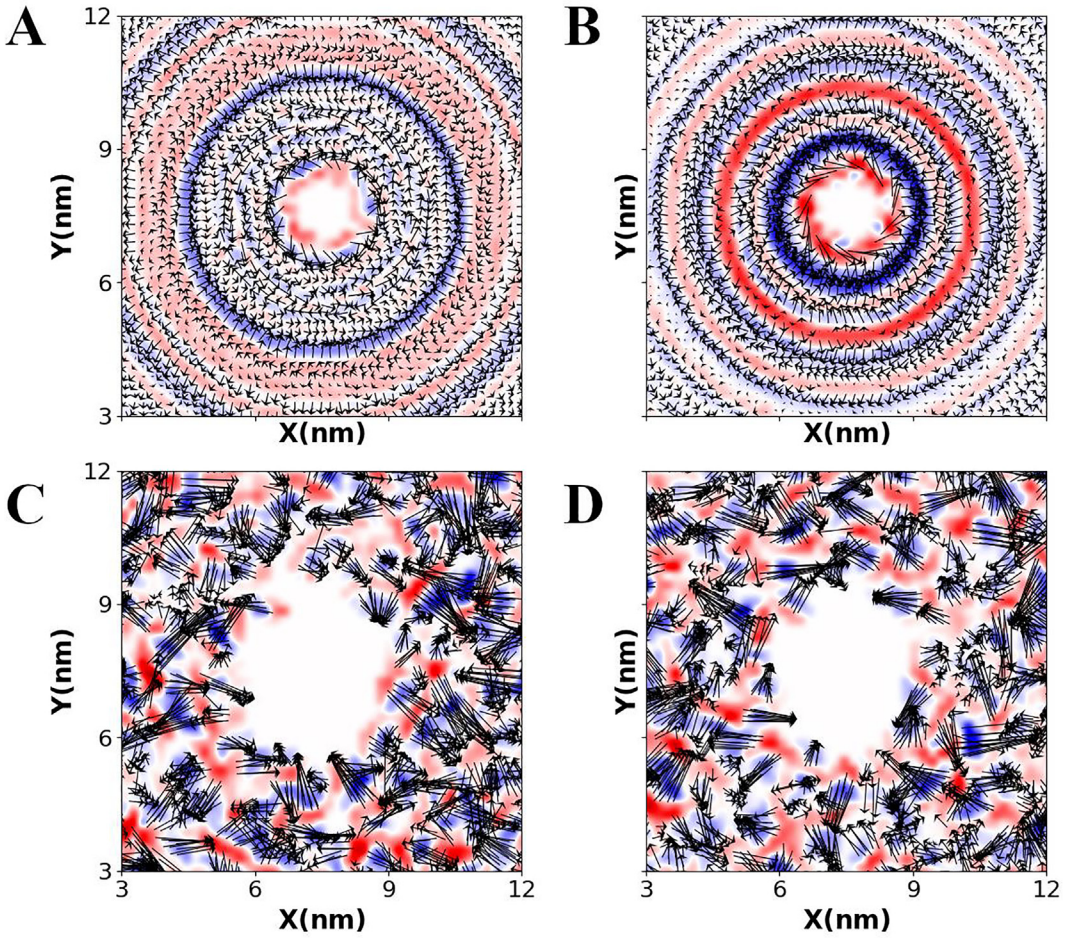

Fig. 4. Schematic representation of the quaternary structure in a DOPC lipid bilayer containing a carbon nanotube (A,B) and a β-barrel protein model (C,D). The vector fields obtained from the orientation of the lipid SN1 tail (A,C) and SN2 tail (B,D) in the upper leaflet are plotted together with the corresponding divergence represented in red-blue color gradient for divergence-convergence, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the macromolecules on the lipid organization is remarkable and distinguishable. In both cases the lipid tails become highly ordered around the macromolecule and they distort the nominal kind of organization seen in previous sections. The CNT creates a series of concentrical circular ripples around itself and generates a continuous pattern in the lipid tails. The divergence heatmap also displays rings, indicating that there will be circular regions defined by coordinate orientation of lipids. On the other hand, the beta-barrel generates local clumps of lipids with the same orientation, but there is no overall cohesion, contrary to the former case. Similarly, the divergence heatmap also displays protein-induced lumps with no clear global organizational pattern.

Thus, the case of the quaternary structure is defined straightforwardly from that of proteins of nucleic acids: it is the response and interaction of lipid membranes to other macromolecules, which is distinctive depending on the case.

4. Conclusions

A hierarchical set of structural levels is proposed for lipid bilayers in analogy to the primary, secondary, tertiary and quaternary structural levels of nucleic acids and proteins. The primary structure is identified with the lipid composition of each leaflet, the secondary structure is defined by the interaction between the two leaflets conforming the membrane, the tertiary structure arises from the lipid domains obtained from the analysis of the properties calculated throughout the membrane surface, as well as from the interaction or coupling between such domains, and the quaternary structure comes from the impact of macromolecules embedded in the lipid bilayers. The coupling matrix proposed to identify the tertiary structure is expected to be a descriptor of how the mechanical domains interact with each other, or how the membrane curvature affects the bilayer thickness, among other connections. The impact of the presence of macromolecules embedded in the membrane, corresponding to the quaternary structure, is huge. The pattern of the lipid tails orientation is highly specific around rigid structures such as carbon nanotubes and less ordered around more flexible structures. Alternative approaches to the definition of hierarchical structural levels for membrane models or to the definition of membrane domains can be defined, our approach representing an illustrative example on this interesting perspective. Part of our treatment is based on a set of local quantitative descriptors, but different sets of properties could also be employed with the same objective. POPC, POPE-POPG, DMPC unaltered membranes, and DOPC membrane with different macromolecules embedded in it are employed to illustrate the proposed approach. The different compositions do not have a significant impact in the different levels of structure proposed and analized in the present work. This result will not necessarily be valid for alternative properties such as the number of contacts between lipid heads and water molecules -which is expected to strongly depend on the lipid head- or the local lateral displacement of lipids, amongst other properties that have not been analyzed in the present work. The connection of the domains obtained from our analysis with biological activity is still unexplored. The method and analytical methods proposed here have a lot of room for improvement, but we hope they will have an impact in the understanding of membrane models as well as in the design of compartmentalized structures for new biotechnological applications.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has received financial support from the Spanish Agencia Estatal de Investigación (AEI) and the European Regional Development Fund - ERDF (PID2019-111126RB-100, RTI2018-098795-A-I00, and PID2019-111327GB-I00) and by the Xunta de Galicia (ED431F 2020/05 and Centro singular de investigación de Galicia accreditation 2019-2022, ED431G 2019/03) and the European Union (ERDF). R.G.-F. thanks Ministerio de Ciencia, Innovación y Universidades for a "Ramón y Cajal" contract (RYC-2016-20335). P. F. G. thanks the Spanish Ministry of Economy and Competitiveness and the European Social Fund for his predoctoral research grant, reference BES-2016-076761. All calculations were carried out at the Centro de Supercomputación de Galicia (CESGA).

Contributions

Á.P., and R. G-F. conceived the study. A.B. performed the simulations and analysis. All authors discussed the results, contributed to write the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.05.042.

References

[1] Watson JD, Crick FHC. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature 1953;171:737–8. https://doi.org/10.1038/171737a0.
[2] Franklin RE, Gosling RG. Evidence for 2-chain Helix in crystalline structure of sodium deoxyribonucleate. Nature 1953;172:156–7. https://doi.org/10.1038/172156a0.
[3] Kendrew JC. The three-dimensional structure of a protein molecule. Sci Am 1961;205:96–110. https://doi.org/10.1038/scientificamerican1261-96.
[4] Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 1958;181:662–6. https://doi.org/10.1038/181662a0.
[5] Singer SJ, Nicolson GL. The fluid mosaic model of the structure of cell membranes. Science (80-) 1972;175:720–31. https://doi.org/10.1126/science.175.4023.720.
[6] Kusumi A, Suzuki KGN, Kasai RS, Ritchie K, Fujiwara TK. Hierarchical mesoscale domain organization of the plasma membrane. Trends Biochem Sci 2011;36:604–15. https://doi.org/10.1016/J.TIBS.2011.08.001.
[7] Alberts B, Johnson A, Lewis JJ, Morgan DL, Raff M, Roberts K, et al. Molecular Biology of the Cell. 6th ed. New York, New York, USA: Garland Science; 2014.
[8] Nussinov R, Tsai CJ, Jang H. Ras assemblies and signaling at the membrane. Curr Opin Struct Biol 2020;62:140–8. https://doi.org/10.1016/j.sbi.2020.01.009.
[9] Corradi V, Sejdiu BI, Mesa-Galloso H, Abdizadeh H, Noskov SY, Marrink SJ, et al. Emerging Diversity in Lipid-Protein Interactions. Chem Rev 2019;119:5775–848. https://doi.org/10.1021/acs.chemrev.8b00451.
[10] Stieger B, Steiger J, Locher KP. Membrane lipids and transporter function. Biochim Biophys Acta - Mol Basis Dis 2021;1867:166079. https://doi.org/10.1016/j.bbadis.2021.166079.
[11] Sens P, Plastino J. Membrane tension and cytoskeleton organization in cell motility. J Phys Condens Matter 2015;27:. https://doi.org/10.1088/0953-8984/27/27/273103273103.
[12] Kielian M, Jungerwirth S. Mechanisms of Enveloped virus entry into cells. Mol Biol Med 1990;7:17–31.
[13] Dimitrov DS. Virus entry: Molecular mechanisms and biomedical applications. Nat Rev Microbiol 2004;2:109–22. https://doi.org/10.1038/nrmicro817.
[14] Münz C. The autophagic machinery in viral exocytosis. Front Microbiol 2017;8:269. https://doi.org/10.3389/fmicb.2017.00269.
[15] Garcia-Fandino R, Piñeiro Á. Delving Into the Origin of Destructive Inflammation in COVID-19: A Betrayal of Natural Host Defense Peptides? Front Immunol 2021;11:3532. https://doi.org/10.3389/fimmu.2020.610024.
[16] Szlasa W, Zendran I, Zalesińska A, Tarek M, Kulbacka J. Lipid composition of the cancer cell membrane. J Bioenerg Biomembr 2020;52:321–42. https://doi.org/10.1007/s10863-020-09846-4.
[17] Alves AC, Ribeiro D, Nunes C, Reis S. Biophysics in cancer: The relevance of drug-membrane interaction studies. Biochim Biophys Acta - Biomembr 2016;1858:2231–44. https://doi.org/10.1016/J.BBAMEM.2016.06.025.
[18] Glukhov E, Stark M, Burrows LL, Deber CM. Basis for selectivity of cationic antimicrobial peptides for bacterial versus mammalian membranes. J Biol Chem 2005;280:33960–7. https://doi.org/10.1074/jbc.M507042200.

[19] Sohlenkamp C, Geiger O. Bacterial membrane lipids: Diversity in structures and pathways. FEMS Microbiol Rev 2015;40:133–59. https://doi.org/10.1093/femsre/fuv008.
[20] Cas MD, Roda G, Li F, Secundo F. Functional lipids in autoimmune inflammatory diseases. Int J Mol Sci 2020;21:3074. https://doi.org/10.3390/ijms21093074.
[21] Leitinger N. The role of phospholipid oxidation products in inflammatory and autoimmune diseases: Evidence from animal models and in humans. Subcell Biochem 2008;49:325–50. https://doi.org/10.1007/978-1-4020-8830-8_12.
[22] Moradi S, Nowroozi A, Shahlaei M. Shedding light on the structural properties of lipid bilayers using molecular dynamics simulation: a review study. RSC Adv 2019;9:4644–58. https://doi.org/10.1039/C8RA08441F.
[23] Garcia-Fandiño R, Piñeiro Á, Trick JL, Sansom MSP. Lipid Bilayer Membrane Perturbation by Embedded Nanopores: A Simulation Study. ACS Nano 2016;10:3693–701. https://doi.org/10.1021/acsnano.6b00202.
[24] De Jong DH, Singh G, Bennett WFD, Arnarez C, Wassenaar TA, Schäfer LV, et al. Improved parameters for the martini coarse-grained protein force field. J Chem Theory Comput 2013;9:687–97. https://doi.org/10.1021/ct300646g.
[25] Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. J Comput Chem 2008;29:1859–65. https://doi.org/10.1002/jcc.20945.
[26] Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J Chem Theory Comput 2016;12:405–13. https://doi.org/10.1021/acs.jctc.5b00935.
[27] Hsu PC, Bruininks BMH, Jefferies D, Telles C, de Souza P, Lee J, et al. Charmm-gui martini maker for modeling and simulation of complex bacterial membranes with lipopolysaccharides. J Comput Chem 2017;38:2354–63. https://doi.org/10.1002/jcc.24895.
[28] Lindahl, Abraham, Hess, Spoel van der. GROMACS 2020.4 Manual 2020. https://doi.org/10.5281/ZENODO.4054996.
[29] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2015;1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001.
[30] Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys 2007;126:. https://doi.org/10.1063/1.2408420014101.
[31] Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–90. https://doi.org/10.1063/1.448118.
[32] Computer VL. "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys Rev 1967;159:98–103. https://doi.org/10.1103/PhysRev.159.98.
[33] Chialvo AA, Debenedetti PG. An automated Verlet neighbor list algorithm with a multiple time-step approach for the simulation of large systems. Comput Phys Commun 1992;70:467–77. https://doi.org/10.1016/0010-4655(92)90108-B.
[34] Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys 1981;52:7182–90. https://doi.org/10.1063/1.328693.
[35] Gowers R, Linke M, Barnoud J, Reddy T, Melo M, Seyler S, et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proc 15th Python Sci Conf 2016:98–105. https://doi.org/10.25080/majora-629e541a-00e.
[36] Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. J Comput Chem 2011;32:2319–27. https://doi.org/10.1002/jcc.21787.
[37] Lam SK, Pitrou A, Seibert S. Numba: A LLVM-based Python JIT Compiler. Proc. Second Work. LLVM Compil. Infrastruct. HPC - LLVM '15, New York, New York, USA: ACM Press; n.d.
[38] Seabold S, Statsmodels PJ. Econometric and Statistical Modeling with Python. Proc 9th Python Sci Conf 2010:92–6. https://doi.org/10.25080/majora-92bf1922-011.
[39] Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature 2020;585:357–62. https://doi.org/10.1038/s41586-020-2649-2.
[40] Hunter JD. Matplotlib: A 2D graphics environment. Comput Sci Eng 2007;9:90–5. https://doi.org/10.1109/MCSE.2007.55.
[41] Deserno M. Fluid lipid membranes: From differential geometry to curvature stresses. Chem Phys Lipids 2015;185:11–45. https://doi.org/10.1016/j.chemphyslip.2014.05.001.
[42] Ergüder MF, Deserno M. Identifying systematic errors in a power spectral analysis of simulated lipid membranes. J Chem Phys 2021;154:214103. https://doi.org/10.1063/5.0049448.
[43] Sud M, Fahy E, Cotter D, Brown A, Dennis EA, Glass CK, et al. LMSD: LIPID MAPS structure database. Nucleic Acids Res 2007;35:D527–32. https://doi.org/10.1093/NAR/GKL838.
[44] Magana M, Pushpanathan M, Santos AL, Leanse L, Fernandez M, Ioannidis A, et al. The value of antimicrobial peptides in the age of resistance. Lancet Infect Dis 2020;20:e216–30. https://doi.org/10.1016/S1473-3099(20)30327-3.
[45] Huan Y, Kong Q, Mou H, Yi H. Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields. Front Microbiol 2020;11:582779. https://doi.org/10.3389/fmicb.2020.582779.
[46] Ernst R, Ballweg S, Levental I. Cellular mechanisms of physicochemical membrane homeostasis. Curr Opin Cell Biol 2018;53:44–51. https://doi.org/10.1016/j.ceb.2018.04.013.