



OPEN SMARTINI3 parametrization of multi-scale membrane models via unsupervised learning methods

Alireza Soleimani^{1,2} & Herre Jelger Risselada^{1,2}✉

In this study, we utilize genetic algorithms to develop a realistic implicit solvent ultra-coarse-grained (ultra-CG) membrane model comprising only three interaction sites. The key philosophy of the ultra-CG membrane model SMARTINI3 is its compatibility with realistic membrane proteins, for example, modeled within the Martini coarse-grained (CG) model, as well as with the widely used GROMACS software for molecular simulations. Our objective is to parameterize this ultra-CG model to accurately reproduce the experimentally observed structural and thermodynamic properties of Phosphatidylcholine (PC) membranes in real units, including properties such as area per lipid, area compressibility, bending modulus, line tension, phase transition temperature, density profile, and radial distribution function. In our example, we specifically focus on the properties of a POPC membrane, although the developed membrane model could be perceived as a generic model of lipid membranes. To optimize the performance of the model (the fitness), we conduct a series of evolutionary runs with diverse random initial population sizes (ranging from 96 to 384). We demonstrate that the ultra-CG membrane model we developed exhibits authentic lipid membrane behaviors, including self-assembly into bilayers, vesicle formation, membrane fusion, and gel phase formation. Moreover, we demonstrate compatibility with the Martini coarse-grained model by successfully reproducing the behavior of a transmembrane domain embedded within a lipid bilayer. This facilitates the simulation of realistic membrane proteins within an ultra-CG bilayer membrane, enhancing the accuracy and applicability of our model in biophysical studies.

Lipid bilayer membranes are believed to be a vital component of the living cell machinery¹. In addition to their role in various biochemical functions such as facilitating synergy among different lipids, cellular signaling and cellular transport, lipid bilayers also function as barriers, separating cells and organelles from their external environment^{2–4}. Currently, multiple computational models are utilized to simulate biomembranes across different scales^{5–20}. Specifically, coarse-grained (CG) models, where a single interaction site represents a group of particles, have garnered considerable attention in the field in recent years due to significant alleviation of the spatial and temporal limitations of atomistic molecular simulations²¹. Pushing the degree of coarse-graining to an extreme, certain models have been developed to represent an entire bilayer section using one single interaction site (e.g., Refs. ^{22–24}). Clearly, such a simplified model does not preserve the lateral nature of biological bilayers, which involves the distribution of material across two distinct, adjoined leaflets and therefore the occurrence of so-called lipid flip-flops between them²⁵. The bilayered structure of biological membranes plays a crucial role in various significant biological processes, including pore formation^{26,27}, membrane curvature formation²⁸, membrane fusion and fission^{29,30}, and notably in lipid droplet formation, where the bilayer separates into two monolayers to encapsulate a hydrophobic droplet of neutral lipids³¹. To maintain the bilayered nature of biological membranes, several ultra-coarse-grained models have been suggested, representing a lipid molecule with only three interaction sites. These models employ either tabulated potentials derived through systematic coarse-graining approaches³² or adopt an explicit analytical form for the coarse-grained potential^{16,33–35}. In these phenomenological models, the lipid topology is typically depicted either as a linear chain comprising two (or more) hydrophobic tail beads and one hydrophilic head bead (e.g., Refs. ^{16,33,34}) or as a head group bead connected to two individual hydrophobic tails (e.g., Refs. ³⁵). There exists a pressing necessity to advance the development of ultra-coarse-grained (ultra-CG) models, aiming to facilitate highly efficient simulations of more complex and realistic biological systems and to also go beyond phenomenological levels within ultra-CG models of lipid membranes. In pursuit of this objective, these models should accurately replicate both the structural and thermodynamic properties of lipid membranes. Simultaneously, they should support multi-scale approaches, facilitating the incorporation of membrane proteins modeled at significantly higher, or even atomistic, detail.

¹Institute for Theoretical Physics, Georg-August-University Göttingen, 37077 Göttingen, Germany. ²Department of Physics, Technical University Dortmund, 44221 Dortmund, Germany. ✉email: jelger.risselada@tu-dortmund.de

Systematic coarse-graining approaches utilized in the development of realistic ultra-coarse-grained (ultra-CG) models include force-matching approaches^{36,37} and inverse Boltzmann and inverse Monte Carlo approaches^{38,39}. These approaches parameterize coarse-grained force fields by replicating the structural component of the partition function of the fine-grained system by either matching relevant radial distribution functions or (combined) forces within the fine-grained system. However, since the partition function only describes a single thermodynamic state point at equilibrium, i.e., a unique combination of pressure & temperature values, such a systematic structure-focused parameterization may not inherently translate into an optimal reproduction of elastic properties or accurate description of phase transitions. Furthermore, the model obtained often characterizes a specific system and may not be directly transferable to various systems. Alternatively, ultra-CG models are constructed using a building block approach, i.e., all nonbonded (and bonded potentials) are parametrized by the same predetermined potential function, and distinctions between interaction types are achieved through the tuning of only a few free parameters, for example, σ and ϵ in case of the famous Lennard-Jones potential. This approach enhances transferability, allowing the model to be applied to different systems. It also promotes compatibility with coarse-grained representations of other membrane constituents, including various lipid types and membrane proteins, potentially even at a different model resolution. Regrettably, the 'bottom-up' development of force-fields is a challenging task because the values of free parameters are not straightforwardly derived and demand extensive fine-tuning to optimally match the desired structural and thermodynamic properties. Consequently, models developed in this manner are typically utilized as phenomenological models only. In this study, we employ Multi-objective Evolutionary Algorithms (Genetic Algorithms), i.e., one of the unsupervised learning methods in machine learning, to construct a realistic three-site coarse-grained membrane model of POPC that optimally reproduces experimentally known structural properties as well as thermodynamic properties. We demonstrate that this ultra-coarse-grained model enables the realistic behavior of lipid membranes, including self-assembly into bilayers, vesicle formation, membrane fusion, as well as the formation of the gel phase. Furthermore, we highlight the compatibility of our model with the widely utilized Martini coarse-grained model for biomolecular simulations. This compatibility allows for the simulation of realistic membrane proteins embedded in an ultra-coarse-grained membrane environment. Significantly, to broaden the accessibility of ultra-coarse-grained (ultra-CG) models to a wider audience, we provide topology files of the potentials derived in this work in a tabulated form, which are directly compatible with the popular molecular dynamics engine GROMACS. Finally, we argue that the here-presented approach can be straightforwardly extended to more complex lipid membrane systems.

Methodology

Force field

In this study, each lipid molecule is characterized by one hydrophilic head and two hydrophobic tails, representing choline and phosphate groups, and the fatty chain of a lipid molecule, respectively. To achieve this representation, we employed a three-site coarse-grained (CG) lipid model, as illustrated in Figure 1A. It is notable that the lipid model utilized in this study is discussed elsewhere in the literature³⁵. The Hamiltonian of the three-site CG model, considering both bonded and nonbonded interactions, is presented below:

$$U = U_{bond} + U_{angle} + U_{nonbonded}, \quad (1)$$

where $U_{bond} = 0.5k_{bond}(r - r_0)^2$ and $U_{angle} = 0.5k_{angle}(\theta - \theta_0)^2$ denote the harmonic stretching and bending potential, respectively, whereas $U_{nonbonded}$ represents the nonbonded potential for intermolecular interactions. Here, k_{bond} and k_{angle} are the force constants, r_0 is the equilibrium bond length (1nm), and θ_0 is the equilibrium angle (30°). Recent studies in the field^{40–42} have underscored the importance of the attractive potential between lipid tails in implicit solvent models as a crucial factor in replicating the equilibrium structure and self-assembly properties of bilayers. Therefore, the three-site CG model introduced in this study also incorporates an attractive potential between hydrophobic tails. In general, the nonbonded interactions are divided into two components: a repulsive short-ranged interaction, rapidly diminishing with distance, utilized to simulate the excluded volume effect between CG sites, and an attractive long-ranged interaction that incorporates the effective potential of solvent effects. Ultimately, the following combined potential form is employed to characterize the nonbonded interactions among various CG sites:

$$U_{nonbonded}^{TT} = \frac{\epsilon}{5.5} \left[0.5 \left(\frac{r_{min}}{r} \right)^6 - 6 \left(\frac{r_{min}}{r} \right)^{0.5} \right] \quad (2)$$

$$U_{nonbonded}^{HH/HT} = \frac{0.4\epsilon}{5.5} 0.5 \left(\frac{r_{min}}{r} \right)^6, \quad (3)$$

where ϵ represents the potential well depth and r_{min} is the distance of the minimum potential. To mitigate potential artifacts at the cutoff position, the nonbonded potentials are modified using a shift function.

As a result, both the potential and forces exhibit continuity and a smooth decay to zero between the starting position of the shift function, denoted as r_1 , and the cutoff distance r_c (see Figure 1B). The GROMACS shift function used in this work is applied to the force function $F(r)$ and is given by⁴³:

$$S(r, r_1, r_c) = \frac{1 + A(r - r_1)^2 + B(r - r_1)^3}{r^{-(\alpha+1)}}, \quad (4)$$

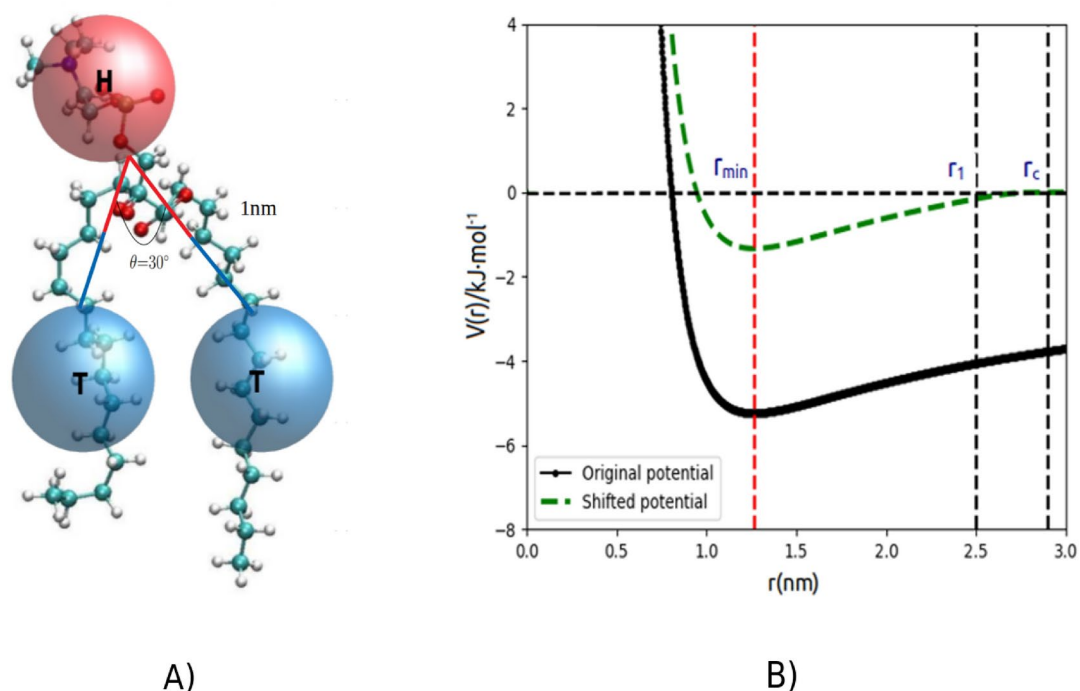


Figure 1. (A) The schematic drawing depicts the coarse-grained model of a lipid molecule. The hydrophilic bead, represented by the letter “H,” corresponds to the cholesterol and phosphate groups of the lipid, whereas the two hydrophobic tail beads labeled “T” symbolize the fatty acid chains. (B) Original and shifted potential functions: Intermolecular potential energy versus distance of a pair of beads.

where α is the interaction power which is 0.5 and 6 for attraction and repulsion in our three-site CG model, respectively. In addition, the constants A and B are obtained according to the conditions that the function should be smooth at r_1 and r_c , hence:

$$A = -\frac{(\alpha + 4)r_c - (\alpha + 1)r_1}{r_c^{\alpha+2}(r_c - r_1)^2} \quad (5)$$

$$B = \frac{(\alpha + 3)r_c - (\alpha + 1)r_1}{r_c^{\alpha+2}(r_c - r_1)^3} \quad (6)$$

Therefore, the total force function is obtained as follows:

$$F_s(r, r_1, r_c) = \frac{1}{r^{\alpha+1}} + A(r - r_1)^2 + B(r - r_1)^3 \quad (7)$$

Molecular dynamics simulations

Langevin dynamics (LD) was employed for all simulations with a time step of 0.15 ps, utilizing the molecular dynamics (MD) package GROMACS⁴⁴. The equation of motion in LD for a system of N particles with masses M is obtained as follows:

$$M\ddot{X} = -\nabla U^{LD} - \gamma\dot{X} + \sqrt{2k_B\gamma TR(t)}, \quad (8)$$

where $U(x)$ represents the interaction potential of the particles; ∇ is the gradient operator; \dot{X} and \ddot{X} denote velocity and acceleration, respectively; γ is the friction coefficient, and the third term is the noise term with T as the temperature, k_B Boltzmann's constant and $R(t)$ as delta-correlated stationary Gaussian process with zero-mean. Periodic boundary conditions were applied across all three dimensions. Steepest descent algorithm was used to minimize the energy of all configurations, succeeded by a 200 ps molecular dynamics (MD) simulation in the canonical (NVT) ensemble to equilibrate the systems at a temperature of 315 K using the Nose-Hoover thermostat. Pressure coupling was implemented using the Berendsen barostat. It is noted that various bilayer properties, including area per lipid, area compressibility, and bending modulus, were measured by imposing zero surface tension in the xy plane with a compressibility of 1×10^{-5} 1/bar. Simultaneously, pressure was maintained at 1 bar in the z plane with a compressibility of 0. A relaxation time of 1 ps was employed for both temperature and pressure coupling.

Computational analysis of membrane properties

Area per lipid

Area per lipid is regarded as a key parameter in the characterization of membranes⁸. The area per lipid plays a crucial role in fine-tuning force fields for membrane systems, making it an indispensable factor. This suggests that accurately calculating structural properties, including the area per lipid, is essential for the molecular dynamics simulation of membranes⁴⁵. Furthermore, the average area per lipid is associated with various membrane properties, encompassing acyl chain ordering, molecular packaging, and area compressibility⁴⁶. Moreover, structural dimensions of bilayers, particularly hydrophobic matching and area per lipid, are pivotal parameters in research domains like in vitro studies of integral membrane proteins or peptides, where the frequent reconstitution of membrane lipids is essential^{47–52}. In the context of phospholipid membranes, experimental studies have indicated that the typical range for the area per lipid is reported to be between 0.55 and 0.75 nm²⁵³. As an example, Kucerka et al. employed hybrid electron density models and determined the area per lipid for POPC, the most abundant lipid in animal cell membranes, to be 0.683 ± 0.015 nm²⁵⁴. Experimentally determining membrane properties can be challenging. Consequently, atomistic and coarse-grained molecular dynamics simulations have been extensively employed as an alternative, robust, and reliable method for calculating various membrane properties. In molecular dynamics simulations conducted under an NPT ensemble, the dimensions of the simulation box may fluctuate due to pressure coupling. Therefore, in this study, the average area per lipid was determined by calculating the size of the simulation box in the xy plane using the following relation⁵⁵:

$$A_{\text{Lipid}} = \frac{\langle L_x L_y \rangle}{N_{\text{lipids}}}, \quad (9)$$

where L_x and L_y represent the membrane plane simulation box dimensions and N_{lipids} denotes the number of lipids present in one leaflet of the bilayer. Typically, the time evolution of the average area per lipid during the simulation is monitored and visualized graphically. Indeed, fluctuations around an average value are anticipated for the area per lipid. Upon reaching convergence, the average area per lipid is expected to stabilize, showing minimal changes as the system reaches equilibrium. Indeed, Frigini et al.⁴⁶ emphasized that the time-dependent behavior of the area per lipid can serve as a valuable criterion for evaluating whether a membrane system has achieved a steady state. They conducted an investigation into the impact of glyphosate in both neutral (GLYP) and charged (GLYP2) forms on a fully hydrated DPPC lipid bilayer. This was accomplished by observing the time evolution of the area per lipid. Simulation results indicated that, for both charged glyphosate and neutral glyphosate at various G:L ratios, the systems reached stability after approximately 5–10 nanoseconds (ns) of simulation time⁴⁶. In a comparative study among multiple phospholipids, including POPC including POPC(C₄₂H₈₂NO₈P), DOPE(C₄₁H₇₈NO₈P), and POPE (C₃₉H₇₆NO₈P), Zhang et al. have reported that POPC and POPE phospholipids exhibit the highest (0.683 nm²) and lowest (0.588 nm²) area per lipid, respectively⁵⁶. Angladon et al. conducted coarse-grained molecular dynamics simulations at 310 K and reported that the area per POPC molecule is 0.676 nm², which agrees well with the reference value of 0.683 nm²⁵⁷. Wang et al. developed an implicit solvent coarse-grained (CG) model for quantitative simulations of POPC bilayers. In the smallest membrane system comprising 72 lipids, equilibrated for about 30 ns in the NP_{zz}AT ensemble, the calculated area per lipid was found to be 0.683 nm². This alignment with the saturated area measured in experiments indicates that the solvent-free CG force field can effectively reproduce the saturated area per lipid obtained from experimental data⁴⁰. Arnarez et al. utilized the Dry Martini force field to replicate various lipid membrane properties, including the area per lipid, bending modulus, and the coexistence of liquid-ordered and disordered domains. For instance, the area per lipid for a POPC bilayer in Dry Martini (0.64 ± 0.1 nm²) was smaller in comparison to that of the Wet Martini (0.66 ± 0.1 nm²). In contrast, the experimental data ranged from 0.62 to 0.68 nm² and was measured between 297 and 323 K¹⁸. Employing a new coarse-grained (CG) model for lipid and surfactant systems, Marrink et al. determined the equilibrium area per lipid for DPPC to be 0.64 nm² at 323 K⁵⁸. Guo et al. utilized a coarse-grained model for peroxidized phospholipids, built upon the MARTINI lipid force field. They conducted simulations with lipids in a box, considering lateral x and y directions and the transverse z direction, at a temperature of 300 K. As a result, the equilibrium area per lipid for POPC and DOPC phospholipids was measured to be 0.640 ± 0.0008 and 0.669 ± 0.0006 nm², respectively⁵⁹. Srivastava et al. determined the area per lipid for highly coarse-grained lipid models of DLPC, DOPC, and mixed DOPC/DOPS to be 0.5763 nm², 0.662 nm², and 0.614 nm², respectively. These values are slightly lower than those calculated in atomistic systems, which are 0.582 nm² and 0.674 nm² for DLPC and DOPC, respectively. This discrepancy is attributed to the softer effective potential energy landscape in coarse-grained systems, leading them to exhibit a smaller area per lipid value than corresponding atomistic systems³². Porasso et al.⁶⁰ employed molecular dynamics simulations to investigate the temporal evolution of various lipid properties, including the surface area per lipid for two lipid bilayer systems, namely DPPC (DiPalmitoylPhosphatidylCholine) and DPPS (DiPalmitoylPhosphatidylSerine). The average area per lipid for DPPC was determined to be 0.663 ± 0.008 nm² in the absence of salt at a temperature of 330 K^{61–63}. Furthermore, the area per lipid for the bilayer formed by 288 DPPS in the presence of CaCl₂ at 0.25N was calculated as 0.522 ± 0.007 nm²⁶⁰. It is important to highlight that in our evolutionary runs, the goal is to achieve a specific target area per lipid of 0.68 nm², a value commonly reported for the area per lipid of a POPC membrane in both experiments and simulations. Nevertheless, converged solutions within the broader range of 0.55–0.77 nm² are also taken into consideration, aligning with experimental data obtained for PC membranes, as discussed earlier.

Area compressibility modulus

Area compressibility stands as a crucial elastic characteristic often used to gauge the reliability and efficiency of membrane models. Compressibility, defined as the volume change of a material under stress, applies specifically to the lipid bilayer's resistance to isotropic area dilation, characterized by an area compressibility modulus, K_A ⁸. Experimental techniques, such as micropipette aspiration, have reported the area compressibility modulus for POPC bilayers in the range of 208–237 mN/m at 21 °C⁶⁴. Furthermore, alternative experimental values ranging from 180 to 330 mN/m, acquired through Infrared measurements at 25 °C, have been reported for POPC and SDPC in other studies⁶⁵. In MD simulations, various methods exist to compute the area compressibility modulus of a bilayer membrane. In this study, area compressibility or the stretching modulus K_A is calculated through an area fluctuation analysis according to the following equilibrium statistical relation:

$$k_A = \frac{k_B T}{4(\langle L_x^2 \rangle - \langle L_x \rangle^2)}, \quad (10)$$

where $\langle L_x \rangle$ and $\langle L_x^2 \rangle$ denote the mean value and the quadratic fluctuation of the lateral size of the anisotropic simulation box. Guo et al. computed the stretching modulus for Martini coarse-grained POPC and DOPC bilayers, each comprising 512 lipids, to be 379 ± 21 and 357 ± 19 mN/m, respectively. However, these values were observed to overestimate the accepted experimental reference values. Conversely, in larger systems containing 8192 lipids, better agreement was achieved with compressibilities of 245 ± 42 and 230 ± 27 mN/m⁵⁹. Additionally, Xu et al. determined the area compressibility modulus of a tensionless coarse-grained bilayer membrane model through the thermal fluctuations of the membrane area, yielding $K_A = 157.5$ mN/m, which was smaller than the experimental data for a DPPC bilayer, i.e., $K_A = 231 \pm 20$ mN/m³⁵. Similarly, utilizing fluctuations in the membrane area per lipid, Marrink et al. observed that for a large bilayer patch comprising 6400 lipids, the area compressibility was determined to be $K_A = 260 \pm 40$ mN/m which is lower compared to the area compressibility of $K_A = 400 \pm 30$ mN/m for a smaller system consisting of 256 lipids. The observed discrepancy was attributed to the contribution of undulatory modes in the large system, leading to a decrease in the area compressibility. The value obtained for the large system aligns well with the experimental value reported for DPPC, which is $K_A = 231 \pm 20$ mN/m⁵⁸. Furthermore, Wang et al. investigated the temperature dependence of the area compressibility by analyzing the standard deviation in the area per lipid. They identified a correlation between the compressibility modulus and the phase of the system. They observed that systems with a compressibility modulus lower than 210 mN/m existed in the liquid phase, whereas those with a compressibility modulus higher than 420 mN/m were in the gel phase⁶⁶. This discovery suggests that the area compressibility can serve as an indicator of membrane phase transitions. Moreover, as an alternative approach, Saeedimasine et al. calculated the area compressibility modulus by determining the derivative of surface tension with respect to areal strain. Consequently, the calculated area compressibility modulus for Pure-POPC was found to be 188 mN/m at the atomistic level. In contrast, at the coarse-grained level, the values varied between 282 and 297 mN/m for different system sizes⁶⁷. Finally, it is noteworthy to mention that our evolutionary trials are designed to achieve a specific target area compressibility of 230 mN/m for a POPC membrane. This value falls within the reported range for POPC membrane area compressibility obtained through both experimental and computational techniques. However, we also consider solutions within the broader range of 180–330 mN/m, as discussed earlier, since these values align with experimental data for PC membranes.

Line tension

Line tension, alternatively known as edge energy, refers to the excess free energy arising from the existence of free edges³⁵. The line tension of pores in phospholipid bilayers plays a crucial role in pore-mediated molecular transport techniques. The existence of a pore within the bilayer acts as a gateway for extracellular molecules to enter the cell. This pore facilitates passive molecular transport, enabling molecules to traverse the membrane without requiring additional energy expenditure⁶⁸. Various methods, encompassing both experimental and computational approaches, have been utilized to measure line tension. In general, line tension or edge tension in diverse membrane compositions has been experimentally measured and observed to fall within the range of 8–42 pN^{69–72}. For example, Portet et al. introduced a novel technique that allows for the measurement of edge tensions in lipid membranes. This method involves the electroporation of giant unilamellar vesicles and the subsequent analysis of the dynamics of pore closure. To assess the edge tension, the researchers applied this technique to membranes with various compositions, including dioleoylphosphatidylcholine (DOPC) and mixtures of DOPC with cholesterol. As a result, the line tension of DOPC was obtained to be 27.7 ± 2.5 pN and 36.4 ± 1.9 pN for membranes made of DOPC/cholesterol at a 5:1 molar ratio. Additionally, the reported range of line tensions for a phosphatidylcholine (PC) bilayer membrane typically falls between 6.5 and 30 pN⁷³. In this study, molecular dynamics simulations are employed to calculate the line tension γ_L from the anisotropy of the pressure within the simulation box. Specifically, this calculation is performed for a bilayer membrane with two free edges in the x-direction, while the membrane is periodically extended in the y-direction (Figure 2)⁷⁴:

$$\gamma_L = -\frac{L_x L_z}{2} \left(P_{yy} - \frac{P_{xx} + P_{zz}}{2} \right), \quad (11)$$

where P_{xx} and P_{yy} represent the tangential and P_{zz} denotes the normal component of the diagonal pressure tensor, L_x and L_z are the simulation box dimensions and the prefactor $\frac{1}{2}$ is included to account for the fact that the system has two free edges. Employing a similar methodology, Miyazaki et al. obtained the line tension of a pure ribbon-like POPC bilayer membrane to be 35.1 pN⁷⁵. Similarly, Y. Jiang et al. utilized pressure anisotropy

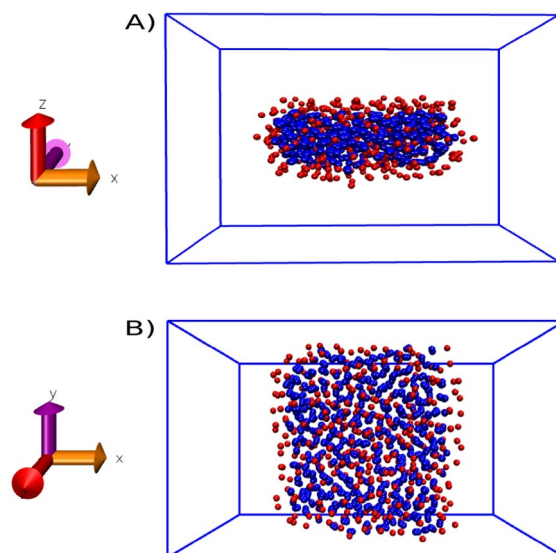


Figure 2. Simulation setup for calculation of line tension in a membrane consisting of 376 lipids. (A) Side view. (B) Top view.

in the simulation box and calculated the line tension within the range of 10–30 pN for the edge of a pure dimyristoylphosphatidylcholine (DMPC) membrane⁷⁶. Alternatively, employing an implicit solvent coarse-grained (CG) model for POPC bilayers, Wang et al. utilized a straightforward formula, i.e. $\gamma_L = -\frac{1}{2} \langle \sigma_{xx} \rangle L_y L_z$ to compute the line tension, where σ_{xx} represents the xx-component of the stress tensor, L_y and L_z represent the length of the simulation box in y and z dimensions, respectively. As a result, the calculated line tension value ($\gamma_L = 29 \pm 12$ pN) fell well within the range of both experimental and atomistic simulation reference values for PC membranes⁴⁰. Given the scarcity of reliable statistical data concerning the experimentally determined value of the POPC membrane line tension, as highlighted in literature⁷⁷, our evolutionary runs are aimed at achieving a specific target line tension of 10 pN. This value aligns with the experimental range of PC membrane line tension, typically reported between 6.5 and 30 pN.

Bending modulus

The membrane's bending modulus plays a crucial role in regulating the out-of-plane bending deformation of a membrane as described in the Canham-Helfrich elastic model⁵⁹. Bending modulus values were obtained through experimental measurements. Specifically, Rawicz et al. utilized the micropipette vesicle aspiration technique to measure the bending modulus of fluid phase phosphatidylcholine (PC) membranes, i.e. POPC and DOPC, resulting in values of approximately 9.0×10^{-20} J and 8.5×10^{-20} J, equivalent to around $20 k_B T$ and $19 k_B T$, respectively⁶⁴. In computer simulations, as well as in the present study, the bending modulus for lipid bilayers is commonly determined by computing the membrane undulation spectrum. While the theoretical concept of using the membrane undulation spectrum is straightforward, its practical application is highly sensitive to errors. Moreover, it is only applicable in the “long-wavelength” limit and exhibits slow convergence, as noted elsewhere⁷⁸. Utilizing a Helfrich-type continuum model, the bilayer can be represented as a single mathematical surface, denoted as $u(x, y)$, with a minimum energy state. This model facilitates the characterization of the bilayer. Consequently, the Hamiltonian for thermally excited energy fluctuations takes the following form⁷⁹:

$$H\{u(x, y)\} = \frac{1}{2} \int \int (K_c |\nabla^2 u(x, y)|^2 + \gamma |\nabla u(x, y)|^2) dx dy, \quad (12)$$

where k_c and γ represent bending modulus and surface tension, respectively.

The membrane fluctuations are analyzed by employing a Fourier decomposition of the surface:

$$u(\mathbf{r}) = \sum_{\mathbf{q}} u(\mathbf{q}) e^{i\mathbf{q} \cdot \mathbf{r}}, \quad (13)$$

where $\mathbf{r}=(x, y)$ is a two-dimensional real space vector and $\mathbf{q}=(q_x, q_y)$ being the two-dimensional reciprocal space vector. Due to the quadratic nature of energy in the amplitudes (Eq.12), the equipartition theorem is applied, stating that each Fourier mode has an average energy of $k_B T/2$ ⁸⁰. In the absence of any in-plane tensions ($\gamma=0$), thermal fluctuations as a function of wave number q are expressed by the following equation⁸¹:

$$\langle |u_q|^2 \rangle = \frac{k_B T}{L^2 k_c q^4} \quad (14)$$

,where $\langle |u_q|^2 \rangle$ represents the ensemble average of the energy of the undulatory modes, k_c denotes the bending modulus, k_B is the Boltzmann's constant, T is the temperature and L is the periodic simulation box side. If the membrane model accurately replicates the continuum behavior dictated by the Helfrich energy density, a deviation from the q^{-4} behavior is anticipated at short wavelengths. This divergence arises because the continuum behavior is only applicable to high-wavelength undulations¹. Various methods can be utilized to extract the height spectrum from simulations of bilayer membranes. These methods enable the extraction of the bending modulus, with the Real Space (RS) method being a commonly employed approach for this purpose. In the Real Space (RS) method, the membrane height field, represented as $u(x, y)$, is interpolated onto a uniform grid of nodes in a two-dimensional ($N \times N$) grid (see Figure 3). To mitigate technical complications, such as “empty” grid cells, it is advisable to select a larger value for N in the Real Space (RS) method. Increasing the size of the grid enhances the accuracy of the interpolation and improves the resolution of the height-height spectrum. Moreover, to map the membrane position onto the defined grid, a straightforward approach involves selecting a reference position in each lipid, such as any of the particles or beads belonging to the lipid or the average position of the phosphocholine head group. The height of the reference position for all lipids within a given grid point is then computed, and the average value of the heights is assigned to that grid point. Another approach to mapping lipid molecules onto a uniform grid, without the need for a reference point, involves directly mapping every microscopic degree of freedom. This entails mapping each individual heavy atom or coarse-grained bead to its nearest grid point, irrespective of the lipid to which it belongs. Finally, to acquire the mean power spectrum, the power spectra of all bilayer configurations are averaged. Indeed, numerous publications in the literature have employed the computational approach described above to calculate the bending modulus of membranes. Guo et al. conducted calculations of the bending modulus utilizing a coarse-grained model specifically developed for peroxidized phospholipids, built upon the MARTINI lipid representation. Their findings revealed bending modulus values of approximately 30.0 and 28.3 $k_B T$ for POPC and DOPC, respectively⁵⁹. Yuan et al. employed a one-particle-thick fluid membrane model, wherein each particle represented a cluster of lipid molecules, to derive the bending rigidity of membranes. Their computations resulted in bending rigidity values falling within the experimental range of 12–40 $k_B T$ ²². Srivastava et al. estimated the bending modulus for highly coarse-grained DLPC membranes to be 6.65×10^{-20} J, approximately equivalent to 15 $k_B T$. Similarly, the bending modulus for a DOPC model was found to be $(7.64 \pm 0.671) \times 10^{-20}$ J, corresponding to 18–19 $k_B T$ ³². Chacon et al. calculated a bending modulus of 21 $k_B T$ for POPC membranes at 320 K using coupled undulatory (CU) modes to analyze membrane fluctuations. This approach yielded excellent estimates of the bending modulus in line with proposed theoretical predictions⁸². In a distinct study by Xu et al., a bending modulus value of 4.35×10^{-20} J, roughly equivalent to 10 $k_B T$, was reported for a coarse-grained DPPC bilayer³⁵. In conclusion, our evolutionary runs aim to reach a specific target bending rigidity of 12 $k_B T$ for a pure POPC membrane, as measured in experiments using techniques like electro-deformation or Fluctuation analysis⁸³. Additionally,

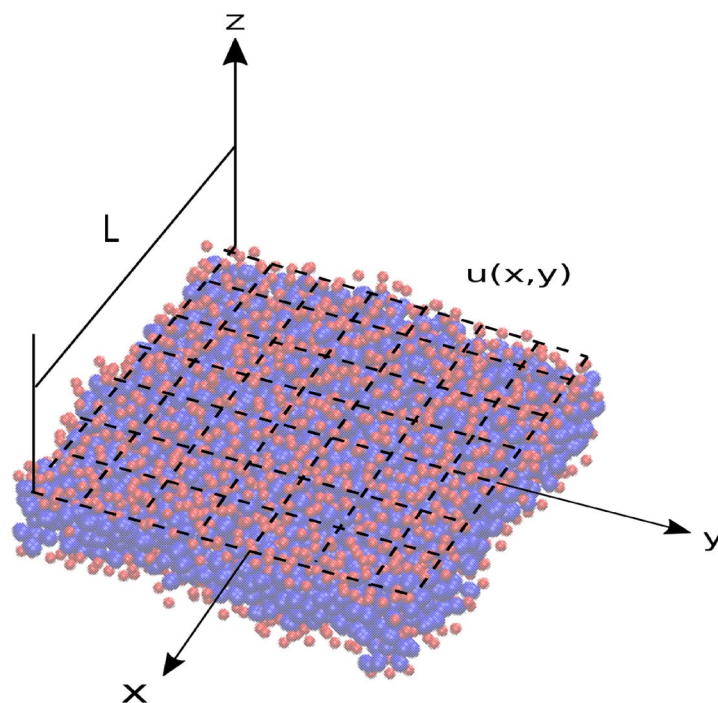


Figure 3. The schematic depicts the height function in a bilayer membrane, $u(x, y)$, fitted to bead positions observed in selected frames throughout the simulation trajectory.

solutions converging to experimentally observed bending rigidity values between 8 and 42 $k_B T$ are deemed valid and fall within an acceptable range.

Phase transition temperature

Considerable research efforts have been directed toward investigating the phase transition in biological membranes, with a particular focus on the ordered (gel) and disordered (liquid-crystalline) lipid phases. These phases play a crucial role in understanding the overall dynamics and structural properties of membranes, given their direct association with the individual lipid components⁸⁴. Numerous experimental studies have explored the phase behavior of phospholipids, employing a variety of techniques. Suurkuusk et al. conducted a comprehensive investigation using calorimetric and fluorescent studies on multi-lamellar vesicles of DPPC. In their study, they observed a heating scan of multi-lamellar vesicles that exhibited a phase transition process reminiscent of a first-order transition, a finding subsequently confirmed through fluorescent measurements⁸⁵. Additionally, Davis et al. employed deuterium magnetic resonance to examine the liquid-crystalline, gel phase, and phase transition of multi-lamellar labeled-DPPC vesicles. Their quantitative measurements based on deuterium spectra revealed a sharp decline in moment values as the vesicles were heated, indicating a first-order transition⁸⁶. In addition to experimental approaches, computational methods have played a significant role in investigating the phase transition in lipid bilayer membranes. For example, the phase transition temperature was determined by observing a sharp rise in enthalpy during the transition from the gel to liquid crystalline states⁸⁷. Furthermore, efforts were made to address the issue of substantial hysteresis, which can pose challenges in accurately determining the phase transition temperature for the liquid-gel transition. To overcome this challenge, simulations were conducted on a system comprising two distinct halves, with one half in the gel phase and the other in the liquid-disordered phase (refer to Figure 4). This configuration effectively diminishes or eliminates the nucleation barrier necessary to initiate either the liquid-disordered or gel phase, leading to a considerable reduction in hysteresis⁸⁸. On the other hand, de Vries et al. conducted an atomistic molecular dynamics study to investigate the phase behavior of lecithin (DPPC) in water. Their simulation results revealed the spontaneous formation of a ripple gel phase when a fully hydrated bilayer was cooled below the phase transition temperature⁷. Accordingly, Leekumjorn et al. employed atomistic molecular dynamics simulations to thoroughly examine the structural properties of saturated DPPC and DPPE lipid bilayers in the vicinity of

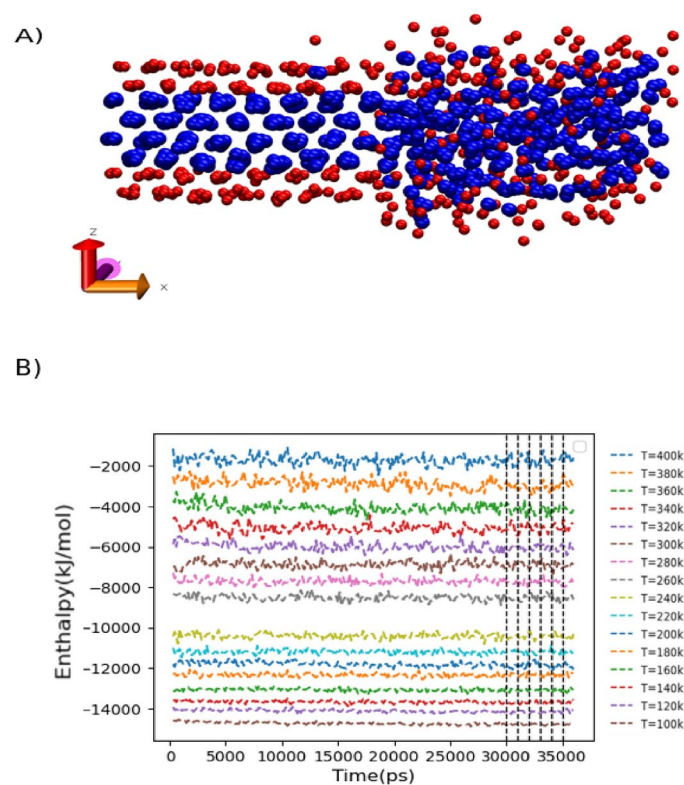


Figure 4. Phase transition temperature calculation. (A) The simulation snapshot depicts the initial configuration of a bilayer, with one half of the lipids in the gel phase (left) and the other half (right) in the fluid phase. (B) Enthalpy versus time at different temperatures ranging from 100 to 400 K. A sudden increase in enthalpy within the temperature range from 240 to 260 K is indicative of a phase transition from the gel to the liquid-disordered phase. The phase transition temperature is then assigned to the middle of the identified phase transition zone, i.e., 250 K.

the main phase transition. Despite their similar chemical structures, with the only difference in their choline and ethanolamine groups, DPPC and DPPE exhibited contrasting transformation mechanisms from an ordered (gel) to a disordered (liquid-crystalline) state. Leekumjorn et al. used the area per lipid and bilayer thickness as criteria to determine the first-order phase transition temperature of DPPC and DPPE bilayers through their simulations. According to their findings, the transition temperature was approximately 305 K for DPPC and 320 K for DPPE. These values were compared to experimental measurements reporting transition temperatures of 315 K for DPPC and 337 K for DPPE. Additionally, using molecular dynamics with Alchemical Steps, Fathizadeh et al. calculated the phase transition temperature from the gel to fluid phases in pure DPPC bilayers, obtaining values within the range of 303.15–323.15 K⁸⁴. Coarse-grained approaches have proven effective in determining the phase transition temperature of lipid bilayers. Marrink et al. conducted simulations using a coarse-grained (CG) model to investigate the transformation between a gel and a fluid phase in DPPC bilayers. They employed a cooling technique on bilayer patches and proposed a four-stage reversible process consisting of nucleation, growth, limited growth, and optimization. Extrapolating their findings to a macroscopic bilayer, they obtained a transition temperature of 295 ± 5 K, which was in semi-quantitative agreement with the experimental phase transition temperature value for DPPC (315 K)⁸⁹. Kociurzynski et al. determined the phase transition temperature of pure DPPC by observing a sharp increase in the area per lipid during the transition from the gel to liquid crystalline states. Their simulation results indicated a transition temperature of approximately 305 K, showing a slight discrepancy of about 10 K compared to the experimental value of 315 K for DPPC. This difference suggested a small deviation between the simulated and experimental transition temperatures^{89,90}. Mirzoev et al. employed a systematic structure-based coarse-graining approach with a grained DMPC lipid model to examine the state point dependence of lipid bilayers. Their simulations covered a temperature range of 200–350 K in 25 K intervals. Through the analysis of the area per lipid and lateral diffusion coefficient, they observed a clear transition between 225 and 250 K, signifying a gel-liquid phase transition⁹¹. Eventually, Wang et al. utilized coarse-grained solvent-free membrane models to explore the phase transformation in DOPC and POPC membranes. Their simulations indicated that the phase transition in a DOPC membrane occurs within the temperature range of 265–295 K, while in a POPC membrane, it takes place between 280 and 310 K. The values obtained from the simulations were slightly lower than the experimental transition temperatures reported for DOPC (around 253 K) and POPC (around 270 K)^{73,92,93}. As a result, our evolutionary runs aim to achieve a phase transition temperature of 270 K for a POPC membrane, aligning with experimental observations. Additionally, we consider solutions converging toward the experimentally observed phase transition temperature for a PC membrane, spanning from 250 to 315 K.

Radial distribution functions

Radial distribution functions (RDF) are crucial in the analysis of biological membrane systems using molecular dynamics (MD) simulations, offering valuable insights into the system's short and medium-range structural information⁹⁴. In this work, the Radial Distribution Function (RDF) is defined according to the formulation by Hub et al.⁹⁵:

$$g(r) = \frac{n(r, r + \Delta r)}{\rho \Delta V_{2D}} \quad (15)$$

,where r is the distance in the lateral direction, $\Delta V_{2D} = 2\pi r \Delta r$ being the volume of a shell with radius r and thickness Δr and $n(r, r + \Delta r)$ is the number of the sites within this shell. Given the significant level of coarse-graining in this study, a comparison was exclusively carried out using the Earth Mover's Distance (EMD) technique for the tail-tail radial distribution function (RDF). The comparison involved our three-site coarse-grained (CG) model and the tail-tail RDF derived from an atomistic molecular dynamics simulation of a POPC membrane. The EMD method quantifies the minimal cost required to transform one distribution into another⁹⁶. Hence, if the two RDF distributions—the normalized RDF related to the coarse-grained model and the atomistic RDF—are identical, the resulting EMD value will be 0, signifying their complete similarity.

Density profile

The distribution of mass along the z-axis of the membrane is visualized by computing the mass density profile across the bilayer⁸. The density profile can offer insights into various properties of the membrane, including the distribution of molecules and the presence of structural features⁹⁷. In calculating the mass density profile, it is crucial to account for potential fluctuations in the center of mass (COM) of the membrane during the simulation. Hence, in each frame, the coordinates (x, y, z) of all atoms are determined relative to the instantaneous center of mass (COM), with the z-coordinate set to 0. This approach enables an accurate assessment of the distribution of mass along the membrane's z-axis. Similar to the radial distribution function (RDF), we conducted an Earth Mover's Distance (EMD)-based comparison of the mass density profile between the head groups in our three-site coarse-grained (CG) model and the atomistic model of a POPC membrane. In this context, if the normalized density profile of the coarse-grained model perfectly matches that of the atomistic density profile, the Earth Mover's Distance (EMD) value becomes zero.

Genetic algorithms (GA)

Genetic Algorithms (GA) are developed based on genetic laws and Darwin's theory of evolution aimed at enhancing the optimization of problems with numerous variables⁹⁸. For example, they are utilized in the determination of molecular docking within protein-DNA complexes or in the generation of new pharmacophore

models through electron density topology analysis^{99,100}. Generally, a Genetic Algorithm (GA) typically consists of five distinct phases, as outlined below:

Initial population

Within a Genetic Algorithm (GA), the search space is represented by a population of individuals, where each individual serves as a potential solution to a specific problem. The individuals in a Genetic Algorithm (GA) are encoded as genomes, which are vectors containing a predefined number of parameters. During the initialization phase, each genome is assigned random values within a predetermined range. The random initialization ensures diversity within the initial population, facilitating a broader exploration of the search space. Subsequently, a cost assessment is carried out to evaluate the quality of each genome. This assessment entails comparing the outcomes generated from molecular dynamics (MD) simulations, employing the parameters encoded in the genome, with the target values from a reference database. The cost score is then determined by assessing the degree of agreement between the molecular dynamics (MD) simulation results and the specified target values. Genomes that do not meet predefined criteria or fail to attain a satisfactory cost score are deemed unfit for further consideration. The eliminated unfit genomes are replaced with new genomes in the subsequent generations, maintaining the population size. In Genetic Algorithms (GA), the initial population size is often chosen to be an even number¹⁰¹.

Selection

The selection operator in a Genetic Algorithm (GA) plays a crucial role in exploiting the advantageous traits of the fittest candidate genomes, aiming to improve and enhance them across successive generations. This accelerates the GA's convergence towards a preferred solution for the optimization problem under study¹⁰². Numerous selection methods have been reported in the literature, among which are roulette wheel selection (RWS), tournament selection (TOS), and linear rank selection (LRS). In this study, roulette wheel selection (RWS) is utilized, due to its widespread adoption as a method in Genetic Algorithms (GA) for selecting genomes for the succeeding generation. The roulette wheel selection (RWS) emulates the concept of survival of the fittest. This method efficiently favors genomes with lower cost scores, providing them with a greater probability of being chosen for the next generation¹⁰³. Specifically, the RWS assigns a probability of selection P_i to each genome i based on its cost value. The probability P_i for each genome is defined as:

$$P_i = \frac{F_i}{\sum_{j=1}^{PopSize} F_j}, \quad (16)$$

where F_i corresponds to the cost of the i th genome. Then, a series of N random numbers are generated and compared with the cumulative probability $C_i = \sum_{j=1}^i P_j$ of the population. Ultimately, the fit genome i is chosen and transferred into the new population provided that $C_{i-1} < U(0, 1) \leq C_i$, where $U(0, 1)$ is a uniform random number generated between 0 and 1. In RWS, genomes are allocated to consecutive segments on a line, where the size of each genome's segment is proportionate to its fitness level. Following the generation of a random number, the genome to be selected is determined by identifying the segment on the line that corresponds to the generated random number. The described process continues until the desired number of genomes, often referred to as the mating population, is achieved. This method shares similarities with a roulette wheel, as the size of each segment is directly proportional to its fitness.

Crossover

Undoubtedly, crossover stands out as one of the most crucial operators in Genetic Algorithms (GA), playing a key role in generating offspring that will constitute subsequent generations. The primary objective of the crossover operation in GA is to broaden the exploration of the search space, thereby aiding in the generation of enhanced genomes in future generations¹⁰⁴. The standard crossover operator involves taking two parent genomes and producing two offspring genomes. The underlying concept behind this operation is that the offspring genomes have the potential to display superior characteristics compared to their parents by inheriting the best traits from each parent. This blending of genetic material from both parents enables the exploration of new genetic combinations that may contribute to improved solutions in the evolutionary process. In this study, single-point crossover is utilized, given its widespread acceptance as one of the most commonly employed crossover operations. In single-point crossover, the two parental genomes are divided into two segments at a randomly determined crossover point. A new child genotype is then generated by exchanging the second segment of the first parent with the second segment of the second parent¹⁰⁵. Figure 5 illustrates the single-point crossover (SPC) operation.

Mutation

The mutation operator holds significant importance within a Genetic Algorithm (GA) as it plays a crucial role in preventing premature convergence by introducing a random element in the population's evolution¹⁰⁶. The mutation operator is applied to the genomes generated through the crossover operation and involves altering one or more bits within a specific genome, with a probability denoted as $P_m \in [0, 1]$, referred to as the mutation probability. Hence, the mutation probability governs the frequency at which the bits within a particular genome undergo mutation. Significantly, the mutation probability is adjusted based on the behavior of the population. When the population becomes trapped in a local optimum, the mutation probability is increased. Conversely, when the population is widely spread across the solution space, the mutation probability is decreased. However, it

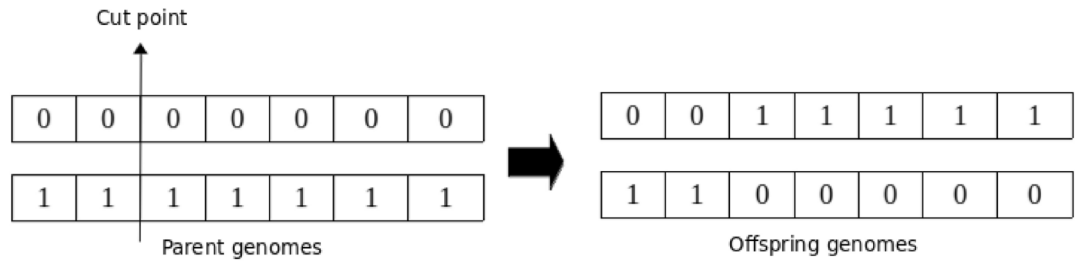


Figure 5. Schematic illustration of single-point crossover. The arrow indicates the crossover cut point.

is usually chosen within the range of 0.001–0.05¹⁰⁷. Various mutation operators have been proposed for different solution representations, including Gaussian mutation, random mutation, and polynomial mutation^{108–110}. In this study, Gaussian mutation is employed as the selected mutation operator, given its widespread recognition as one of the most commonly used mutation operators. Accordingly, for each bit in a given genome, a random value is sampled from a normal distribution with a mean of 0 and a standard deviation of 1. Subsequently, the randomly generated value is compared to a predefined threshold, typically set to 0.05. If the generated value is lower than the threshold, the mutation is considered valid and accepted. However, if the value exceeds the threshold, the mutation is rejected, and no modifications are made to the corresponding bit position within the genome. Once a mutation is accepted, an adaptive mutation operation is executed to introduce additional variability. This operation involves multiplying a sigma value, representing the standard deviation derived from the Gaussian distribution of the bits at a particular position from the previous generation, by the randomly generated number. The resulting value obtained from the adaptive mutation operation is then added to the selected bit within the genome, as illustrated below¹¹¹:

$$x_i \in \mathbb{R} \longrightarrow x_i + \sigma N(0, 1), \quad (17)$$

where i represents the i th bit of genome and σ denotes the previously discussed standard deviation.

Termination

At this stage of the procedure, the average cost of the best genomes corresponding to a given generation is compared to a threshold, i.e., the target average cost. If the obtained average cost is below the threshold (in the case of a minimization problem) or above the threshold (for a maximization problem), the Genetic Algorithm (GA) terminates. Otherwise, the loop continues until a subsequent generation successfully achieves the goal. It is worth noting that there are alternative stopping criteria for the Genetic Algorithm (GA). These criteria include setting an upper limit on the number of generations or the number of cost function evaluations. Additionally, the Genetic Algorithm (GA) may terminate if the probability of achieving significant changes in the subsequent generation becomes exceedingly low, as mentioned in Ref.¹¹².

Cost function

In this study, the weighted sum method, widely recognized as one of the most prominent approaches for addressing multi-objective optimization problems, is employed to calculate the cost of each genome¹¹³. This method involves consolidating multiple objectives into a single aggregated objective function by assigning a weighting coefficient to each respective objective function. Hence, the cost corresponding to each genome is calculated using the following equations:

$$Cost = Cost_1 + Cost_2 \quad (18)$$

$$Cost_1 = \sum_{i=1}^5 w_{1,i} * \left[\frac{(Simulated_Property)_i}{(Target_Property)_i} - 1 \right]^2 \quad (19)$$

$$Cost_2 = \sum_{i=6}^7 w_{2,i} * EMD[(CG_Simulated_Property)_i, (Atomistic_Target_Property)_i] \quad (20)$$

,where w_i represents the selected weighting coefficient for the i th objective function. Therefore, equation (19) calculates the cost value for five membrane properties, i.e., area per lipid, area compressibility, line tension, bending modulus, and phase transition temperature, each represented by a single value. In contrast, equation (20) is utilized to assess the cost value for the other two properties including radial distribution function and density profile, represented by a distribution. Furthermore, it is important to highlight that the weighting coefficient for each property in the first category is set at 0.15, while for the second category, the assigned weighting coefficient for each property is 0.125. Figure 6 depicts the flow of the genetic algorithm.

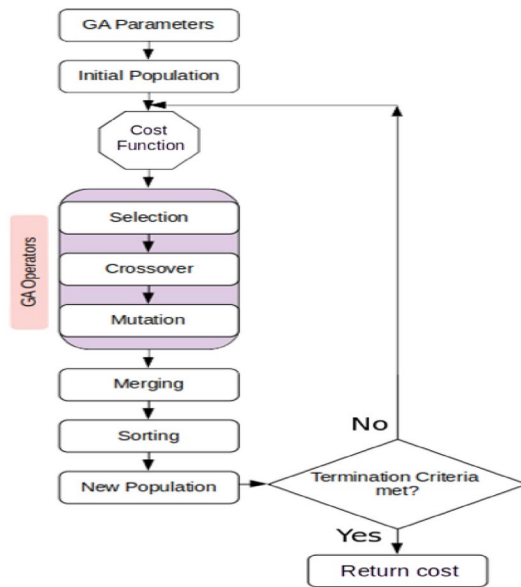


Figure 6. Flowchart of the genetic algorithm.

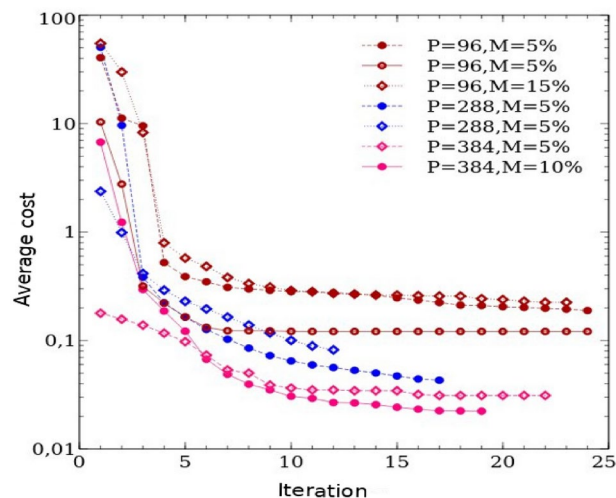


Figure 7. Average cost value as a function of iteration. The legend distinguishes between two datasets labeled with the same symbol; these datasets correspond to results from independent genetic algorithm runs, each initialized with different random seeds to ensure exploration of diverse regions in the parameter space. Symbols P and M in the legend denote the initial population size and mutation rate, respectively.

Results and discussion

Convergence in genetic algorithms (GA) hinges on various factors, with the initial population size and the number of iterations standing out as two crucial elements. As discussed earlier, the convergence of a genetic algorithm (GA) is significantly influenced by the size of the initial population, as it directly impacts the diversity and exploration capabilities of the algorithm, leading to diverse cost values. Therefore, a larger initial population size tends to facilitate a more extensive exploration of the search space, potentially yielding enhanced genomes (solutions) and cost values that are closer to the desired target. Hence, choosing an appropriate population size is crucial for attaining favorable convergence outcomes in a genetic algorithm (GA). Therefore, in this study, we carried out multiple evolutionary runs, adjusting population sizes from 96 to 384 and mutation rates from 5 to 15%. This exploration aimed to determine whether the evolution's convergence resulted in suboptimal (local solution) or optimal (global solution) outcomes. Figure 7 depicts the average cost versus iteration, demonstrating the evolution of average cost values across successive iterations for each generation. As expected, we observed that increasing the population size led to lower cost values, indicating fitter genomes. For example, when employing an initial population of 96 genomes, the genetic algorithm (GA) converged to a cost value of 0.121. In contrast, with a population size of 384, the GA attained a significantly lower cost value of 0.0242,

which is one order of magnitude smaller. This suggests that larger population sizes contribute to smaller cost values in the evolutionary process. Additionally, we recommend a baseline population size of 288 as conducive to achieving convergence in the genetic algorithm (GA). Our simulation results show that commencing with a random population size of 288 enables the genetic algorithm (GA) to meet the desired convergence criteria. Nevertheless, it is crucial to acknowledge that the optimal population size may vary depending on the unique attributes of the problem domain and the characteristics of the search space. Hence, conducting additional experiments and fine-tuning is essential to ascertain the most suitable population size for the genetic algorithm (GA) in a specific problem context. Lastly, Figure 8 displays six different genomes-force field parameter sets and their corresponding membrane properties identified by the GA. Each set is represented by a distinct color and is compared against predefined GA targets. It is important to note that all datasets represent results from independent genetic algorithm runs, each with a distinct random initial population, to ensure thorough exploration of various regions within the parameter space.

The outcomes of our evolutionary runs reveal that the membrane properties of an optimal genome (solution), identified through the implementation of GA, optimally align with a parameterized force field parameter set including $\epsilon=5.788 k_B T$, $r_{min}=1.095$ nm, $r_1=1.729$ nm, $r_c=2.500$ nm, $k_{bond}=1997$ kJ.mol⁻¹nm², and $k_{angle}=264$ kJ.mol⁻¹rad². Interestingly, a wide range of parameter sets can yield fitness scores that are remarkably close, suggesting that the solution space within these simple models lies on a hyper-spherical plane, with diverse combination of parameters yielding equivalent results. It is crucial to highlight that the parameterization of force-fields often results in an underdetermined system because many of the structural target characteristic for membranes, such as thickness, area compressibility, line tension, and bending modulus, are interdependent and correlated. This interconnectedness is why coarse-grained models, like the Martini model, can effectively reproduce the elastic properties of lipid membranes without necessitating detailed parameterization of these properties^{58,114}. It is also worth pointing out that while the optimization strategy we employed allows us to achieve high accuracy for the target properties, we recognize the potential risk of overfitting these specific properties. To mitigate this, it is important to evaluate the robustness of the optimized parameter sets by predicting additional structural and perhaps even dynamic properties (to tune the effective times-scales of the model) that were not included in the current target set of the genetic algorithm.

Next, we underscore the significance of the three-site lipid membrane model by examining key processes that demonstrate its ability to accurately mimic the behavior of real lipid membranes¹¹⁵. These processes encompass the self-assembly of lipids into bilayers, the formation of vesicles through self-assembly, membrane fusion, and the creation of vesicles in the gel phase. These findings highlight the model's utility in biophysical studies, providing insights into the complex dynamics of lipid membranes

A) Self-assembly into bilayers: The fundamental process of lipids self-assembling into bilayers plays a crucial role in the formation of cellular membranes¹¹⁶. Lipids, characterized by hydrophilic head groups and hydrophobic tails, spontaneously organize into a bilayer structure, as illustrated in Figure 9A.

B) Self-assembly into vesicles: Under specific conditions, lipid membranes can undergo self-assembly, forming spherical vesicular structures. In essence, the self-assembly process results in the spontaneous creation of closed vesicles, featuring a lipid bilayer shell that effectively isolates the internal contents from the external environment. Vesicles hold considerable implications in drug delivery systems, as they possess the ability to encapsulate both hydrophilic and lipophilic drugs. This protective enclosure during transportation not only ensures the integrity of the drugs but also facilitates targeted delivery¹¹⁷⁻¹¹⁹ (refer to Figure 9B).

	A	B	C	D	E	F
ϵ	5.788	5.632	8.995	6.067	5.452	16.106
r_{min}	1.095	1.222	1.359	1.042	1.095	1.421
r_1	1.729	2.383	2.258	1.794	1.973	2.028
r_{cut}	2.500	2.623	2.620	2.246	2.289	2.500
k_{bond}	1997	12279	12793	10572	11476	6657
k_{angle}	264	10711	9313	13635	11300	5577
APL	0.677	0.72	0.66	0.60	0.69	0.63
AC	249	327	302	312	235	200
LT	8.90	13.18	11.17	9.89	12.33	8.47
BM	9.25	9.25	10.00	9.34	7.67	12.66
PTT	250	250	270	270	250	290

A)

B)

APL=0.68
AC=230
LT=10
BM=12
PTT=270

Figure 8. (A) Illustration of force field parameters as well as the corresponding membrane properties for 6 different force field parameter sets, each uniquely highlighted by a distinct color. In all depicted 6 sets of force field parameters, the obtained normalized density profile and radial distribution function exhibited complete similarity, i.e., EMD=0, with the reference atomistic density profile and radial distribution function. (B) Predefined GA point targets, i.e., area per lipid (APL= 0.68 nm²), area compressibility (AC= 230 mN/m), line tension (LT= 10pN), bending modulus (BM= 12 k_BT) and phase transition temperature (PTT= 270K). Alongside the GA point targets, we as well accept genomes that capture membrane properties consistent with experimental data for PC membranes.

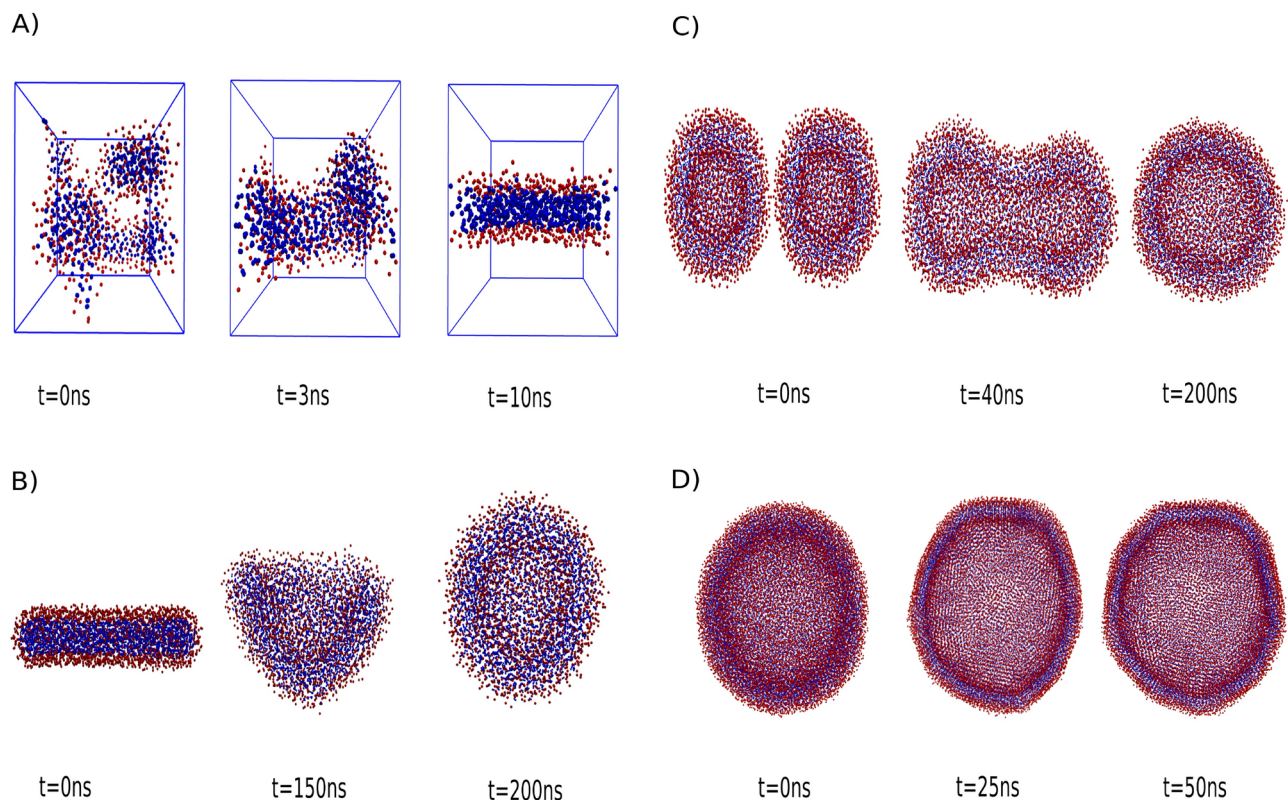


Figure 9. Realistic behavior of lipid membranes. (A) Self assembly into bilayer. (B) Self assembly into vesicle. (C) Membrane fusion. (D) Formation of vesicle in the gel phase.

(C) Membrane fusion: The dynamic process of membrane fusion occurs when two vesicles make contact and merge, ultimately giving rise to the formation of a continuous vesicle¹²⁰. This process holds paramount significance in various biological phenomena, encompassing cellular trafficking, exocytosis, and the fusion of enveloped viruses with host cell membranes^{121,122}. Membrane fusion entails a sequence of intricate molecular rearrangements and fusion events that bring two vesicles into close proximity, destabilize their structures, and ultimately result in their merging, as depicted in Figure 9C.

(D) Formation of vesicles in the gel phase: The existence of lipid membranes in distinct phases is temperature and lipid composition-dependent¹²³. In the gel phase, lipids are already in an ordered and rigid state. However, because of the soft nature of the model, local rearrangements or disruptions within the bilayer can still occur, which allows for the formation of vesicles even in this phase. The occurrence of vesicle formation in the gel phase is attributed to the local disruption of the lipid bilayer. This disruption enables the bilayer to bend and close upon itself, leading to the formation of vesicles¹²⁴. This behavior holds significant implications, particularly in areas like cryopreservation, where the formation of vesicles in the gel phase serves as a protective mechanism for cellular structures against damage induced by freezing (refer to Figure 9D).

Multiscale simulation

As highlighted earlier, we believe it is crucial for the next generations of coarse-grained models to possess state-of-the-art capabilities that facilitate mixed-resolution multi-scale approaches. This ensures seamless integration of membrane proteins, allowing for heightened precision, even at the atomistic level. Therefore, in this study, we introduced an additional technique for conducting multiscale simulations by employing tabulated potentials. These potentials serve as a means to seamlessly integrate various resolution levels. To evaluate the effectiveness of our approach, we conducted a simulation in which a Martini KALP peptide is embedded in our parameterized coarse-grained (CG) bilayer membrane. The primary goal was to ascertain the stability of the peptide within the membrane. To achieve this, we initially established nonbonded interaction tables to delineate the interactions between various components within the system, as follows:

- (i) Interaction tables were generated for the CG membrane beads using a nonbonded cut-off of 2.5 nm, employing the potential forms discussed earlier.
- (ii) Interaction tables were generated for the Martini peptide beads using a nonbonded cut-off of 1.2 nm, utilizing a modified LJ potential.
- (iii) Interaction tables were generated between the Martini peptide and CG membrane beads, utilizing a nonbonded cut-off of 1.2 nm and implementing a modified LJ potential. Accordingly, the termination criteria for the GA are met when a Gaussian distribution of angles between the bilayer membrane normal and

a vector connecting the first bead to the last bead in the peptide is achieved, with a mean of 30° and a standard deviation of 10° . Therefore, considering all hydrophilic head groups and hydrophobic tail groups of both the peptide and membrane, the GA needs to parameterize eight nonbonded interactions (ϵ , r_{min} , r_1 , r_{cut} , k_{bond} and k_{angle} in LJ potential) to achieve its two objectives. In other words, each genome consists of eight parameters (See Figure 10).

The outcome of the evolutionary process, represented by the cost as a function of iteration, illustrates that the GA successfully achieved the two predetermined objectives within six iterations. Moreover, the angle distribution obtained aligns perfectly with the expected distribution. Furthermore, the simulation snapshots illustrate that despite the initial insertion of the peptide being perpendicular to the membrane, it ultimately adapted to a tilted orientation of 30° due to the parameterization of the genetic algorithm (See Figure 11).

Our findings demonstrate a promising synergy between transmembrane peptides and ultra-CG models, achieved through the introduction of only two simple interaction types. A distinguishing feature of our model is its utilization of real units and dimensions, which is crucial for achieving this compatibility. Additionally, the integration of the GROMACS software, which supports the implementation of the Martini model, significantly enhances the model's applicability.

In our study, we have merely begun to explore the vast potential of hybrid simulation models. We anticipate that such an approach can be straightforwardly extended to more complex membrane proteins, such as viral spike proteins or multi-pass membrane proteins, thereby expanding its utility in biophysical studies. It is crucial to highlight that, given the implicit modeling of solvents, protein-protein interactions outside of the membrane are not realistically represented within a hybrid setup using the standard Martini models. A quick hack to circumvent this issue is to render the protein-protein and protein-membrane interactions within the cytosolic part as being repulsive only. This simple solution could enable the study of scenarios such as the clearance of the membrane fusion site or the expansion of a formed fusion pore, especially when the apposed membranes are overcrowded with sterically repelling proteins that have realistic dimensions, shapes, and membrane behavior. For instance, one could examine the effect of varying protein density on the formation or expansion of a fusion pore, or in the case of enveloped viruses, study the additional release of RNA through a formed fusion pore by combining the model with implicit coarse-grained models for modeling RNA¹²⁵. Moreover, knowledge of protein-protein interaction site may be ad hoc modeled on a level of non-bonded interactions within Martini, similar to the concept of Martini Go models¹²⁶. Certainly, such a setup remains an ultra-CG setup, and as

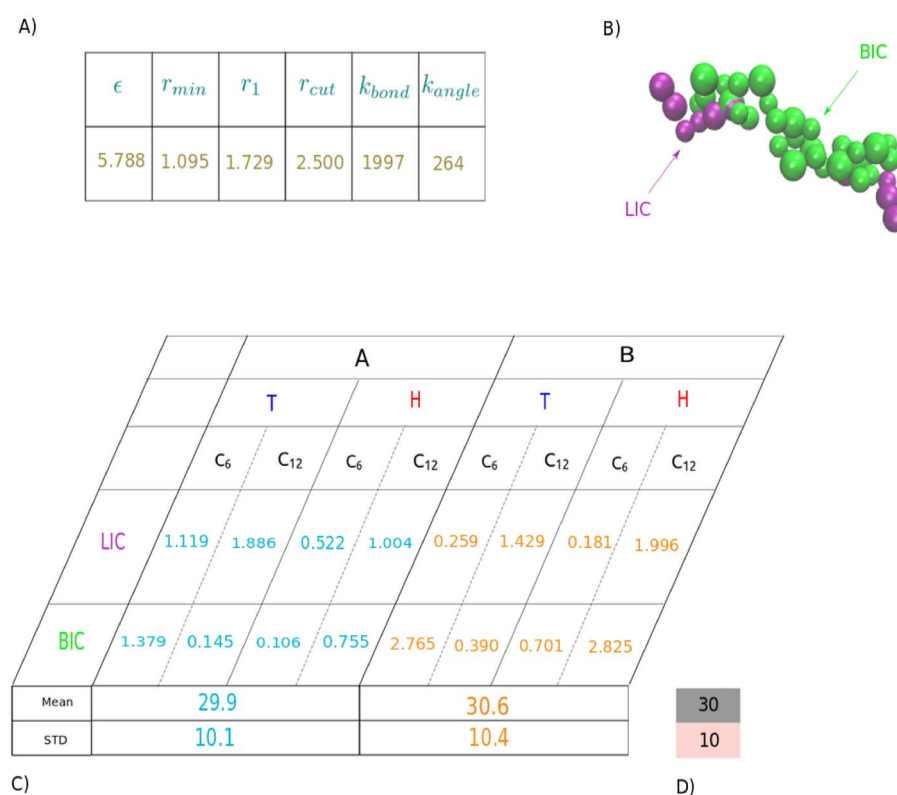


Figure 10. (A) Membrane force field parameters. (B) Illustration of Martini KALP Peptide with hydrophilic (LIC) regions represented in purple and hydrophobic (BIC) regions depicted in green. (C) Nonbonded membrane-peptide interactions – $C_6/\text{kJ mol}^{-1}\text{nm}^6$ and $C_{12}/\text{kJ mol}^{-1}\text{nm}^{12}$ – as well as the obtained properties from simulations corresponding to two various force field combinations. (D) Predefined GA targets, i.e., probability distribution of the angle defined between peptide and membrane normal with mean and standard deviation of 30° and 10° , respectively.

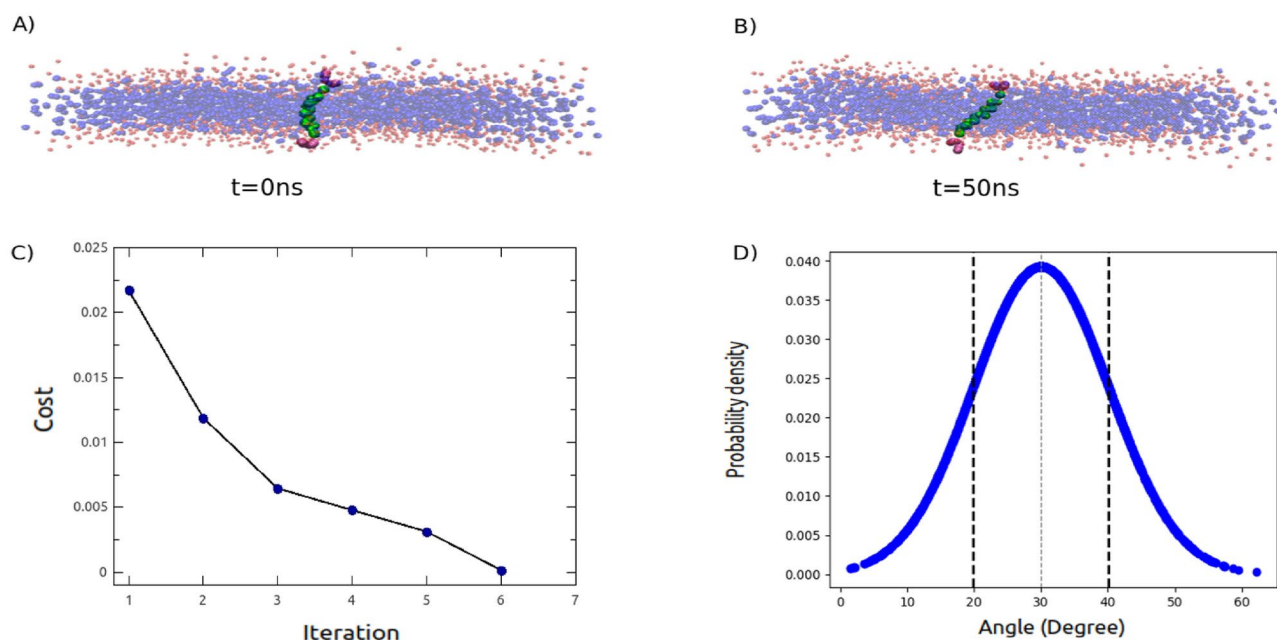


Figure 11. Peptide-Membrane Interaction. (A) Snapshot of a KALP peptide within a membrane at the initial of the simulation. (B) Simulation snapshot of a KALP peptide within a membrane at $t=50$ ns utilizing parameterized interactions via GA. (C) Cost versus iteration. (D) GA target, i.e. probability distribution of the angle defined between peptide and membrane normal with mean and standard deviation of 30° and 10° , respectively.

such, the study and system should be adapted to conform with the more limited accuracy of these models. Nonetheless, we anticipate the utility of these models in phenomenologically describing the behavior of large, complex protein-membrane systems with greater accuracy than simplified toy models of proteins.

Conclusion

In this work, evolutionary algorithms, specifically genetic algorithms, i.e., one of the unsupervised learning methods in machine learning, are employed to construct an implicit solvent ultra-coarse-grained POPC membrane model comprising three sites. Our goal is to parameterize the ultra-coarse-grained model to reproduce the experimentally observed structural and thermodynamic characteristics of PC membranes within real units and dimensions. This involves properties such as the area per lipid, area compressibility, bending modulus, line tension, phase transition temperature, density profile, and radial distribution function. To accomplish this, we conduct a series of evolutionary runs using diverse random initial population sizes, ranging from 96 to 384. As a result, our findings illustrate a direct correlation between increasing the initial population size and achieving lower cost values. Furthermore, we demonstrate that our ultra-coarse-grained membrane model exhibits realistic behaviors observed in lipid membranes. This includes self-assembly into bilayers, vesicle formation, membrane fusion, and gel phase formation. To enhance the accessibility of the ultra-coarse-grained membrane model for a broader audience, topology files of the derived potentials are supplied in a tabulated format. These files are meticulously designed to seamlessly integrate with the GROMACS molecular dynamics engine, facilitating immediate implementation and usage. Moreover, we highlight the compatibility of the developed ultra-coarse-grained model with the widely used Martini coarse-grained model. This compatibility enables the simulation of realistic membrane proteins within an ultra-coarse-grained bilayer membrane. This compatibility opens up avenues for researchers to explore complex membrane systems, delve into protein-membrane interactions, and gain valuable insights applicable to various biological processes and the development of therapeutics.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 12 July 2024; Accepted: 7 October 2024

Published online: 28 October 2024

References

1. Sadeghi, M., Weikl, T. R. & Noé, F. Particle-based membrane model for mesoscopic simulation of cellular dynamics. *J. Chem. Phys.* **148**, 044901 (2018).
2. Martinez-Seara, H. & Rog, T. *Biomolecular Simulations* 407–429 (Springer, 2013).

3. Moradi, S., Nowroozi, A. & Shahlai, M. Correction: Shedding light on the structural properties of lipid bilayers using molecular dynamics simulation: A review study. *RSC Adv.* **9**, 7687–7687 (2019).
4. Sadeghi, M. & Noé, F. Hydrodynamic coupling for particle-based solvent-free membrane models. *J. Chem. Phys.* **155**, 114108 (2021).
5. Lindahl, E. & Edholm, O. Mesoscopic undulations and thickness fluctuations in lipid bilayers from molecular dynamics simulations. *Biophys. J.* **79**, 426–433 (2000).
6. Marrink, S. & Mark, A. Effect of undulations on surface tension in simulated bilayers. *J. Phys. Chem. B* **105**, 6122–6127 (2001).
7. de Vries, A. H., Yefimov, S., Mark, A. E. & Marrink, S. J. Molecular structure of the lecithin ripple phase. *Proc. Natl. Acad. Sci.* **102**, 5392–5396 (2005).
8. Moradi, S., Nowroozi, A. & Shahlai, M. Shedding light on the structural properties of lipid bilayers using molecular dynamics simulation: A review study. *RSC Adv.* **9**, 4644–4658 (2019).
9. Gu, R.-X., Baoukina, S. & Tieleman, D. P. Phase separation in atomistic simulations of model membranes. *J. Am. Chem. Soc.* **142**, 2844–2856 (2020).
10. Goetz, R. & Lipowsky, R. Computer simulations of bilayer membranes: Self-assembly and interfacial tension. *J. Chem. Phys.* **108**, 7397–7409 (1998).
11. Drouffe, J.-M., Maggs, A. & Leibler, S. Computer simulations of self-assembled membranes. *Science* **254**, 1353–1356 (1991).
12. Noguchi, H. & Takasu, M. Self-assembly of amphiphiles into vesicles: A Brownian dynamics simulation. *Phys. Rev. E* **64**, 041913 (2001).
13. Wang, Z.-J. & Frenkel, D. Modeling flexible amphiphilic bilayers: A solvent-free off-lattice Monte Carlo study. *J. Chem. Phys.* **122**, 234711 (2005).
14. Farago, O. “Water-free” computer model for fluid bilayer membranes. *J. Chem. Phys.* **119**, 596–605 (2003).
15. Cooke, I. R., Kremer, K. & Deserno, M. Tunable generic model for fluid bilayer membranes. *Phys. Rev. E* **72**, 011506 (2005).
16. Cooke, I. R. & Deserno, M. Solvent-free model for self-assembling fluid bilayer membranes: Stabilization of the fluid phase based on broad attractive tail potentials. *J. Chem. Phys.* **123**, 224710 (2005).
17. Mori, T., Miyashita, N., Im, W., Feig, M. & Sugita, Y. Molecular dynamics simulations of biological membranes and membrane proteins using enhanced conformational sampling algorithms. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1858**, 1635–1651 (2016).
18. Arnarez, C. et al. Dry Martini, a coarse-grained force field for lipid membrane simulations with implicit solvent. *J. Chem. Theory Comput.* **11**, 260–275 (2015).
19. Sharma, P., Desikan, R. & Ayappa, K. G. Evaluating coarse-grained MARTINI force-fields for capturing the ripple phase of lipid membranes. *J. Phys. Chem. B* **125**, 6587–6599 (2021).
20. Liu, M., Gan, J., Gao, L., & Wang, W. Molecular Dynamics Simulation of Self-assembly and Electroporation of Lipid Bilayer Membrane in Martini Force Field. *2019 IEEE 19th International Conference on Nanotechnology (IEEE-NANO)*. pp 68–71 (2019).
21. Arnarez, C. et al. Dry Martini, a coarse-grained force field for lipid membrane simulations with implicit solvent. *J. Chem. Theory Comput.* **11**, 260–275 (2015).
22. Yuan, H., Huang, C., Li, J., Lykotraftis, G. & Zhang, S. One-particle-thick, solvent-free, coarse-grained model for biological and biomimetic fluid membranes. *Phys. Rev. E* **82**, 011905 (2010).
23. Fu, S.-P., Peng, Z., Yuan, H., Kfoury, R. & Young, Y.-N. Lennard-Jones type pair-potential method for coarse-grained lipid bilayer membrane simulations in LAMMPS. *Comput. Phys. Commun.* **210**, 193–203 (2017).
24. Feng, S., Hu, Y. & Liang, H. Entropic elasticity based coarse-grained model of lipid membranes. *J. Chem. Phys.* **148**, 164705 (2018).
25. Allhusen, J. S. & Conboy, J. C. The ins and outs of lipid flip-flop. *Acc. Chem. Res.* **50**, 58–65 (2017).
26. Awasthi, N. & Hub, J. S. Simulations of pore formation in lipid membranes: reaction coordinates, convergence, hysteresis, and finite-size effects. *J. Chem. Theory Comput.* **12**, 3261–3269 (2016).
27. Tieleman, D. P. & Marrink, S.-J. Lipids out of equilibrium: Energetics of desorption and pore mediated flip-flop. *J. Am. Chem. Soc.* **128**, 12462–12467 (2006).
28. Bubnis, G., Risselada, H. J. & Grubmüller, H. Exploiting lipid permutation symmetry to compute membrane remodeling free energies. *Phys. Rev. Lett.* **117**, 188102 (2016).
29. Smirnova, Y. G., Marrink, S.-J., Lipowsky, R. & Knecht, V. Solvent-exposed tails as prestalk transition states for membrane fusion at low hydration. *J. Am. Chem. Soc.* **132**, 6710–6718 (2010).
30. Mattila, J.-P. et al. A hemi-fission intermediate links two mechanistically distinct stages of membrane fission. *Nature* **524**, 109–113 (2015).
31. Zoni, V. et al. To bud or not to bud: a perspective on molecular simulations of lipid droplet budding. *Front. Mol. Biosci.* **6**, 124 (2019).
32. Srivastava, A. & Voth, G. A. Hybrid approach for highly coarse-grained lipid bilayer models. *J. Chem. Theory Comput.* **9**, 750–765 (2013).
33. Cooke, I. R., Kremer, K. & Deserno, M. Tunable generic model for fluid bilayer membranes. *Phys. Rev. E* **72**, 011506 (2005).
34. Brannigan, G., Philips, P. F. & Brown, F. L. H. Flexible lipid bilayers in implicit solvent. *Phys. Rev. E* **72**, 011915 (2005).
35. Xu, R., Wang, Z.-L. & He, X.-H. Mesoscale simulation of vesiculation of lipid droplets. *Chin. J. Chem. Phys.* **27**, 663 (2015).
36. Izvekov, S., Parrinello, M., Burnham, C. J. & Voth, G. A. Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching. *J. Chem. Phys.* **120**, 10896–10913 (2004).
37. Izvekov, S. & Voth, G. A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **109**, 2469–2473 (2005).
38. Lyubartsev, A. P. & Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E* **52**, 3730 (1995).
39. Lyubartsev, A. P. Multiscale modeling of lipids and lipid bilayers. *Eur. Biophys. J.* **35**, 53–61 (2005).
40. Wang, Z.-J. & Deserno, M. A systematically coarse-grained solvent-free model for quantitative phospholipid bilayer simulations. *J. Phys. Chem. B* **114**, 11207–11220 (2010).
41. Ingólfsson, H. I. et al. The power of coarse graining in biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4**, 225–248 (2014).
42. Brini, E. et al. Systematic coarse-graining methods for soft matter simulations—a review. *Soft Matter* **9**, 2108–2119 (2013).
43. Van Der Spoel, D. & van Maaren, P. J. The origin of layer structure artifacts in simulations of liquid water. *J. Chem. Theory Comput.* **2**, 1–11 (2006).
44. Abraham, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).
45. Patra, M. et al. Molecular dynamics simulations of lipid bilayers: Major artifacts due to truncating electrostatic interactions. *Biophys. J.* **84**, 3636–3645 (2003).
46. Frigini, E. N., Cascales, J. L. & Porasso, R. D. Molecular dynamics simulations of glyphosate in a DPPC lipid bilayer. *Chem. Phys. Lipid.* **213**, 111–117 (2018).
47. Shahlai, M., Madadkar-Sobhani, A., Fassihi, A. & Saghale, L. Exploring a model of a chemokine receptor/ligand complex in an explicit membrane environment by molecular dynamics simulation: the human CCR1 receptor. *J. Chem. Inf. Model.* **51**, 2717–2730 (2011).

48. Shahlai, M. et al. Homology modeling of human CCR5 and analysis of its binding properties through molecular docking and molecular dynamics simulation. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1808**, 802–817 (2011).
49. Shahlai, M. & Mousavi, A. A conformational analysis study on the melanocortin 4 receptor using multiple molecular dynamics simulations. *Chem. Biol. Drug Des.* **86**, 309–321 (2015).
50. Kinnun, J. J., Mallikarjunaiah, K., Petrache, H. I. & Brown, M. F. Elastic deformation and area per lipid of membranes: atomistic view from solid-state deuterium NMR spectroscopy. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1848**, 246–259 (2015).
51. Brown, M. F. Curvature forces in membrane lipid-protein interactions. *Biochemistry* **51**, 9782–9795 (2012).
52. Botelho, A. V., Huber, T., Sakmar, T. P. & Brown, M. F. Curvature and hydrophobic forces drive oligomerization and modulate activity of rhodopsin in membranes. *Biophys. J.* **91**, 4464–4477 (2006).
53. Sodt, A. J. & Head-Gordon, T. An implicit solvent coarse-grained lipid model with correct stress profile. *J. Chem. Phys.* **132**, 05B611 (2010).
54. Kučerka, N., Tristram-Nagle, S. & Nagle, J. F. Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains. *J. Membr. Biol.* **208**, 193–202 (2006).
55. Grote, F. & Lyubartsev, A. P. Optimization of slippers force field parameters describing headgroups of phospholipids. *J. Phys. Chem. B* **124**, 8784–8793 (2020).
56. Zhang, P. et al. Molecular dynamics simulation of transport mechanism of graphene quantum dots through different cell membranes. *Membranes* **12**, 753 (2022).
57. Angladon, M.-A., Fossépré, M., Leherter, L. & Vercauteren, D. P. Interaction of POPC, DPPC, and POPE with the μ opioid receptor: A coarse-grained molecular dynamics study. *PLoS ONE* **14**, e0213646 (2019).
58. Marrink, S. J., De Vries, A. H. & Mark, A. E. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B* **108**, 750–760 (2004).
59. Guo, Y., Baulin, V. A. & Thalmann, F. Peroxidised phospholipid bilayers: Insight from coarse-grained molecular dynamics simulations. *Soft Matter* **12**, 263–271 (2016).
60. Porasso, R. D. & López Cascales, J. J. A criterion to identify the equilibration time in lipid bilayer simulations. *Pap. Phys.* **4**, 1–9 (2012).
61. Nagle, J. F. et al. X-ray structure determination of fully hydrated L alpha phase dipalmitoylphosphatidylcholine bilayers. *Biophys. J.* **70**, 1419–1431 (1996).
62. Rand, R. & Parsegian, V. Hydration forces between phospholipid bilayers. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* **988**, 351–376 (1989).
63. Nagle, J. F. & Tristram-Nagle, S. Structure of lipid bilayers. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* **1469**, 159–195 (2000).
64. Rawicz, W., Olbrich, K. C., McIntosh, T., Needham, D. & Evans, E. Effect of chain length and unsaturation on elasticity of lipid bilayers. *Biophys. J.* **79**, 328–339 (2000).
65. Binder, H. & Gawrisch, K. Effect of unsaturated lipid chains on dimensions, molecular order and hydration of membranes. *J. Phys. Chem. B* **105**, 12378–12390 (2001).
66. Wang, Y., Gkeka, P., Fuchs, J. E., Liedl, K. R. & Cournia, Z. DPPC-cholesterol phase diagram using coarse-grained Molecular Dynamics simulations. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1858**, 2846–2857 (2016).
67. Saeedimazine, M., Montanino, A., Kleiven, S. & Villa, A. Role of lipid composition on the structural and mechanical features of axonal membranes: a molecular simulation study. *Sci. Rep.* **9**, 1–12 (2019).
68. Shigematsu, T., Koshiyama, K. & Wada, S. Line tension of the pore edge in phospholipid/cholesterol bilayer from stretch molecular dynamics simulation. *J. Biomech. Sci. Eng.* **15**, 00422 (2015).
69. García-Sáez, A. J., Chiantia, S., Salgado, J. & Schwille, P. Pore formation by a Bax-derived peptide: Effect on the line tension of the membrane probed by AFM. *Biophys. J.* **93**, 103–112 (2007).
70. May, S. A molecular model for the line tension of lipid membranes. *Eur. Phys. J. E* **3**, 37–44 (2000).
71. Melikyan, G., Matinyan, N. & Arakelian, V. The influence of gangliosides on the hydrophilic pore edge line tension and monolayer fusion of lipid membranes. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1030**, 11–15 (1990).
72. Portet, T. et al. Visualization of membrane loss during the shrinkage of giant vesicles under electropulsation. *Biophys. J.* **96**, 4109–4121 (2009).
73. Wang, Z.-J. & Deserno, M. Systematic implicit solvent coarse-graining of bilayer membranes: Lipid and phase transferability of the force field. *New J. Phys.* **12**, 095004 (2010).
74. Wohler, J., den Otter, W. K., Edholm, O. & Briels, W. J. Free energy of a trans-membrane pore calculated from atomistic molecular dynamics simulations. *J. Chem. Phys.* **124**, 154905 (2006).
75. Miyazaki, Y., Okazaki, S. & Shinoda, W. Free energy analysis of membrane pore formation process in the presence of multiple melittin peptides. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1861**, 1409–1419 (2019).
76. Jiang, F. Y., Bouret, Y. & Kindt, J. T. Molecular dynamics simulations of the lipid bilayer edge. *Biophys. J.* **87**, 182–192 (2004).
77. Akimov, S. A. et al. Pore formation in lipid membrane I: Continuous reversible trajectory from intact bilayer through hydrophobic defect to transversal pore. *Sci. Rep.* **7**, 12152 (2017).
78. Brannigan, G. & Brown, F. L. Solvent-free simulations of fluid membrane bilayers. *J. Chem. Phys.* **120**, 1059–1071 (2004).
79. Brandt, E. G., Braun, A. R., Sachs, J. N., Nagle, J. F. & Edholm, O. Interpretation of fluctuation spectra in lipid bilayer simulations. *Biophys. J.* **100**, 2104–2111 (2011).
80. Levashov, V. A., Egami, T., Aga, R. S. & Morris, J. R. Equipartition theorem and the dynamics of liquids. *Phys. Rev. B* **78**, 064205 (2008).
81. Sadeghi, M. & Noé, F. Large-scale simulation of biomembranes incorporating realistic kinetics into coarse-grained models. *Nat. Commun.* **11**, 1–13 (2020).
82. Chacón, E., Tarazona, P. & Bresme, F. A computer simulation approach to quantify the true area and true area compressibility modulus of biological membranes. *J. Chem. Phys.* **143**, 07B610_1 (2015).
83. Niggemann, G., Kummrow, M. & Helfrich, W. The bending rigidity of phosphatidylcholine bilayers: Dependences on experimental method, sample cell sealing and temperature. *Journal de Physique II*(5), 413–425 (1995).
84. Leekumjorn, S. & Sum, A. K. Molecular studies of the gel to liquid-crystalline phase transition for fully hydrated DPPC and DPPE bilayers. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1768**, 354–365 (2007).
85. Suurkuusk, J., Lentz, B., Barenholz, Y., Biltonen, R. & Thompson, T. A calorimetric and fluorescent probe study of the gel-liquid crystalline phase transition in small, single-lamellar dipalmitoylphosphatidylcholine vesicles. *Biochemistry* **15**, 1393–1401 (1976).
86. Davis, J. H. Deuterium magnetic resonance study of the gel and liquid crystalline phases of dipalmitoyl phosphatidylcholine. *Biophys. J.* **27**, 339–358 (1979).
87. Sun, L. & Böckmann, R. A. Membrane phase transition during heating and cooling: molecular insight into reversible melting. *Eur. Biophys. J.* **47**, 151–164 (2018).
88. Stevens, M. J. Coarse-grained simulations of lipid bilayers. *J. Chem. Phys.* **121**, 11942–11948 (2004).
89. Marrink, S. J., Risselada, J. & Mark, A. E. Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model. *Chem. Phys. Lipid.* **135**, 223–244 (2005).
90. Kociurzynski, E. A. Phase transition of glycolipid membranes studied by coarse-grained simulations. *Langmuir* **31**, 9379–9387 (2015).

91. Mirzoev, A. & Lyubartsev, A. P. Systematic implicit solvent coarse graining of dimyristoylphosphatidylcholine lipids. *J. Comput. Chem.* **35**, 1208–1218 (2014).
92. Johansson, E. A. Phase equilibria and formation of vesicles of dioleoylphosphatidylcholine in glycerol/water mixtures. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1149**, 285–291 (1993).
93. Litman, B. J., Lewis, E. N. & Levin, I. W. Packing characteristics of highly unsaturated bilayer lipids: Raman spectroscopic studies of multilamellar phosphatidylcholine dispersions. *Biochemistry* **30**, 313–319 (1991).
94. Mu, X., Wang, D., Feng, T. & Kübel, C. Radial distribution function imaging by STEM diffraction: Phase mapping and analysis of heterogeneous nanostructured glasses. *Ultramicroscopy* **168**, 1–6 (2016).
95. Hub, J. S., Salditt, T., Rheinstädter, M. C. & De Groot, B. L. Short-range order and collective dynamics of DMPC bilayers: A comparison between molecular dynamics simulations, x-ray, and neutron scattering experiments. *Biophys. J.* **93**, 3156–3168 (2007).
96. Rubner, Y., Tomasi, C. & Guibas, L. J. The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vision* **40**, 99–121 (2000).
97. Hamed, F. & Mohammad-Aghaie, D. Molecular insight into the mutual interactions of two transmembrane domains of human glycine receptor (TM23-GlyR), with the Lipid Bilayers. *Physical Chemistry Research* **8**, 373–397 (2020).
98. Angibaud, L., Briquet, L., Philipp, P., Wirtz, T. & Kieffer, J. Parameter optimization in molecular dynamics simulations using a genetic algorithm. *Nucl. Instrum. Methods Phys. Res., Sect. B* **269**, 1559–1563 (2011).
99. Becue, A., Meurice, N., Leherte, L. & Vercauteren, D. P. *Models, Mysteries and Magic of Molecules* 301–323 (Springer, 2008).
100. Wang, J. & Kollman, P. A. Automatic parameterization of force field by systematic search and genetic algorithms. *J. Comput. Chem.* **22**, 1219–1228 (2001).
101. Sakae, Y., Straub, J. E. & Okamoto, Y. Enhanced sampling method in molecular simulations using genetic algorithm for biomolecular systems. *J. Comput. Chem.* **40**, 475–481 (2019).
102. Zames, G. et al. Genetic algorithms in search, optimization and machine learning. *Inf. Technol. J.* **3**, 301–302 (1981).
103. Behera, N. Principles and Methods for Data Science. In *Handbook of Statistics* Vol. 43 (eds Srinivasa, Rao AS. & Rao, C.) 349–378 (Elsevier, 2020).
104. Koohestani, B. A crossover operator for improving the efficiency of permutation-based genetic algorithms. *Expert Syst. Appl.* **151**, 113381 (2020).
105. Kaya, Y., & Uyar, M. others A novel crossover operator for genetic algorithms: ring crossover. arXiv preprint [arXiv:1105.0355](https://arxiv.org/abs/1105.0355) (2011)
106. Abdoun, O., Abouchabaka, J., & Tajani, C. Analyzing the performance of mutation operators to solve the travelling salesman problem. arXiv preprint [arXiv:1203.3099](https://arxiv.org/abs/1203.3099) (2012)
107. Srinivas, M. & Patnaik, L. M. Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE Trans. Syst. Man Cybern.* **24**, 656–667 (1994).
108. Deb, K. *Multi-objective evolutionary optimisation for product design and manufacturing* 3–34 (Springer, 2011).
109. Rocke, D. Genetic Algorithms+ Data Structures= Evolution programs (3rd. *Journal of the American Statistical Association* **95**, 347 (2000).
110. Deb, K., & Agrawal, S. A niched-penalty approach for constraint handling in genetic algorithms. *Artificial neural nets and genetic algorithms*. pp 235–243 (1999)
111. Bell, O. Applications of Gaussian Mutation for Self Adaptation in Evolutionary Genetic Algorithms. arXiv preprint [arXiv:2201.00285](https://arxiv.org/abs/2201.00285) (2022)
112. Safe, M., Carballido, J., Ponzoni, I., & Brignole, N. *On Stopping Criteria for Genetic Algorithms. Advances in Artificial Intelligence – SBIA 2004*. Berlin, Heidelberg, pp 405–413 (2004).
113. Kim, I. Y., & de Weck, O. Adaptive weighted sum method for multiobjective optimization. *10th AIAA/ISSMO multidisciplinary analysis and optimization conference*. p 4322 (2004).
114. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & De Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).
115. Peetla, C., Stine, A. & Labhasetwar, V. Biophysical interactions with model lipid membranes: Applications in drug discovery and drug delivery. *Mol. Pharm.* **6**, 1264–1276 (2009).
116. Sych, T., Mély, Y. & Römer, W. Lipid self-assembly and lectin-induced reorganization of the plasma membrane. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170117 (2018).
117. Xu, R. et al. Extracellular vesicles in cancer-implications for future improvements in cancer care. *Nat. Rev. Clin. Oncol.* **15**, 617–638 (2018).
118. Burdick, D., Le Gall, A. H. & Rodriguez-Boulan, E. Vesicular transport: implications for cell polarity. *Biocell: official journal of the Sociedades Latinoamericanas de Microscopia Electronica... et. al* **20**, 343–353 (1996).
119. Song, H. et al. Nanoengineering facilitating the target mission: targeted extracellular vesicles delivery systems design. *J. Nanobiotechnol.* **20**, 1–23 (2022).
120. Jahn, R., Lang, T. & Südhof, T. C. Membrane fusion. *Cell* **112**, 519–533 (2003).
121. Rothman, J. E. Intracellular membrane fusion. *Adv. Second Messenger Phosphoprotein Res.* **29**, 81–96 (1994).
122. Joardar, A., Pattnaik, G. P. & Chakraborty, H. Mechanism of membrane fusion: interplay of lipid and peptide. *J. Membr. Biol.* **255**, 211–224 (2022).
123. Heberle, F. A. & Feigenson, G. W. Phase separation in lipid membranes. *Cold Spring Harb. Perspect. Biol.* **3**, a004630 (2011).
124. Talló, K., Pons, R., González, C. & López, O. Monitoring the formation of a colloidal lipid gel at the nanoscale: Vesicle aggregation driven by a temperature-induced mechanism. *J. Mater. Chem. B* **9**, 7472–7481 (2021).
125. Dawson, W. K., Maciejczyk, M., Jankowska, E. J. & Bujnicki, J. M. Coarse-grained modeling of RNA 3D structure. *Methods* **103**, 138–156 (2016).
126. Poma, A. B., Cieplak, M. & Theodorakis, P. E. Combining the MARTINI and structure-based coarse-grained approaches for the molecular dynamics studies of conformational transitions in proteins. *J. Chem. Theory Comput.* **13**, 1366–1374 (2017).

Acknowledgements

The authors would like to thank Marcus Müller for fruitful discussions. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2033 - 390677874 - RESOLV.

Author contributions

A.S. conducted molecular dynamics simulations and calculations, and drafted the initial manuscript. J.R. conceptualized the research. The two authors participated in result discussions and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.J.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024