Force field development for performing coarse-grained molecular dynamics simulations of biological membranes

Kyoto University, Graduate School of Science

Division of Biological Science, Department of Biophysics

Ugarte La Torre Diego Renato

2021, July

Abstract

Chapter 1 – Introduction

Although molecular dynamics (MD) simulations have proven to be invaluable tools for studying the structural dynamics of biomolecular systems, the time scale achievable by all-atom MD simulations still presents a significant challenge for studying several biological phenomena. To address this constraint, coarse-grained (CG) modeling lowers the number of degrees of freedom by grouping atoms into CG beads, decreasing the computational cost of simulations while preserving as much as possible of the properties of interest. Since biomolecular structures are hierarchical, there are a variety of coarse-graining resolutions. Higher-resolution CG models are generally more reliable, but they are often more computationally expensive. Thus, depending on the target system, some models are more suitable than others. For example, when studying large systems, such as in biological membranes, implicit solvent models decrease enormously computational calculations by integrating the effect of the solvent into the interaction potentials.

This thesis presents the development of a new implicit solvent lipid model for performing coarse-grained MD simulations of biological membrane systems and is organized in the following way. In chapter 2, a CG implicit solvent lipid model is introduced. Then, in chapter 3, a way to combine the coarse-grained lipid model with α-carbon protein models is explained, followed by its application to various membrane proteins. Finally, in chapter 4, some concluding remarks are given.

Chapter 2 – Coarse-grained implicit solvent lipid force field with a compatible resolution to the Cα protein representation

MD simulations of biological membrane systems often need to consider both the lipids involved in the membrane and the proteins around it. This requires contemplating lipid and protein force fields not only as independent units but also as a unified model. However, combining different CG models is not a trivial task, and it requires a certain degree of compatibility between them. More specifically, it is desired that the different models to-be-combined share a similar resolution, i.e., a similar mapping between all-atom particles to CG beads. In this chapter, a lipid model with a similar resolution to the widely-used Cα representation of proteins, iSoLF, is presented.

In iSoLF, two-tailed lipid molecules are mapped to single-chain molecules composed of five beads: two hydrophilic head beads (H1 and H2) and three hydrophobic tail beads (T1, T2, and T3). This mapping produces a similar resolution to the one in the Cα representation of proteins, which is a feature that is desired. The interaction potential between different CG lipid molecules is composed of four terms: V_{Bond}, V_{Angle}, $V_{Repulsion}$, and $V_{Attraction}$. The two terms, V_{Bond} and V_{Angle}, represent the local bond and bond-angle interactions of lipid beads and are parameterized by inverting all-atom-sampled distributions using the Boltzmann Inversion method. In contrast, the other two terms, $V_{Repulsion}$, and $V_{Attraction}$, represent the non-local interactions and were modeled using the same interacting potential described previously by Cooke et al. [Phys. Rev. E (2005) 72]. The parameters for these non-local interactions were tuned for reproducing the area-per-lipid and the hydrophobic thickness of lipid bilayers made of POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) or DPPC (1,2-dipalmitoyl-sn-glycero-3-phosphatidylcholine).

After obtaining an optimal set of parameters, different features of the model were tested. The first test consisted of the spontaneous formation of a membrane. In implicit solvent models, as is the case in iSoLF, the hydrophobic effect is encoded in the non-local interactions, and it is crucial that a bilayer conformation is favored. By starting from a random configuration of lipids, the simulations showed that the lipid model correctly assembled a lipid bilayer, as desired. The second test consisted of obtaining the right phase behavior of each lipid at 30 °C. The simulations showed that POPC lipids were in a liquid phase, whereas for DPPC, lipids adopted a gel phase. Contrasting these results with experimental measurements for these lipids, the model produced the expected behavior. Finally, the last test consisted of a simulation of a vesicle. Since vesicles behave locally as lipid bilayers, it was expected that the model could also stabilize these conformations. By making the local curvature of vesicles large enough, they could be simulated without any breakage. Altogether, this showed that iSoLF is a suitable model for performing CG MD simulations of biological membranes.

Chapter 3 – Modeling lipid-protein interactions for coarse-grained lipid and Cα protein models

Lipid-protein interactions are key components in the simulation of membrane proteins because they encompass each molecule's hydrophobic and hydrophilic nature, especially in implicit solvent systems. Therefore, the modeling of these interactions needs to be examined carefully. In this chapter, the modeling and parameterization of a Lennard-Jones-like energy function representing hydrophilic and hydrophobic interactions between lipids and proteins are presented.

One advantage of the Lennard-Jones potential is its separability into short- and long-range interactions. This idea was used previously by Kim and Hummer [J. Mol. Bio. (2008)

375] in the context of protein-protein interaction. Here, a similar approach was followed by defining the lipid-proteins interactions as $V_{LPI} = V_{rep} + V_{HP}$, where V_{rep} corresponds to the Weeks-Chandler-Andersen potential and V_{HP} to the hydrophilic-hydrophobic interactions.

In order to reduce the complexity of the parameterization of the V_{LPI}, some simplifications were introduced. The coefficients of the repulsive term, V_{rep}, were calculated using combination rules with the corresponding parameters from iSoLF (lipid model) and AICG2+ (protein model). On the contrary, the hydrophilic-hydrophobic interactions, V_{HP}, were tuned against an experimental hydrophobicity scale for the twenty amino acids and free energy profiles for the insertion of amino acids inside a lipid bilayer calculated in all-atom simulations.

The parameterized interaction was tested by performing simulations of various proteins with different positionings inside lipid-containing systems. The first group of proteins consisted of transmembrane proteins with either α-helices or β-sheets spanning the membrane. The calculated tilt angle (orientation with respect to the normal to the lipid bilayer) and insertion depth (height relative to the center of mass of the lipid bilayer) for these proteins had good agreement with the reference values obtained from the OPM database. Importantly, during these tests, the need for special coefficients for the N and C termini of a protein was put into evidence because for single α-helix peptides, charged residues at both ends help stabilize their orientation. The second group of proteins consisted of water-soluble globular proteins. The expected behavior was observed for these proteins, i.e., they did not interact with the lipid bilayer. Lastly, the final group consisted of peripheral and other proteins. The binding of peripheral proteins into the membrane surface was observed, as expected. On the other hand, crambin, an initially classified peripheral protein, was inserted into the membrane, obtaining a stable configuration in the middle of the bilayer. This insertion agreed with previous reports in the literature. In general, the predicted behavior for various groups of proteins within/outside lipid environments could be reproduced by the lipid-protein interaction.

Chapter 4 – Conclusions

In this work, a novel force field for performing CG MD simulations of biological membranes is presented. In the first chapter, the current status of the field was examined. In chapter two, the details about the development of iSoLF, a CG lipid model for phospholipids, were explained. By defining an adequate mapping from lipid atoms to CG beads, a five-beads single-tailed lipid molecule was constructed for both POPC and DPPC lipids. Lastly, in chapter three, the iSoLF lipid force field was combined with a Cα protein model, AICG2+, in order to perform MD simulations of proteins inside lipidic environments. In these systems, the hydrophobic/hydrophilic interactions are the main driving force. Furthermore, the parameterized lipid-protein interaction was tested by simulating various classes of proteins.

In conclusion, the iSoLF lipid force field, together with the lipid-protein interaction, represent a useful model for performing CG MD of large membrane systems. Currently, parameters for only two lipids, POPC and DPPC, are available. However, this will be addressed in the next iteration of the force field.

Table of Contents

Chapter 1

Introduction

Molecular dynamics (MD) has been a useful tool for performing simulations of various types of molecular systems. Since its early adoption around 1950, practitioners of this technique have been able to study in-silico molecular systems for which new insights could only be obtained indirectly from experiments (**1**). For this reason, together with the technological development of computing, MD has become one of the most popular simulation techniques (**2**), being used nowadays in various of the most powerful high-performance computing clusters around the globe (**3-5**). In classical MD, each atom is represented with one particle, and the Hamiltonian that describes the dynamics of the system can be defined in a quantum-mechanical or classical way, depending on the study. However, as with any method, MD also has limitations that constraint its applicability to arbitrarily any system of interest. Among these limitations, one of the most severe ones is related to the increase in the consumption of computational resources as the complexity of the system under study increases (**6**). Practically, this forbids us from performing MD simulations of highly complicated systems, such as those encountered usually in materials science or biology. Nevertheless, numerous alternatives that alleviate this problem have also been proposed (**7,8**).

When performing MD simulations of biomolecular systems, the complexity mentioned above arises not only from the many biomolecules usually involved but also due to the environment in which the biological phenomena take place. Even for small biomolecules, most

of the computational cost comes from the simulation of water molecules. Suppose one is interested in studying specific interactions between a small number of atom groups inside a determined context, for example, in the hydration protein pockets or the metal-protein interactions in a holo conformation. In that case, the presence of the solvent may not represent any limitation and, in fact, might be desired. However, if dynamic properties that operate at long distances and large time scales become the target of the study, unless some simplifications are made, MD simulations will not yield results in a sensible time.

One of the most widely-used methods for reducing the complexity of MD systems consists of decreasing the number of degrees of freedom by grouping atoms into coarse-grained (CG) beads (**9**). In CG methods, the way in which atoms are grouped defines a specific mapping, and the number of mapped atoms per CG bead defines a resolution. Both mapping and resolution are, in general, different for each CG model. In the literature, over time, several CG models targeting different biological systems have been developed. For example, for lipid bilayers, there are several CG representations of one phospholipid, ranging from highly coarse-grained models that map one lipid molecule to a single CG bead (**10**) to quite more conservative coarse-grained models which use more than ten CG beads for representing one lipid molecule (**11**). For the latter, the MARTINI force field is the canonical example (**12**). In this force field, a four-to-one mapping is used. That is, four non-hydrogen atoms are grouped together into one single CG bead. Since it has CG representations for a large portion of the liposome, it has become the most popular force field used in the community for studying from middle to large biological membrane systems. However, simulations are still computationally demanding even at this resolution due to the inherently long equilibration times required by membrane systems, often requiring substantial computational resources.

On the other side of the CG spectrum, there are models that opt for a coarser representation of lipids. For example, Cooke. Kremer and Deserno (**13**) developed a lipid force field in which each two-tailed phospholipid is mapped into a three-beads single-tailed CG molecule. Interestingly, this model still reproduces satisfactorily various properties of lipid bilayers despite its high degree of coarse-graining. Additionally, another advantage of this model is the treatment of the solvent. In highly CG models like this, solvent molecules, i.e., water and ions, are represented implicitly in the interaction potentials. In other words, the net effect coming from the lipid-water and lipid-ion interactions are captured in the lipid-lipid interactions. This level of coarse-graining, summed with an implicit representation of the solvent, decreases the computational cost of MD simulations enormously. Thus, these models are commonly used to study phenomena occurring at large scales, like the fusion of lipid vesicles or the invagination of lipid membranes (**14**).

Apart from the two levels of coarse-graining mentioned above, it is quite surprising that, in the literature, there are few models with an intermediate resolution for lipids. A model with an intermediate representation might, at first glance, appear to be of no use. However, there is a specific reason for requiring such a model. In the field of protein CG MD simulation, one of the preferred resolutions for modeling proteins consists of representing each amino acid with one CG bead centered at the Cα position. If these models were to be applied to membrane proteins, a lipid model with a compatible resolution would be desired. In fact, this compatible resolution corresponds to mapping phospholipids to CG molecules composed of five or six beads. Having such a model would permit to expand the applicability of Cα protein models to more realistic membrane environments.

This thesis presents the development of a new implicit solvent lipid model for performing coarse-grained MD simulations of biological membrane systems and is organized in the following way. In Chapter 2, the development of the new implicit solvent lipid model is treated in detail. Then, in Chapter 3, a way to combine the coarse-grained lipid model with Cα protein models is presented, and the application to proteins with different orientations with respect to the membrane is shown. Finally, in Chapter 4, this work ends with some concluding remarks and future directions.

Chapter 2

Coarse-grained implicit solvent lipid force field with a compatible resolution to the Cα protein representation

2.1 Coarse-grained lipid force fields

Though molecular dynamics (MD) simulations have become invaluable for investigating the structural dynamics of biomolecular systems, the time scale required for study many biological phenomena is prohibitively large (**15-17**). To address this constraint, coarse-grained (CG) modeling decreases the number of degrees of freedom by grouping atoms into CG beads, lowering the computational cost of simulations while preserving as much of the properties of interest as possible (**18-20**). Since biomolecular systems are inherently hierarchical, there are a variety of coarse-graining resolutions. Higher-resolution CG models are generally more precise, but they are often more computationally demanding. Thus, depending on the target, a CG model may be more suitable than others. For example, explicit solvent CG models are more accurate, whereas implicit solvent CG models trade-off accuracy in favor of computation efficiency by average the effect of the solvent into the force field interactions (**21**).

CG MD simulations often target biological membrane systems, for which various types of CG lipid models have been produced over the last two decades. Goetz and Lipowsky created an explicit solvent CG amphiphile model in 1998 and effectively simulated lipid bilayer

self-assembly (**22**). Later, Noguchi and Takasu were the first to create an implicit solvent CG model of amphiphiles that exhibited the proper physical properties of a bilayer membrane (**23**). Then, Cooke et al. established a somewhat simplified implicit solvent CG model for lipids based on pairwise interactions (**24**). Both of these implicit solvent models employ three CG beads per lipid, making them simple and generic in essence since they do not need to be parameterized for any particular molecule. Numerous physical properties of membrane structures, such as the gel-liquid phase transfer, phase separation, membrane fusion, and budding, have been effectively investigated using these simple models. Several higher-resolution CG lipid models have been established as a distinct class of models, including the pioneering work of MARTINI by Marrink et al. in 2004 (**25-32**). This class of models employs more than ten CG beads per lipid and specifically represents the two-alkyl-tail geometry, allowing the representation of particular phospholipids. Additionally, the MARTINI model, among others, has been successfully extended to a variety of biomolecules (**33-36**). It is worth mentioning that, with a few exceptions (**27,31,37**), the majority of these models employ explicit solvent molecules, making them more computationally demanding compared to the aforementioned implicit solvent models.

Notably, the majority of biological membrane structures are often composed of membrane proteins. Thus, compatibility with CG protein models is critical for applying CG lipid models to a large number of these biological systems. To model interactions between lipids and proteins, it is critical that both the CG lipid and protein representations have a reasonably similar resolution. For lipids, proteins, and other molecules, for example, the MARTINI force field consistently utilizes a mapping of one CG particle to approximately four non-hydrogen (heavy) atoms. Among the numerous CG representations for proteins, a widely-used representation consists of assigning one CG particle to each amino acid, centered at the

Cα position (**38-42**). Therefore, since there are on average 8.4 ± 2.4 non-hydrogen atoms per amino acid, the Cα representation achieves a reduction in order of magnitude for the degrees of freedom. POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) and DPPC (1,2-dipalmitoyl-sn-glycero-3-phosphatidylcholine) are two representative phospholipids that include 52 and 50 non-hydrogen atoms, respectively. Consequently, using from five to six CG particles to represent each lipid molecule would produce a resolution consistent with the Cα protein model. In the literature, among several lipid models (**43-49**), there are just a handful of them with a resolution of 5-6 CG particles per lipid, but not necessarily design to be combined with or extended to membrane proteins.

The main objective of using five CG beads is not to model generic lipid molecules but to parameterize the model for various phospholipid molecules, d in particular, POPC and DPPC. It is well established that pure POPC lipid membranes are in the liquid disordered process at physiological temperatures (30 °C, for example), while pure DPPC lipid membranes are in the gel phase **(50,51**). It is important to reproduce these two phases because biological membranes are composed of a combination of unsaturated and saturated phospholipids, as well as membrane proteins and others. Thus, a new implicit solvent CG model that reproduces the correct phase behavior around 30 °C and possesses a resolution that works well with existing Cα protein models was developed.

2.2 Lipid model

In the model, a two-tailed glycerophospholipid is mapped to a CG linear chain molecule (**Figure 2.1a**). Each CG lipid comprises five beads, two polar head beads (H1 and H2), and three hydrophobic tail beads (T1, T2, and T3). The H1 bead represents the

characteristic chemical group bounded to the phosphate. The H2 bead represents the phosphate, glycerol, and ester carbonyls. The T1, T2, and T3 beads correspond to the first five carbon atoms in each tail, the next five carbon atoms in each tail, and the remaining carbon atoms in each tail, respectively. As previously said, this five-bead mapping provides a resolution comparable to that of Cα protein models.

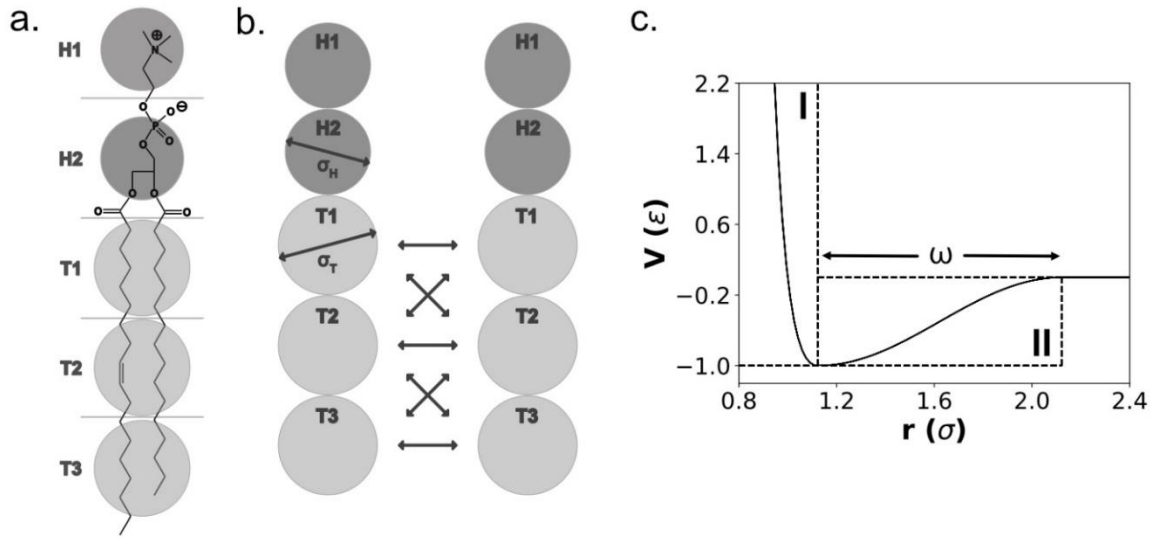

Figure 2.1. (**a**) Mapping of a lipid (POPC in the figure) to a five-beads CG molecule. H1 and H2 represent the polar head beads. T1, T2, and T3 represent the hydrophobic tail beads. (**b**) Diagram of the attractive interactions between tail beads. Notably, no attractive interaction is applied between T1 and T3 beads. (**c**) Interaction potential for the lipid tail beads. (I) correspond to the repulsive portion of the potential and (II) to the attractive part.

In the lipid model, the energy function is composed of four terms:

$$V = V_{Bond} + V_{Angle} + V_{Repulsion} + V_{Attraction} \tag{2.1}$$

The first term of the potential, V_{Bond}, represents the bonding interaction between adjacent CG bead pairs of the same lipid molecule and is modeled as follows:

$$V_{Bond} = \sum_{i=1}^{n_{bonds}} k_{bond,i}\left(b_i - b_{0,i}\right)^2 \tag{2.2}$$

Here, $k_{bond,i}$ is the force constant, b_i is the *i*-th virtual bond length between consecutive CG beads, $b_{0,i}$ is the point of minimum energy for the virtual bond, and n_{bonds} is the total number of virtual bonds. The next term, V_{Angle}, represents the virtual bond-angle interaction between two consecutive virtual bonds in the same lipid molecule:

$$V_{Angle} = \sum_{i=1}^{n_{angles}} k_{angle,i}\left(\theta_i - \theta_{0,i}\right)^2 \tag{2.3}$$

Here, $k_{angle,i}$ is the force coefficient, θ_i is the *i*-th angle, $\theta_{0,i}$ is the equilibrium value for the *i*-th angle, and n_{angles} is the total number of angles.

The two remaining terms of **Eq. (2.1)** represent the intermolecular interactions between different lipid molecules and have the same functional form as in reference **24**. The repulsive term, $V_{Repulsion}$, corresponds to the Weeks-Chandler-Andersen potential:

$$V_{Repulsion} = \sum_{i<j}^{n_{nl-pairs}} \begin{cases} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} + \frac{1}{4}\right], & r_{ij} \le \sqrt[6]{2}\sigma_{ij} \\ 0, & r_{ij} > \sqrt[6]{2}\sigma_{ij} \end{cases} \tag{2.4}$$

with ε_{ij} representing the force scaling coefficient, σ_{ij} the repulsive distance for the ij pair of beads, and $n_{nl-pairs}$ the total number of non-local pairs. This potential is applied to all the pairs of beads that do not participate in either a virtual bond or a bond-angle interaction. In order to reduce the complexity of the model, σ_{ij} and ε_{ij} are defined with combination rules. On the one hand, σ_{ij} is defined as $(\sigma_i + \sigma_j)/2$ where σ_i (σ_j) represents the van der Waals diameter of the *i*-th (*j*-th) CG particle. In each lipid molecule, there are two values for σ_i: one for the head beads (σ_H) and one for the tail beads (σ_T) (**Figure 2.1b**). As it will be discussed later, these values are constraint by $\sigma_H = 0.65\sigma_T$ in order to prevent the formation of persistent

pores in the membrane. On the other hand, ε_{ij} is defined as $\sqrt{\varepsilon_i \varepsilon_j}$, where ε_i depends is characteristic for each lipid.

Finally, $V_{Attraction}$ represents the interaction between the hydrophilic tail beads and has the following form:

$$V_{Attraction} = \sum_{i<j}^{n_{nl-pairs}} \begin{cases} -\varepsilon_{ij}, & r_{ij} \leq \sqrt[6]{2}\sigma_{ij} \\ -\varepsilon_{ij} cos^2\left[\frac{\pi}{2\omega_{ij}}\left(r_{ij} - \sqrt[6]{2}\sigma_{ij}\right)\right], & \sqrt[6]{2}\sigma_{ij} < r_{ij} \leq \sqrt[6]{2}\sigma_{ij} + \omega_{ij} \\ 0, & \sqrt[6]{2}\sigma_{ij} + \omega_{ij} < r_{ij} \end{cases} \quad (5)$$

The key feature of this potential is a tunable attraction width, which has been shown to represent the mid-range attractive hydrophobic interactions effectively (**52**). As for the coefficients ε_{ij} and σ_{ij}, they have the same values of the repulsive potential. However, ω_{ij} is defined as $(\omega_i + \omega_j)/2$ and represents the width of the pair potential well. This attractive interaction is only applied between tail beads of different lipid molecules except for T1-T3 pairs (**Figure 2.1b**). The decision to exclude T1-T3 prevents the formation of unrealistic membrane conformations. The total interaction between two lipid tail beads is depicted in **Figure 2.1c**. Hereafter, this *i*mplicit *so*lvent *l*ipid *f*orce field will be termed iSoLF.

2.3 Coarse-grained molecular dynamics simulations

All the CG MD simulations were performed using a modified version of CafeMol v3.1 (**53**) and the standard underdamped Langevin equation. For each CG bead, the mass was set equal to the sum of the masses of individual atoms involved in the CG bead.

Periodic boundary conditions and a semi-isotropic pressure coupling in the xy-direction were used during the optimization of the force-field parameters, the estimation of the physical

properties of plane membranes, the evaluation of the temperature dependence of lipids, the observation of pore formation, and the observation of the phase behavior of POPC/DPPC membranes. To integrate the equations of motion, a method developed by Gao, Fang, and Wang (40) was used. The thermostat's friction coefficient was set to 0.1 (1/CafeMol-time) for the Langevin dynamics. In CafeMol v3.1, one CafeMol-time unit is approximately 49 fs. The friction coefficient for the barostat was set equal to 0.1 (1/CafeMol-time), and the compressibility of the box equal to 0.01 ($Å^3 \cdot mol/kcal$). MD time steps of 0.2 (CafeMol-time) were used to simulate pure POPC systems and 0.1 for simulations involving DPPC and vesicles. With a time step of 0.2, DPPC-containing systems were unstable. This occurred because the equilibrium distance for the DPPC T2-T3 bond was larger than the other equilibrium lengths, enabling beads to pass through. Thus, by halving the integration step, repulsive interactions prevented beads from crossing virtual bonds. For the force field parameterization, each simulation was performed for 1×10^6 and 2×10^6 MD steps for POPC and DPPC, respectively. For the simulations of the temperature dependence, 1.2×10^6 and 2.4×10^6 MD steps were used for POPC and DPPC systems, respectively. Here, only the first sixth of the data points was discarded (**Figure A1**). For the mixed POPC/DPPC system, three simulations were performed. For the separated system, one of 2×10^6 MD steps, and for the mixed systems, two of 3.4×10^6 MD steps. For these trajectories, the first 0.4×10^6 MD steps were discarded.

For the simulations corresponding to the spontaneous formation of a lipid bilayer and the vesicle equilibration, the default setup of CafeMol was used. Additionally, a fixed-size box with periodic boundary conditions, and the NVT ensemble were used. For the lipid bilayer formation, the simulation was 2×10^6 MD steps long. Whereas for the vesicle simulations, an

initial heating was performed for 0.6×10^6 MD steps, followed by one long run of 3.4×10^6 MD steps, and three short runs of 1×10^6 MD steps, all of them at constant temperature.

The initial configuration for the spontaneous lipid bilayer formation was built by sequentially placing lipid molecules in a box. First, the H1 bead was randomly placed, and then the rest of the molecule (beads H2, T1, T2, and T3) were added following a straight line randomly oriented. If while completing the lipid, any of its beads had a distance lower than 1.3σ from any other bead already placed, it was discarded, and the process started again (H1 bead placement).

The initial configuration for the vesicle simulation was built following a simple geometric method that permits to distribute points on the surface of a sphere while keeping them separated as evenly as possible by mapping Fibonacci lattice (**55**). Considering the physical dimensions of a POPC lipid, the prepared vesicle of radius 15 nm contained in the inner and outer leaflets 2976 and 4400 lipids, respectively.

Finally, the initial configuration for the self-assembly of lipids was prepared by randomly placing 5000 POPC lipids in a box of $500\text{Å} \times 500\text{Å} \times 500\text{Å}$. Periodic boundary conditions were not used. Instead, each wall of the box had assigned a repulsive potential. This simulation was performed for 8×10^5 MD steps.

2.4 All-atom molecular dynamics simulations

The bottom-up parameterization and the calculation of physical properties required all-atom MD simulations. These simulations were performed using GROMACS (**56**) version 5.1.1,

the lipid force field Slipids (**57,58**), and the TIP3P water model (**59**). Pre-equilibrated bilayers of POPC and DPPC were downloaded from the Slipids website (**60**). The simulation protocol consisted of a short energy minimization phase using the steepest descent method, an NVT equilibration phase at a constant temperature of 303K for 200 ps using the v-rescale thermostat, and an NPT equilibration phase at a constant pressure of 1.013 bar and a constant temperature of 303K for 5 ns using the Parrinello-Rahman barostat. In these simulations, lipid and water molecules were coupled separately, using time constants of 0.5ps and 10ps for the thermostat and barostat, respectively. Then, after all the equilibration phases were completed, production runs were performed for 200 ns.

Finally, for comparing the running times of the CG and all-atom models, all-atom and CG membrane patches of POPC were equilibrated with the protocols mentioned above. Then, production runs were performed in order to estimate the 2D MSD using 1 CPU core of an Intel i7-5930K processor and no GPUs.

2.5 Calculation of properties

For the planar membranes, the area per lipid (A_L), the order parameter (S_θ), the hydrophobic thickness, and the 2D diffusion coefficient were calculated. The formula employed for A_L was:

$$A_L = 2\frac{A_{xy}}{n_{lipids}} \tag{2.6}$$

where A_{xy} is the cross-sectional area of the simulation box in the xy-plane and n_{lipids} corresponds to the total number of lipids in the box.

The order parameter, S_θ, was calculated by measuring the angle θ_i between the z-axis and the line joining the center of mass of the tail beads T1 and T3 (**Figure 2.2a**) by the following formula:

$$\langle S_\theta(t)\rangle = \frac{1}{n_{lipids}} \sum_{i=1}^{n_{lipids}} \frac{1}{2}[3cos^2(\theta_i(t)) - 1] \tag{2.7}$$

where n_{lipids} represents the total number of lipids in the system, and $\langle \cdots \rangle$ means the average over all lipids.

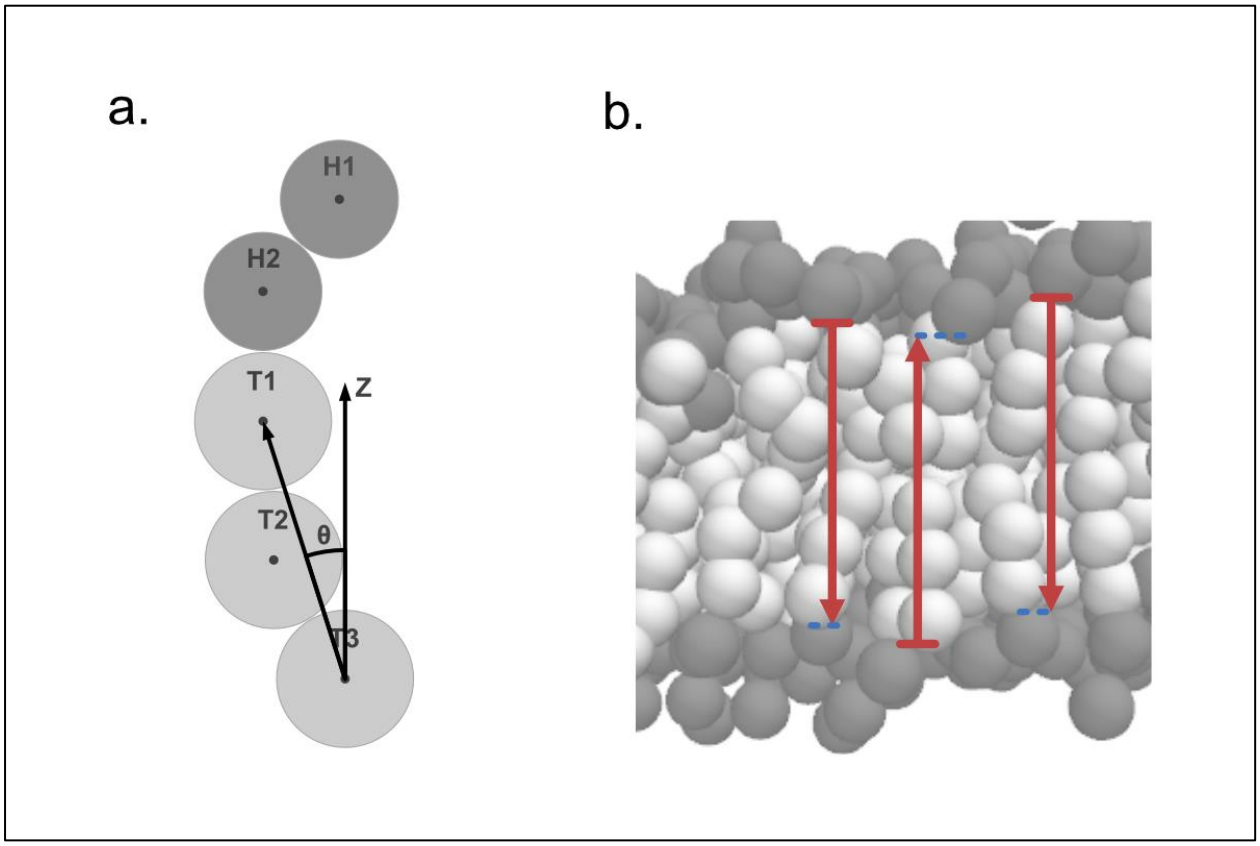

Figure 2.2. Order parameter and membrane thickness calculation. (a) A representation of the angle θ used for the calculation of the order parameter. It corresponds to the angle between the line joining beads T1 and T3, and the z-axis. (b) Local hydrophobic thickness for three lipids. In each lipid molecule in one leaflet, another lipid molecule in the opposite leaflet is found such that the distance in xy-plane between the two lipids is the smallest (blue dashed lines in the figure). The difference in the z-axis between the pair of lipids (indicated by red arrows in the figure) corresponds to the hydrophobic thickness at that point.

The hydrophobic was calculated using a method similar to the one implemented in the GridMAT-MD Software (**61**). First, for each lipid molecule, the middle point between the H2 and the T1 beads was calculated. Then, after finding the lipid with the closest distance in the xy-plane contained in the opposite leaflet, the difference in the z-axis is calculated, corresponding to the hydrophobic thickness at that lipid site. The membrane thickness is defined as the average over all the calculated hydrophobic thicknesses (**Figure 2.2b**).

Finally, in each simulated bilayer, the calculated physical properties were used to describe the phase behavior. A slow lateral diffusion was characteristic of a gel phase. In contrast, a fast lateral diffusion corresponded to the liquid phase. Furthermore, in the liquid phase, two subphases, liquid-ordered and liquid-disordered, were characterized by the lipid order parameter. For each lipid membrane, significant changes in the area per lipid, the order parameter, and the hydrophobic thickness occurred around the same temperature, which was identified as the phase transition temperature.

2.6 Model parameterization

Force field parameters were determined for two target phospholipids, POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) and DPPC (1,2-dipalmitoyl-sn-glycero-3-phosphatidylcholine). POPC and DPPC lipid membranes exhibit at near-physiological temperatures the liquid disordered and gel phases, respectively. In the parameter determination, a partly bottom-up and partly top-down approach was used.

The virtual-bond and bond-angle potential parameters were parameterized following a bottom-up approach. First, all-atom simulations of single-component lipid membranes of

POPC and DPPC were performed. Then, Eq. (**2.2**) and Eq. (**2.3**) were fitted by applying the standard Boltzmann inversion method to the calculated ensembles (**63**). The Boltzmann-inverted potentials (**Figure A2**) were well approximated by harmonic potentials near the minima but deviated considerably far from it. Therefore, a non-linear least-square fitting near the minima was used for fitting. **Table 2.1** lists the obtained parameters. It is important to mention that lipids forming a bilayer were used as reference structures for the parameterization. Thus, the behavior of lipids outside the membrane might not be correctly represented.

Type	Coefficient	POPC	DPPC
Bond	k_{H1-H2}	0.446	0.471
	k_{H2-T1}	1.073	1.320
	k_{T1-T2}	1.001	0.875
	k_{T2-T3}	0.443	0.280
	$b_{0,H1-H2}$	5.580	5.417
	$b_{0,H2-T1}$	5.452	5.824
	$b_{0,T1-T2}$	5.050	6.312
	$b_{0,T2-T3}$	5.095	6.299
Angle	$k_{H1-H2-T1}$	0.600	0.582
	$k_{H2-T1-T2}$	2.383	3.357
	$k_{T1-T2-T3}$	0.880	4.823
	$\theta_{0,H1-H2-T1}$	3.142	3.142
	$\theta_{0,H2-T1-T2}$	3.142	3.142
	$\theta_{0,T1-T2-T3}$	3.142	3.142

Table 2.1. Parameters for the intramolecular interactions of POPC and DPPC. Force coefficients k are in $kcal/Å^2 mol$ for the virtual bond, and in $kcal/mol$ for the virtual bond-angle. Equilibrium distances b_0 are in Å, and equilibrium angles θ_0 are in radians.

The parameterization of the intermolecular repulsive and attractive potentials was performed following a top-down approach, in which the parameters were optimized so CG MD

simulations could reproduce three properties of lipid membranes. A cost function was also defined in order to apply optimization methods:

$$Cost(\varepsilon, \sigma, \omega) = \sum_{i=1}^{3} \left(\frac{p_{i,sim}(\varepsilon, \sigma, \omega) - p_{i,ref}}{p_{i,ref}} \right)^2 \quad (2.8)$$

with $p_{i,ref}$ being the reference value for the i-th property and $p_{i,sim}$ the *i*-th property calculated from CG MD simulations that depend on the force field parameters, ε, σ, and ω. The three properties selected for fitting were the area per lipid (APL), the hydrophobic thickness, and the order parameter in the bilayer membrane (**Fig. 2.3a-2.3c**). For POPC, the first two properties, the reference values, were taken from experimental measurements (**64**), and the last property, from all-atom simulations. In contrast, for DPPC, all the reference values were taken from all-atom simulations.

	POPC	DPPC
ε	0.416	0.464
σ_T	7.111	6.900
ω	9.867	10.318

Table 2.2. Coefficients for the intermolecular interactions of POPC and DPPC. ε is in kcal/mol, and σ and ω are in Å.

Parameters that minimize the cost function are sought. Since calculating the derivatives of the cost function with respect to the force field parameters is computationally very expensive, the gradient-free Nelder-Mead method (**65**) was employed. In this method, up to the desired precision, a boundary enclosing a minimum in the parameter space is refined in each iteration step. With a suitable set of initial values, convergence was achieved within some tenths of iterations (**Fig. 2.3d-2.3f**). The optimized parameters for POPC and DPPC lipids are given in **Table 2.2**.

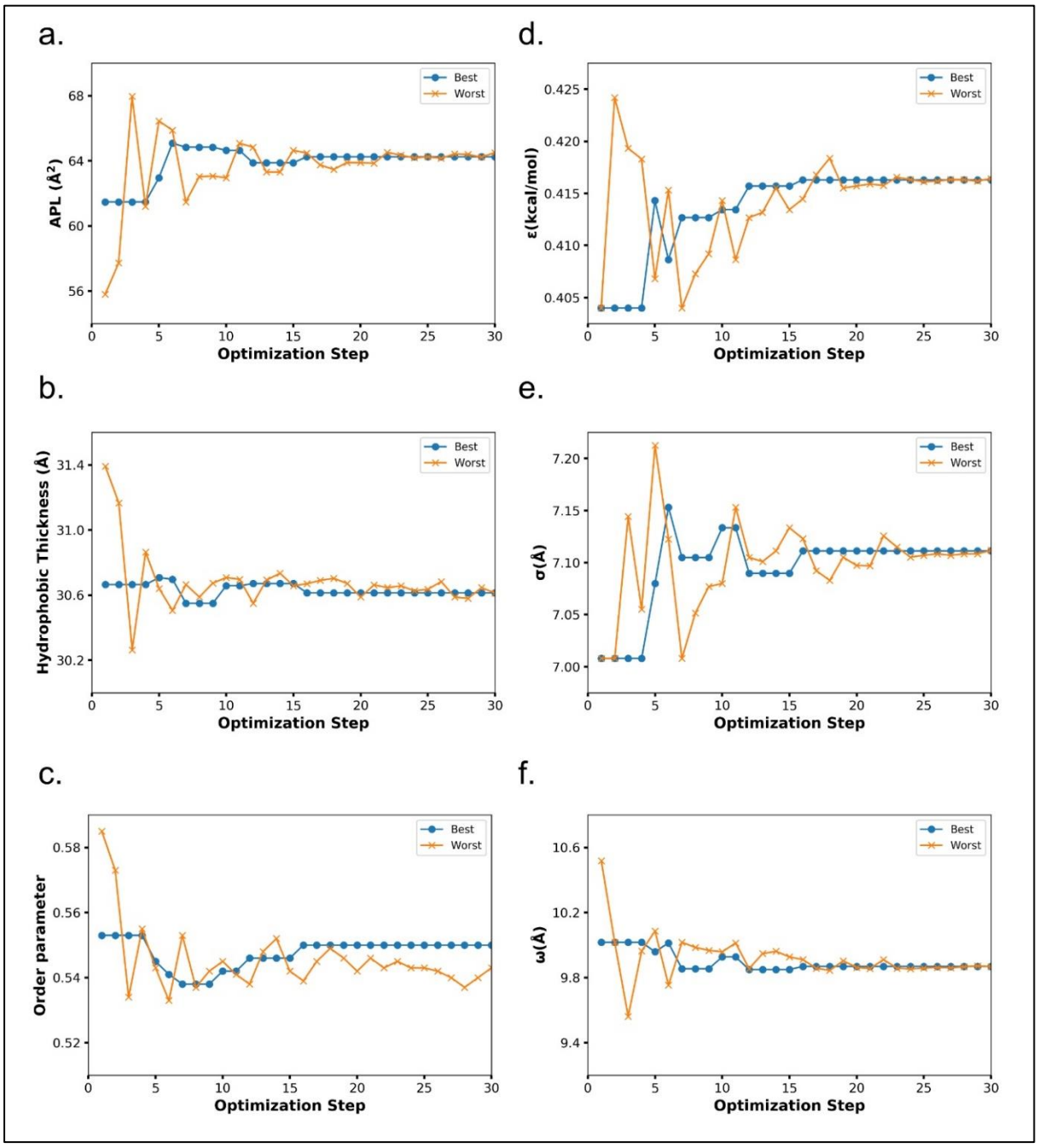

Figure 2.3. The parameter optimization process for POPC using the Nelder-Mead method. (a)-(c) The best and the worst points for the target properties in each optimization step. (d)-(f) The best and the worst points for the force field coefficients in each optimization step.

2.7 Spontaneous membrane formation

With the optimized set of parameters, the spontaneous formation of lipid bilayer membranes with the CG force field was examined. A system containing 200 POPC lipid molecules randomly placed in a box was prepared. Fixing the size of the box, L_x, L_y, and L_z, to 64, 64, and 80 Å, produced the equilibrium APL for POPC at 30 °C. With this setup, the lipids acquired a lipid bilayer configuration within 10^4 MD steps (**Fig. 2.4**). After forming the membrane, no pore was formed, suggesting the membrane conformation is thermodynamically stable.

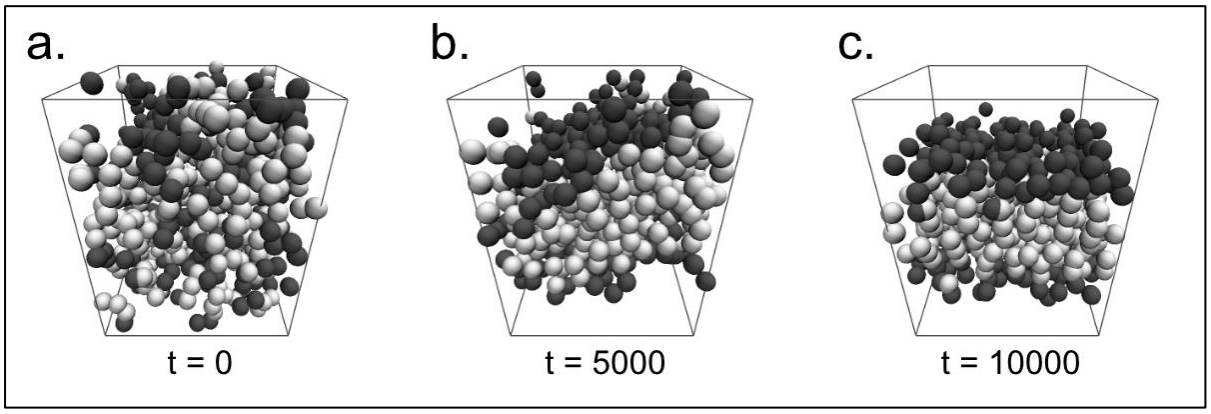

Figure 2.4. Spontaneous lipid bilayer membrane formation. (a) Simulation for 200 POPC lipids at 30 °C, starting from a random configuration. (b) Lipids begin to form a membrane-like conformation after 0.5×10^4 md-time. (c) Lipids adopt a membrane conformation after 1.0×10^4 md-time. Here, lipid head and tail beads are in dark-gray and white color, respectively.

In order to confirm that no artifact was introduced by the usage of the periodic boundary conditions, a larger scale simulation for the self-assembly of 5000 POPC lipids in a cubic box of 50 nm^3 without periodic boundary conditions was performed (**Figure A3**). Starting from a random configuration (**Figure A3a**), the formation of many small clusters was observed. However, not a merged bilayer membrane nor vesicle (**Figure A3b**) was observed. Among

many clusters, relatively larger clusters formed bicelle-like structures were found **Figure A3c**). This might suggest that a much longer time is required in order to obtain a unified bilayer membrane.

Two findings in preliminary studies are worth mentioning. The first one is related to the stability of the CG simulations. When the ensemble of zero surface tension in the xy-direction with variable box size was used, CG MD simulations starting from random conformations were highly unstable, making the system box expand indefinitely. Thus, to avoid this, the NVT ensemble was used in these cases.

The second one is related to the stability of the formed lipid bilayer membrane. By choosing $\sigma_H = 0.85\sigma_T \sim 1.00\sigma_T$ pores started to appear spontaneously in the membranes (**Fig. 2.5a** and **Fig. A4**). These pores were very stable, and the membranes were trapped in these conformations (**Fig. 2.5a**). Interestingly, this behavior was already reported by the original author group (**66**). When the system box size is allowed to change, it occasionally expands and creates a transient cavity in the membrane. Then, this transient cavity induces a tilt of the surrounding lipids, increasing the repulsive energy between head and tail beads. Thus, to reduce the repulsion, the system forms a pore. The stability of the pores depends on the ratio σ_H/σ_T between the head and tail beads. This ratio has a transition point around $\sigma_H/\sigma_T \sim 0.75$ (**Fig. A4**). Since the spontaneous pore formation is not desired, a ratio of $\sigma_H = 0.65\sigma_T$ was chosen. Furthermore, even after a pore was formed by using a ratio of $\sigma_H = 0.85\sigma_T$, just changing the ratio to $\sigma_H = 0.65\sigma_T$ made the pore disappeared (**Fig. 2.5b**). Thus, to make lipid membranes stable, it is necessary a small ratio for the size of the head and tail beads.

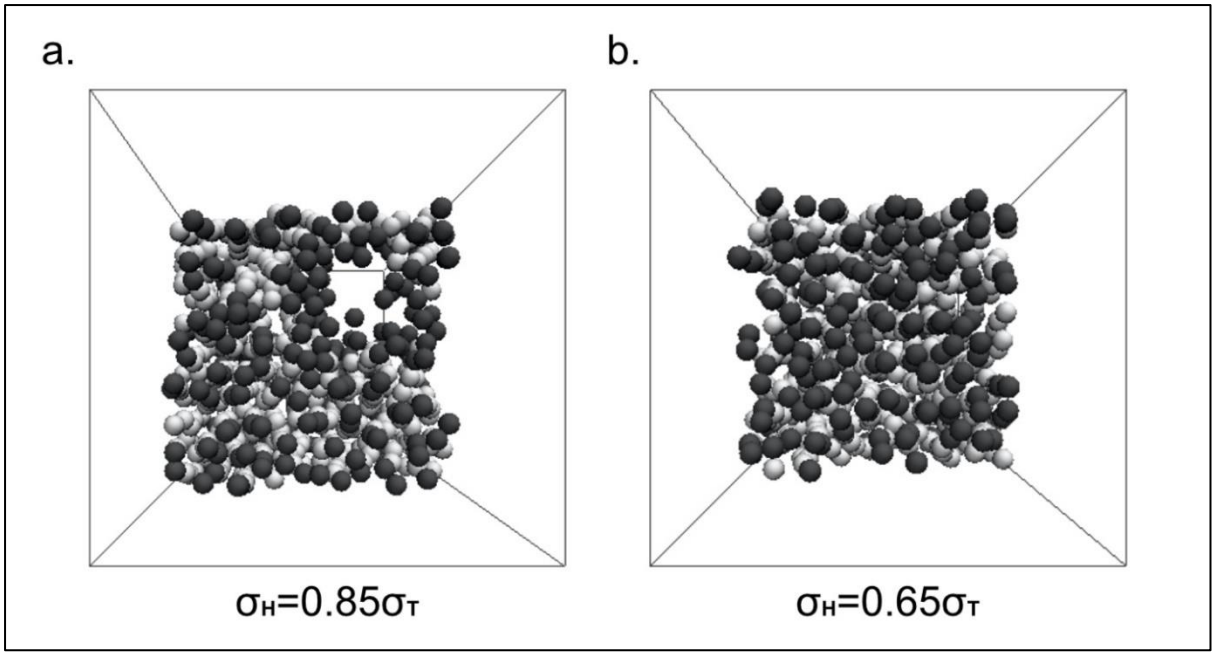

Figure 2.5. Pore formation in lipid membranes. Pores depend on the ratio σ_H/σ_T. (a) A ratio of $\sigma_H = 0.85\sigma_T$ results in the spontaneous formation of a pore. (b) A ratio of $\sigma_H = 0.65\sigma_T$ makes pore disappear. Here, lipid head and tail beads are in dark-gray and white color, respectively.

2.8 Lateral diffusion

The lateral diffusion of POPC and DPPC was also evaluated. To quantify it, the MSD in 2D at 30 °C was computed (**Fig. 2.6a**). It was found that the MSD as a function of time differences fitted well a straight line, suggesting normal 2D diffusion. Comparing the slope of the MSD of the two lipids showed that, at 30 °C, the POPC membrane was in a liquid phase, whereas the pure DPPC membrane was in a gel phase. In order to confirm this, the diffusion coefficient of POPC and DPPC at different temperatures was calculated (**Fig. 2.6b**). An apparent phase transition from gel to liquid phases around 25 °C was observed for POPC and around 95 °C for DPPC. Later, this was further confirmed by a similar behavior in the other properties around the same tested temperatures.

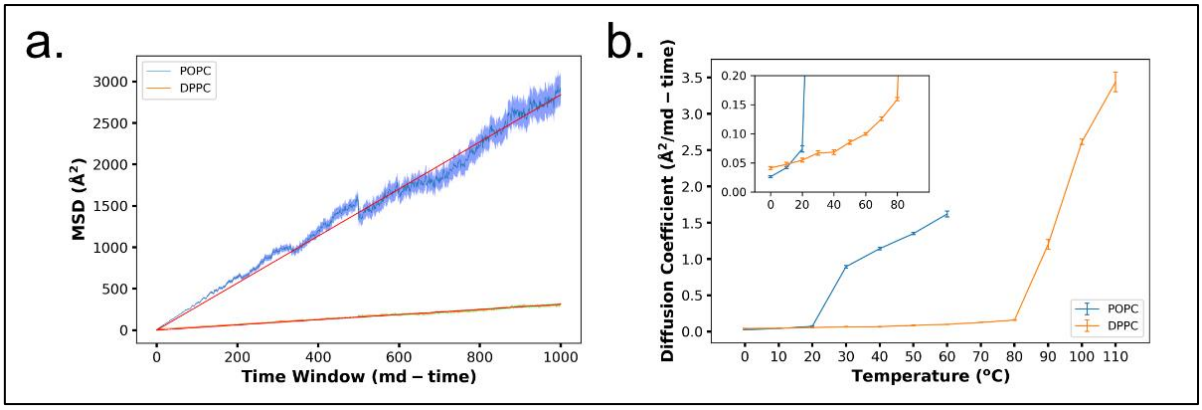

Figure 2.6. 2D diffusion of POPC and DPPC. (a) Mean square displacement (MSD) at a temperature of 30 °C for POPC (blue) and DPPC (orange). POPC membranes presented a liquid phase, whereas the DPPC membrane stayed in a gel phase. Red lines correspond to the fitted equation employed to calculate the diffusion coefficient. (b) 2D diffusion coefficient for POPC and DPPC as a function of temperature. An apparent phase transition occurs at around 25 °C and 95 °C for POPC and DPPC, respectively. The subplot shows a zoomed-in version of the lower portion of the plot.

The 2D MSD was also used for comparing the running time of the coarse-grained lipid model against a standard all-atom model. Using only 1 CPU core for both simulations, the MSD was calculated (**Fig. A5**). From the all-atom simulation, an MSD of 0.174 nm^2 was obtained in about 8 $hours$ 47 $minutes$. On the other hand, an MSD of 36.7 nm^2 was obtained in about 21 $minutes$ with the CG model. Assuming that the MSD increases linearly in time, the CG model achieved a speed-up factor of ~5000 relative to the all-atom model. In perspective, to get an MSD value from all-atom models comparable with the one obtained in CG simulations, about 2 months and 17 days would be needed if the same resources were used.

2.9 Vesicle dynamics

Next, CG MD simulations of a vesicle made of POPC lipids were performed. For this, a small unilamellar vesicle (SUV) with a diameter of ~30 nm (**Fig. 2.7a**) was prepared. Preliminary tests suggested that, perhaps due to a poor setup of the initial structure, starting the

CG MD simulations at room temperature causes an unstable behavior of the vesicle. Thus, to avoid this instability, vesicles started at a temperature of 0K = - 273 °C, and then they were gradually heated until they reached 30 °C. During this process, some lipid in the outer leaflet of the vesicle diffused away without affecting the overall shape of the vesicle. However, by setting the integration time step to half its original value, i.e., 0.1 md-time, the number of escaping lipids considerably decreased. When the vesicle reached 30 °C (**Fig. 2.7b**), the lipids that were not forming part of the vesicle were removed, and then production runs were performed. All four trajectories showed a stable vesicle. Additionally, for one long trajectory, fluctuations in shape were observed, exhibiting an ellipsoid configuration (**Fig. 2.7c**). To further evaluate the stability of the vesicle, the average radius of the ellipsoid that best fitted the vesicle as a function of time was also monitored (**Fig. 2.7d**).

2.10 Temperature dependence

The parameterization of the CG lipid force field was performed at 30 °C, reproducing well both the reference properties and the corresponding phases of POPC and DPPC (**Fig. 2.3** and **Fig. 2.6b**). Then, there is the question about the usability of the force field at different temperatures. After simulating POPC and DPPC lipid membranes at different temperatures from 0 °C to 110 °C, the area per lipid (APL), the hydrophobic thickness, the order parameter (**Fig. 2.8**), and the lateral diffusion coefficient (**Fig. 2.6b**) were calculated.

The area per lipid, the hydrophobic thickness, and the order parameter exhibited changes nearly at the same temperature as that in the lateral diffusion coefficient, both for POPC and DPPC (**Fig. 2.8**), supporting a phase transition from the gel phase to the liquid disordered phase (**67-69**). For the pure POPC membrane, both the area per lipid and the

hydrophobic thickness stayed correlated with the experimental values at temperatures within the range of 30 - 60 °C. However, a phase transition around 25 °C was observed in the CG model (**Fig. 2.3** and **Fig. 2.6b**), whereas, experimentally, it is known to occur at around -2 °C. This shows that the current parameterization is not tuned for reproducing phase transitions.

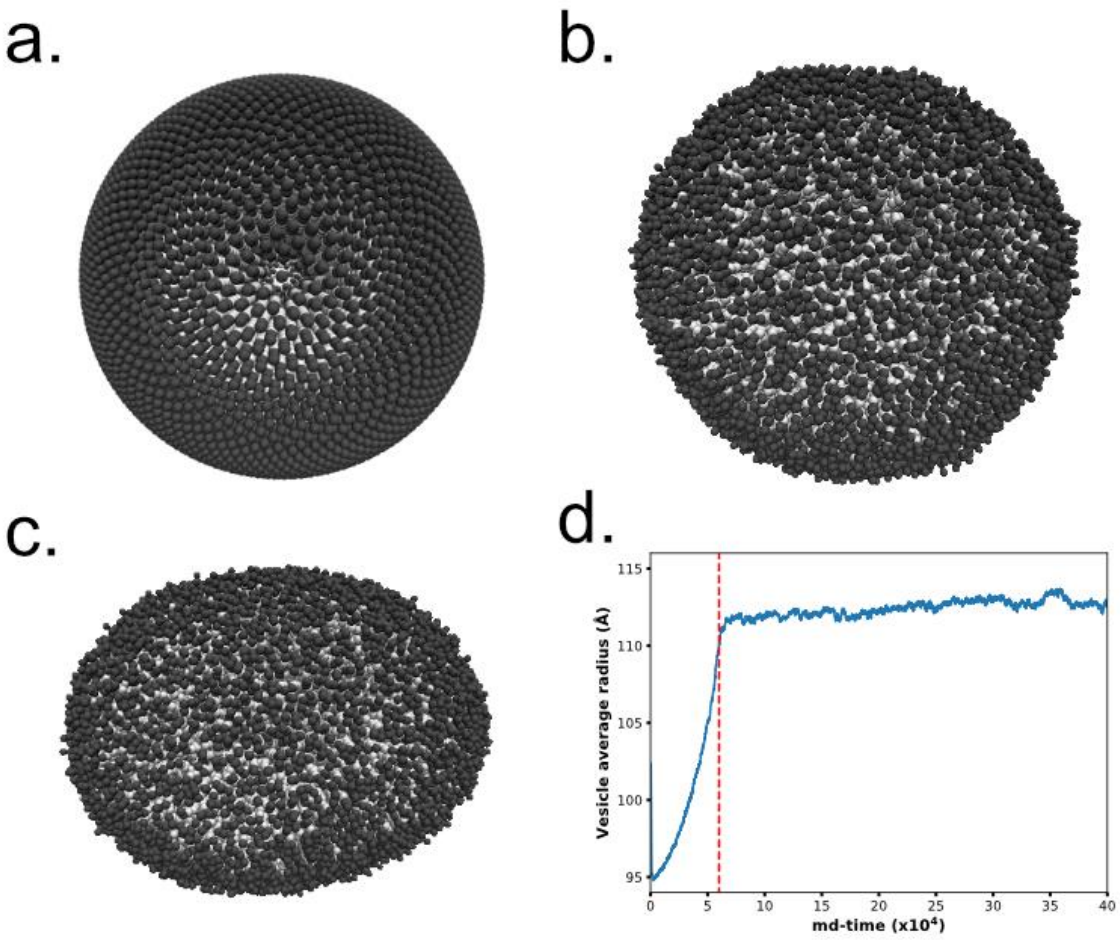

Figure 2.7. CG MD vesicle simulation. (a) Vesicle initial conformation. (b) Vesicle after reaching a temperature of 30 °C. (c) Snapshot of the final conformation adopted by the vesicle. (d) Average radius of the best fit ellipsoid during the heating process (the left of the red vertical line) and the production run. First, the vesicle was heated from 0 K = -273 °C to 30 °C and then kept at a constant temperature of 30 °C. Here, lipid head and tail beads are in dark-gray and white color, respectively.

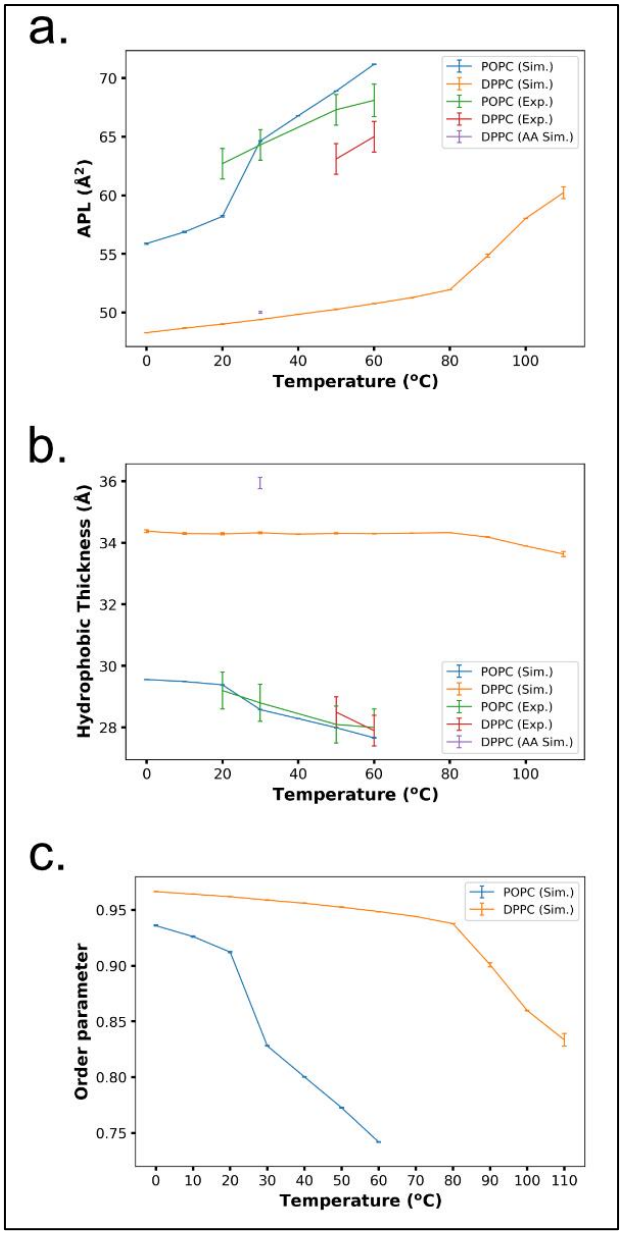

Figure 2.8. Temperature dependence of membrane properties. Comparison of the temperature dependence for (a) the area per lipid (APL), (b) the hydrophobic thickness, (c) and the order parameter. Experimental data are available only in the liquid phase for POPC. Purple squares represent calculated values from all-atom MD simulations of DPPC membranes. The maximum errors are ± 0.04 and ± 0.009 for the APL, ± 0.01 and ± 0.004 for the hydrophobic thickness, and $\pm 7.1 \times 10^{-4}$ and ± 0.006 for the order parameter of POPC and DPPC, respectively.

For DPPC lipid membranes, it has been experimentally determined that the gel-liquid phase transition occurs at a temperature of 41 °C. With the current parameters, a sharp gel-liquid phase transition at around 95 °C (**Fig. 2.6b**) was observed in the simulations. Again, the

phase transition temperature could not be accurately reproduced. Also, at higher temperatures, the estimates of the area per lipid and the hydrophobic thickness deviate from experimental data (**Fig. 2.8**).

Overall, the parameterized lipid model reproduces well the physical properties of lipid bilayers at the calibrated temperature of 30 °C. However, phase transition temperatures can not be reproduced.

2.11 Two-component membrane system

Finally, the behavior of a membrane composed of POPC and DPPC was tested. A membrane consisting of 256 POPCs and 256 DPPCs was simulated at 50 °C. The initial configuration had POPC lipid molecules localized in one half of the membrane and DPPC lipids in the other half (**Fig. 2.9a**). In an early stage of the simulation, two different phases were confirmed: a liquid phase composed of POPC lipids and a gel phase composed mainly of DPPC lipids (**Fig. 2.10a**).

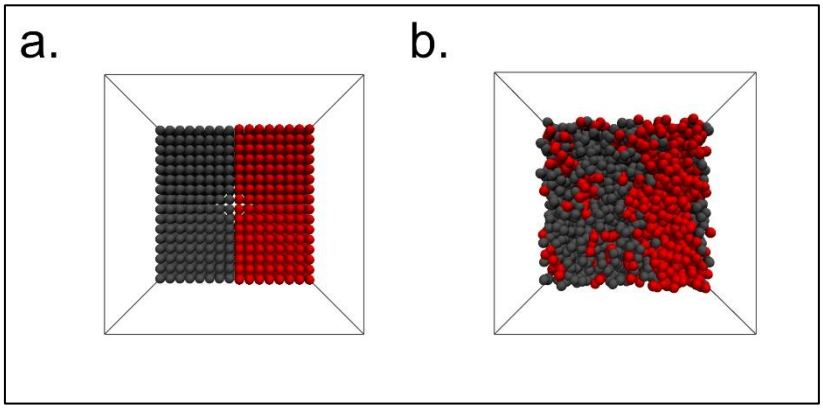

Figure 9. Simulation of a POPC-DPPC system with a ratio of 1:1. (a) Initial configuration. Grey and red molecules represent POPC and DPPC lipids, respectively. (b) POPC lipids stayed in a gel phase when surrounded by DPPC lipids. On the other hand, DPPC lipids stayed in a liquid phase when diffused towards the POPC phase.

As the system evolved, some exchanges of lipids around the interface between the DPPC and POPC phases were observed. It was found that the POPC lipids that moved to the gel phase had almost zero diffusion, whereas the DPPC lipids that moved to the liquid disordered phase presented a faster diffusion. Consistently, the order parameter calculation suggests that DPPC lipids exhibit liquid disordered-like behavior when surrounded by POPC lipids, despite being at a temperature that favors the gel phase of DPPC (**Fig. 2.10b**). In the same way, POPC lipids exhibited a gel-like behavior when surrounded by DPPC lipids.

Another simulation with the same 1:1 ratio for POPC and DPPC but starting from a randomly mixed configuration at 50 °C (**Fig. S6**) was performed. However, in this case, only one phase was observed. The system stayed in the gel phase (**Fig. S6a, c**) and kept that configuration throughout the trajectory. This suggests that both the phase-separated and the fully mixed conformation are stable.

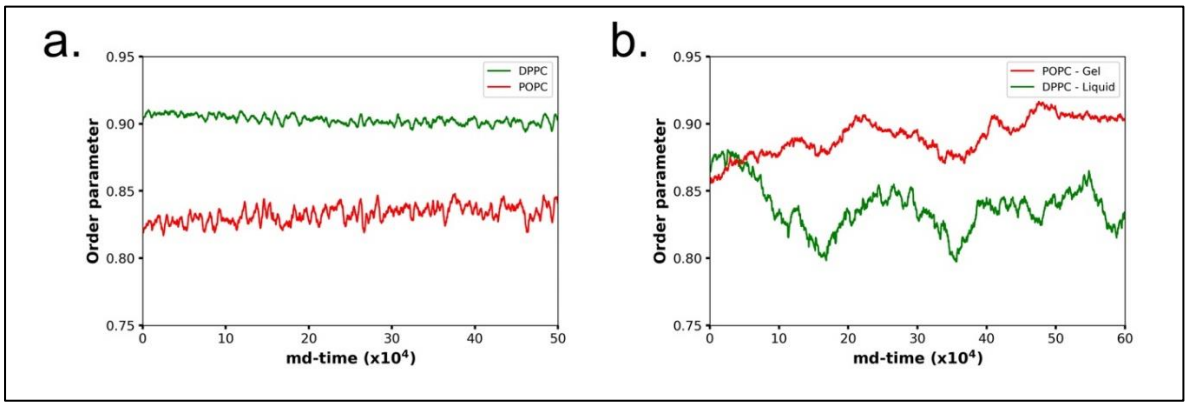

Figure 10. Order parameters calculation for POPC and DPPC lipids in a membrane with a 1:1 ratio. (a) Average order parameter for POPC and DPPC lipids. (b) Order parameter of two specific lipids. The red curve shows the order parameter of a POPC lipid that diffuses into the DPPC gel phase, and the green curve the order parameter of a DPPC lipid that diffuses into the POPC liquid phase.

Finally, a membrane with an increased ratio of 2:1 for POPC and DPPC was also examined. In contrast to the 1:1 ratio membrane, it was found that the fully-mixed membrane

remained in a liquid phase (**Fig. S6b, d**), suggesting that by controlling the ratio of POPC to DPPC in the membrane, a liquid phase for both lipids can be achieved at near-physiological temperatures.

2.12 Conclusions

In this chapter, an extension of the three-bead lipid model developed by Cooke, Kremer, and Deserno into a five-bead model was presented. The model was parametrized for two phospholipids, one unsaturated, POPC, and the other saturated, DPPC lipids. The developed model, iSoLF, could reproduce the area per lipid, the hydrophobic thickness, and the phase behaviors of the target phospholipids at 30 °C. Also, the model membranes of POPC and DPPC presented the correct phase behavior expected from experiments. The spontaneous formation of a lipid bilayer, the temperature dependence of physical properties, the vesicle dynamics, and the POPC/DPPC two-component membrane dynamics using the parameterized CG lipid model were also explored.

While the CG model membranes, both for POPC and DPPC lipids, could reproduce physical properties estimated from experiments or all-atom models at 30 °C, it did not correctly reproduce the gel-liquid phase transition temperature. Probably, a finer tuning of the parameters targeting the phase transition temperature for each target phospholipid might be necessary. These refinements are left for future studies.

There are two more advantages in the current five beads representations in addition to producing a mapping compatible with the Cα representation of proteins. First, compared to the three-bead representation, the five-bead representation approximates the ratio between the

width and the height of a lipid molecule more faithfully. This is of great importance for correctly reproduce the membrane thickness and the area per lipid of membranes. Second, one can naturally assign the negatively charged phosphate group to the H2 bead. This will help to incorporate electrostatic interactions when developing lipid-protein interactions.

Chapter 3

Modeling lipid-protein interactions for coarse-grained lipid and Cα protein models

3.1 Membrane proteins

Cell membranes are complex heterogeneous structures consisting primarily of a large number of membrane proteins and a diverse array of lipid molecules (**70**). Approximately one-fourth of all proteins encoded in the genome correspond to membrane proteins (**71**). Additionally, these proteins are involved in a variety of biological processes, including cell signaling (**72-74**), molecule transport (**75,76**), and energy generation (**77,78**), and are one of the most frequently targeted classes for drugs (**79,80**). Also, the number of membrane protein structures registered in databases has increased significantly in recent years (**81,82**), allowing their study by computational methods such as molecular dynamics (MD) simulations, which enable permit to investigate spatiotemporal information not accessible through other techniques (**83-85**). However, as the size of the device increases, the computational cost of performing MD simulations rapidly increases, reducing the duration and time scales that can be achieved in a reasonable amount of time. This is important since cell membrane systems often need extended periods of equilibration (**86**). One possibility is to use coarse-grained (CG) models, in which multiple atoms are clustered into a single CG particle, effectively decreasing the computational cost by reducing the total number of degrees of freedom in the system (**87-89**). By combining CG models, long simulations of complex membrane systems can be

achieved (**90**). However, it is necessary to characterize the interactions between these CG particles appropriately.

In the previous chapter, by extending a minimal CG lipid model (**52**), a new CG implicit solvent lipid force field named iSoLF (**91**) was developed. However, the lipid-protein interactions were not explicitly addressed. Therefore, in this chapter, a lipid-protein interaction model specialized for combining the CG lipid force field, iSoLF, and the CG protein force field, AICG2+ (**92**), will be presented.

Lipid-protein interactions operate at different levels. On the one hand, non-specific interactions represent the partitioning preference of each protein amino acid between water and lipid environments (**93,94**). For describing these non-specific interactions, both hydrophobicity scales and solvent accessible surface areas (SASA) have been largely used (**95-97**). In CG models, these scales can be used as targets for either parameterizing or testing parameterizations of lipid-protein interactions, as in the well-known MARTINI force field (**98-100**). On the other hand, specific interactions between lipid atoms and local atomic groups of membrane proteins are often related to specific biological functions (**101-102**). These specific interactions have been studied experimentally by X-ray crystallography (**103**), fluorescent methods (**104**), cryo-electron microscopy (**105**), and nuclear magnetic resonance (**106**). While all-atom models are required to reproduce these specific interactions, CG models can be empirically tuned to represent them.

The lipid-protein interactions presented in this chapter are obtained by parameterizing a modified Lennard-Jones energy function capable of representing the hydrophobic and hydrophilic interactions between lipid CG particles and protein amino acids in implicit solvent

CG models. The parameters for these interactions are tuned against the experimental transfer free energy of each amino acid from aqueous solution to the hydrophobic layer of membrane and the theoretical estimate of the free energy profile normal to the membrane surface. Using the tilt angle and the distance of each protein to the center of the membrane, the placement of different proteins is evaluated and compared with reference structures obtained from the Orientation of Proteins in Membranes (OPM) database (**107**).

3.2 Coarse-grained model of lipids and proteins

In this chapter, a lipid and a protein CG force field will be combined. Here, the iSoLF (**91**) model is used, and each lipid is represented as a molecule with five CG particles, in which two particles correspond to hydrophilic heads (H1 and H2) and the three particles represent hydrophobic tails (T1, T2, and T3) **(Fig. 3.1a).** The iSoLF force field is currently parametrized for the two lipids, POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) and DPPC (1,2-dipalmitoyl-sn-glycero-3-phosphatidylcholine). For proteins, each amino acid is approximated as one CG particle placed at its Cα position, employing the AICG2+ force field (**Figure 3.a**). The AICG2+ force field is structure-based. That is, potential parameters are constructed based on the atomic interactions of reference (native) structures and are calibrated to match the fluctuations around the native state (See the original articles for more details (**92**)).

The five-beads representation of a phospholipid and one bead representation of an amino acid in proteins have about the same resolution. Specifically, one amino acid contains on average 8.4 ± 2.4 non-hydrogen atoms (the mean and the standard deviation from the 20 amino acids), whereas each CG particle in iSoLF represents, on average, 10.4 non-hydrogen

atoms for POPC and 10.0 for DPPC. Having compatible resolutions between models is considered to be advantageous when combining them.

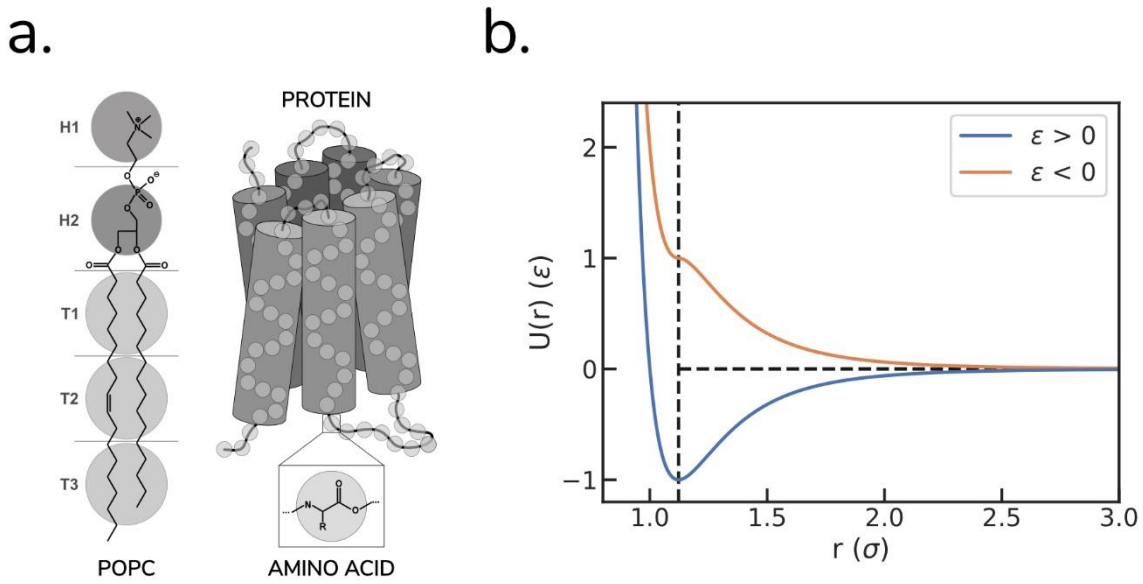

Figure 3.1. Coarse-grained (CG) modeling of lipid-protein interactions. (a) Mapping for the POPC lipid and a representative transmembrane protein into CG beads (b) Modified Lennard-Jones potential function used for the lipid-protein interactions. The force coefficient can be either positive or negative.

3.3 Lipid-protein interactions

Now, the modeling of the lipid-protein interactions (LPI) with a modified Lennard-Jones potential is presented, which is essentially the same as that introduced by Kim and Hummer (**108**). The energy function (V_{LPI}) is composed of a short-range repulsive term (V_{rep}) and a middle-range hydrophobic-hydrophilic term (V_{HP}),

$$V_{LPI} = V_{rep} + V_{HP} \tag{3.1}$$

The repulsive term is modeled by the Weeks-Chandler-Andersen potential,

$$V_{rep} = \sum_{i \in lipids, j \in proteins}^{n_{LPI-pairs}} \begin{cases} 4\varepsilon_{rep,ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} + \frac{1}{4}\right], & r_{ij} \le \sqrt[6]{2}\sigma_{ij} \\ 0, & r_{ij} > \sqrt[6]{2}\sigma_{ij} \end{cases} \quad (3.2)$$

with σ_{ij} representing the repulsive range, $\varepsilon_{rep,ij}$ representing the energy scaling factor for the interaction between the *i*-th lipid bead and the *j*-th protein residue, and $n_{LPI-pairs}$ being equal to the total number of lipid-residue pairs of CG beads. The values of $\varepsilon_{rep,ij}$ and σ_{ij} are defined with the following combination rules,

$$\varepsilon_{rep,ij} = \sqrt{\varepsilon_{rep,i}\varepsilon_{rep,j}} \quad (3.3)$$

$$\sigma_{ij} = \sqrt[6]{\frac{\sigma_i^6 + \sigma_j^6}{2}} \quad (3.4)$$

where the $\varepsilon_{rep,i}$ ($\varepsilon_{rep,j}$) and σ_i (σ_j) values are the corresponding coefficients in the iSoLF (AICG2+) force field. The combination rule in Eq. (4) makes σ_{ij} closer to the larger value of σ_i and σ_j , which is preferred to the arithmetic mean when the two particles' sizes are markedly different (**109**). In fact, as it will be discussed later, this combination rule was found to be important.

The hydrophobic-hydrophilic potential, V_{HP}, is defined as

$$V_{HP} = \sum_{i \in lipids, j \in proteins}^{n_{LPI-pairs}} \begin{cases} -\varepsilon_{HP,ij}, & r_{ij} \le \sqrt[6]{2}\sigma_{ij} \\ 4\varepsilon_{HP,ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], & r_{ij} > \sqrt[6]{2}\sigma_{ij} \end{cases} \quad (3.5)$$

where σ_{ij} is the same as above in Eq. (**3.4**), but the energy scale parameter $\varepsilon_{HP,ij}$ is different from Eq. (**3.3**). $\varepsilon_{HP,ij}$ takes a positive or negative value when the middle-range interaction is attractive or repulsive, respectively (**Figure 3.b**). Since there are five CG particles per one lipid and 20 kinds of natural amino acids, there can be as many as 5 x 20 = 100 energy scale

parameters. By a suitable parameterization of these $\varepsilon_{HP,ij}$, the hydrophobic or hydrophilic nature of amino acids can be approximated inside lipidic environments.

3.4 Coarse-grained molecular dynamics simulations

All the CGMD simulations in this work were performed with a modified version of CafeMol v3.1 (**53**) and the standard underdamped Langevin equation. The mass of each CG bead was set equal to the sum of the masses of atoms it represents, and the friction coefficient equal to 0.8435 1/CafeMol-time (one CafeMol-time unit approximately corresponds to $\sim 49 fs$, although this mapping cannot be used to interpret the time-scale of the large-scale dynamics due to the use of low-friction coefficient, among other reasons). The default dynamics setup of CafeMol was used, a temperature of 303K, the NVT ensemble, and periodic boundary conditions.

For the umbrella sampling simulations, a system composed of 128 POPC lipids arranged in a square bilayer (64 in each leaflet) and one amino acid placed at a different position along the positive z-axis with respect to the middle of the membrane was prepared. The simulation box dimensions were set to $64\text{Å} \times 64\text{Å} \times 200\text{Å}$ and the lipid bilayer was placed in the center, perpendicular to the z-axis. The spring constant for the harmonic bias potential was $1\, kcal/\text{Å} \cdot mol$ and the umbrellas covered distances from 0 to 40Å from the center of the bilayer at constant intervals of 1Å. The simulations consisted of 6×10^6 MD steps, of which the first 1×10^6 MD steps were discarded before calculating free energies using Grossfield's implementation (**110**) of the Weighted Histogram Analysis Method (**111**) (WHAM) v2.0.10.1.

For the test simulations, a system composed of 512 POPC lipids arranged in a square bilayer (256 in each leaflet) and one test target protein was prepared. The box dimensions for these simulations were $128\text{Å} \times 128\text{Å} \times 200\text{Å}$, and as in the umbrella sampling simulations, the membrane was placed at the center of the box. For the transmembrane proteins, a lipid bilayer was superimposed on the structures obtained from the OPM database, and all the lipids molecules containing at least one bead at a distance less than or equal to 3.5Å of any of the protein beads were removed. In the case of peripheral and globular proteins, they were placed at a distance of at least 5Å above the lipid membrane.

3.5 Parameter tuning

The energy scale coefficients for the hydrophobic-hydrophilic interaction $\varepsilon_{HP,ij}$ between the *i*-th CG lipid bead and the *j*-th amino acid were tuned by comparing the free energy at $z = 0\text{Å}$ and $z = 20\text{Å}$ relative to the aqueous environment ($z = \infty$). Here, z is the coordinate normal to the membrane surface, with $z = 0\text{Å}$ corresponding to the center of the membrane. A hybrid parameterization approach was employed, using experimental and theoretical data as target values for adjusting the interaction strength at 0 and 20Å, respectively.

During the tuning process, since each free energy value could be either below or above the target value, the bisection method by defining a simple difference cost function was applied. It is worth noticing that near the target values, the statistical error in the free energy calculation from CGMD simulations could make the bisection algorithm failed. Thus, for the final three iteration steps, the choice of the next iteration point was manually supervised.

3.6 Proteins

Tested proteins were categorized into three groups: 1) The membrane-spanning proteins with either α helices (the first three) or β sheets (the last two); WALP19 (protein data bank id, 2lcn), rhodopsin (5ax0), Zrt/Irt-like protein zinc transporter ZIP (5tsa), the outer membrane domain of intimin (4e1s), and the outer membrane protein OprG (2x27). 2) The water-soluble globular proteins; myoglobin (2spl), pepsin (1psn), and calmodulin (1rfj). 3) The peripheral and other proteins; acutohaemolysin (1mc2) and crambin (1ejg).

3.7 Property calculation

To describe the orientation and positioning, calculation of the projection in the z-axis (z) of the vector joining the center of mass (COM) of the membrane with the COM of the protein was performed, termed insertion depth in this study, and the tilt angle (θ). Even though the tilt angle is a simple concept, there are multiple definitions in the literature **(112-113**). In this work, it is defined as the angle between the z-axis perpendicular to the membrane and the vector joining two reference groups defined differently for each protein type. For the α helical transmembrane proteins, the centers of mass of the first three and last three Cα were set as the reference groups of each α helix, whereas for β sheet transmembrane proteins, only the first and last Cα's of each β strand were used. The mean of the tilt angles of all the transmembrane α helices or β strands were calculated. For the other proteins, the tile angle is not very important. While there is no unique way of defining the reference groups, two Cα's positioned at two opposite sites were used. Those atoms were selected as follows: Ser3 and Lys50 for myoglobin, Ser62 and Gly201 for pepsin, Asp24 and Ala102 for calmodulin, Lys1087 and Ser1116 for

acutohaemolysin, and Pro19 and Gly42 for crambin. Here, each number represents the residue ID in its corresponding PDB file.

3.8 Model parameterization tuning

Here, the parameters in the lipid-protein interaction potential Eq. (**3.1**) are determined. First, the parameters in the repulsive potential (Eq. **3.2**) are defined by simple combination rules (Eqs. **3.3** and **3.4**), and thus no further parameter tuning is necessary. On the other hand, in the hydrophobic-hydrophilic potential (Eq. **3.5**), the energy scale parameters, $\varepsilon_{HP,ij}$, need to be determined. These parameters play a central role in the model and thus need to be tuned carefully. In general, since each $\varepsilon_{HP,ij}$ depends on both the type of the lipid bead i, and the type of amino acid j, we can have a total of 100 different parameters for the interactions between the 20 natural amino acids and one lipid molecule (which in the model is represented with 5 beads). To decrease the parameterization complexity, in this study, it was assumed that the three lipid tail beads (T1, T2, and T3) share the same value for $\varepsilon_{HP,ij}$, reducing the number of parameters to 60.

Next, the reference data for tuning the parameters were selected. A hydrophobicity scale for the 20 amino acids as a bottom line for the interaction was used. Hydrophobicity scales are suitable targets because they capture the partition preferences of amino acids between water and lipid-like solvents. In a recent survey (**114**) of 98 hydrophobicity scales that assessed the potential for the separation of protein/peptide pools into different classes, the classic scale by Engelman et al. (**115**) showed one of the top performances (scale id 28 in Table S6 of the survey). This hydrophobicity scale is based on experimentally determined transfer free energies that are theoretically adjusted to account for transmembrane α-helices. Since this scale aims to

predict transmembrane regions from its amino acid sequence, it is inherently suitable for the current purpose. However, these hydrophobicity scales could only be used to compare the free energy difference of each amino acid between aqueous solutions and the hydrophobic layer of the lipid membrane, not reflecting the interactions of amino acids with the hydrophilic layer of the lipid membrane where the lipid head beads (H1 and H2) mostly reside. Therefore, to complement the experiment-based data, the free energy profiles of each amino acid normal to the membrane surface calculated by all-atom MD simulations (**116**) were also used.

The MD-based free energy profiles show free energy minima or maxima at the hydrophilic layer, representing ideal target values for reproducing with the lipid-protein interaction. Interestingly, these free energy profiles also exhibit, for some amino acids, a characteristic behavior near the interface between hydrophobic and hydrophilic layers, but in this work, due to their over-complexity, it was ignored. In principle, the MD-based free energy at the center of the lipid bilayer could also be used in place of the experimental hydrophobicity scales. However, for better accuracy and robustness, it was decided to use the experiment-based scale, i.e., Engelman's scale, for this purpose. Lastly, for the residues which did not have available theoretical free energy profiles, the curve for the closest amino acid in the hydrophobicity scale was used as a proxy.

In summary, Engelman's hydrophobicity scale was used as target values for the free energy at the center of the lipid bilayer ($z = 0$ Å) relative to an aqueous solution and the all-atom MD free energy estimate at 20 Å from the center of the membrane surface ($z = 20$ Å) as the reference data in the parameterization (red crosses in Figure 2).

To fit the model to the target data, the free energy profile of each amino acid between water and lipid environments was calculated by performing umbrella sampling simulations. In preliminary trials, it was found that a local minimum at the middle of the membrane independently from the chosen value of $\varepsilon_{HP,ij}$ when using the arithmetic mean as a combination rule for the repulsive range, $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ (**Fig. A7**). Representing lipids as single-chain molecules requires tail beads to be significantly large compared to amino acid beads. This difference in size produced a small cavity between T3 beads where hydrophilic amino acids were stabilized. In contrast, with Eq. (**3.4**) as a combination rule, σ_{ij} becomes closer to the largest of the two σ's, effectively destabilizing the artificial local minimum due to an increase in the size of the amino acid beads.

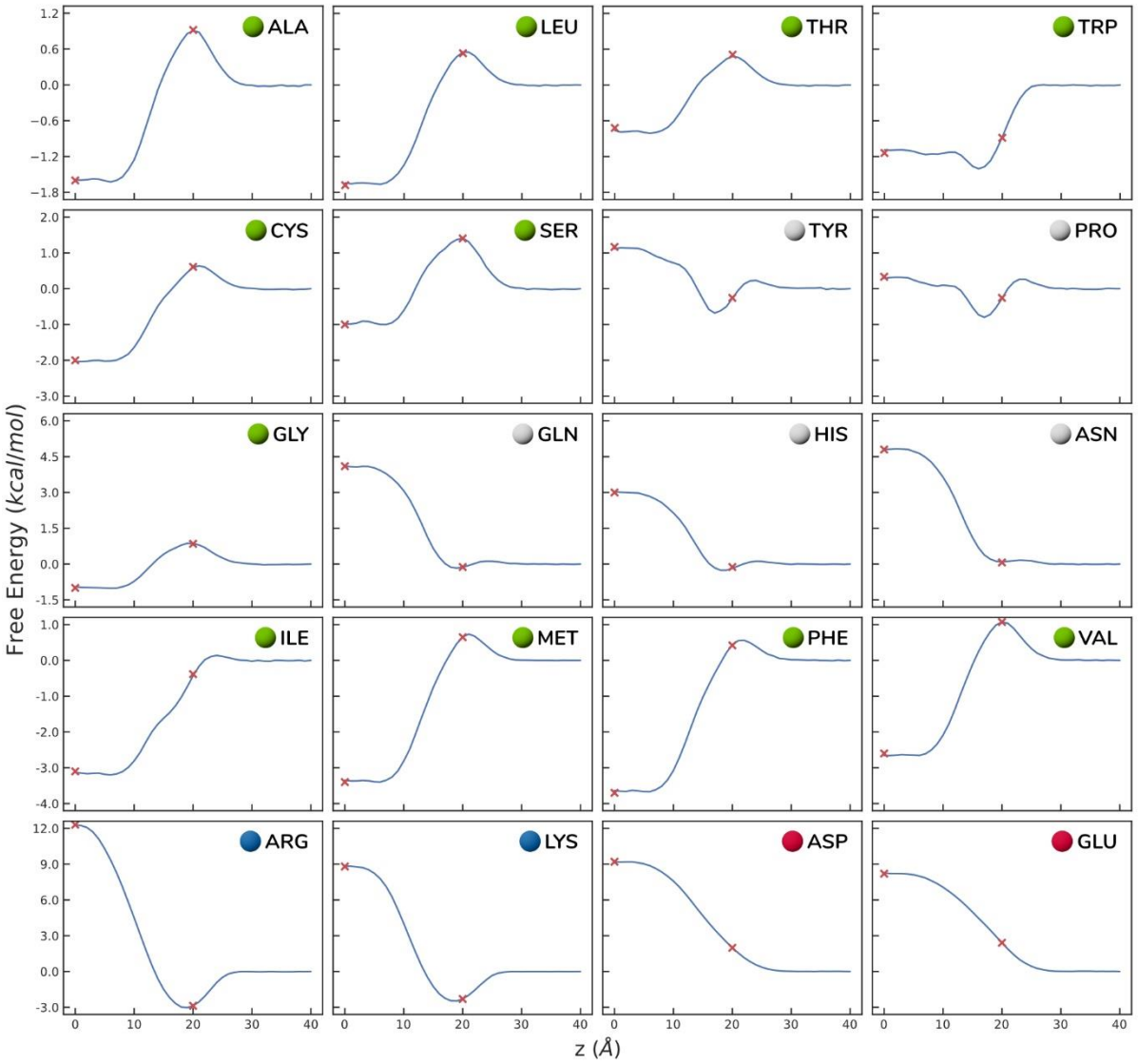

Figure 3.2. Free energy profiles for the 20 amino acids along the z-axis normal to the POPC membrane surface computed by CGMD with the tuned parameters. Here, $z = 0$ and $z \approx 25$Å correspond to the center of the lipid bilayer and the surface of the membrane, respectively. Red crosses mark the target free energy values used for tuning. Based on the free energy profile, each amino acid is classified either as hydrophobic (green), polar (white), positively charged (blue), or negatively charged (red).

RESIDUE	H1	H2	T1, T2, T3
ALA	0.000	-0.995	0.593
ARG	0.000	2.581	-0.680
ASN	0.000	0.571	0.030
ASP	0.000	-0.735	-0.371
CYS	0.000	-0.510	0.626
GLN	0.000	0.650	0.093

GLU	0.000	-1.515	-0.280
GLY	0.000	-0.995	0.539
HIS	0.000	0.579	0.207
ILE	0.000	0.130	0.723
LEU	0.000	-1.050	0.692
LYS	0.000	2.030	-0.335
MET	0.000	-0.760	0.745
PHE	0.000	-0.588	0.766
PRO	0.000	0.302	0.436
SER	0.000	-0.740	0.506
THR	0.000	-0.780	0.560
TRP	0.000	0.894	0.616
TYR	0.000	0.330	0.394
VAL	0.000	-1.148	0.677
N-terminal	0.000	1.000	-0.500
C-terminal	0.000	-1.000	-0.500

Table 3.1. Tuned energy scale parameters $\varepsilon_{\mathrm{HP,ij}}$ for the hydrophobic-hydrophilic interactions between a lipid particle and a protein residue. All values are in kcal/mol.

In a preliminary examination, it was observed that varying $\varepsilon_{HP,ij}$ for the lipid tail bead changed the free energy value at $z = 0$ Å, but not $z = 20$ Å, whereas $\varepsilon_{HP,ij}$ for the lipid head bead changed the free energy value at $z = 20$ Å, but not $z = 0$ Å (**Fig. A8**). This pseudo-independence allowed to set up an initial guess on the target parameters by performing a linear fitting for the head and tail separately. Furthermore, it was observed that the free energy at $z = 20$ Å could be modeled solely by the interaction with H2, and thus, to minimize the complexity, it was decided to turn off the interaction of H1 (this removal can be a major point for future improvement since at this stage proteins cannot recognize lipids with different head groups).

Starting from an initial guess, the free energy profiles were iteratively calculated, comparing their values at $z = 0$ Å and 20 Å, and updating the corresponding values of $\varepsilon_{HP,ij}$

until the difference in free energy at each position was smaller than 0.1 kcal/mol. The parameters for all the residues converged after around ten iterations and are presented in Table 3.1. Figure 3.2 shows the calculated free energy curves for the 20 residues, together with the reference data points used for the fitting (red crosses).

In Table 1, it was found that the tail beads (T1-T3) have negative values (namely, the repulsive interaction) of $\varepsilon_{HP,ij}$ only with the charged amino acids (Arg, Asp, Glu, and Lys), but not the polar ones (Asn, Gln, His, Pro, and Tyr), contrary to what was initially expected. Not surprisingly, the tail beads have positive $\varepsilon_{HP,ij}$ values (the attractive interaction) for all the hydrophobic amino acids. Since the H2 particle contains the phosphate group, $\varepsilon_{HP,ij}$ of the H2 particle is positive and large for the positively charged amino acids (Arg and Lys) and is negative for the negatively charged amino acids (Asp and Glu). The $\varepsilon_{HP,ij}$ values of the H2 particle with other amino acids are either positive or negative.

Finally, an extra set of parameters for representing the interactions between the termini of the protein with the lipid was added. By treating the N- and C-terminus as positively and negatively charged amino acids, respectively, small transmembrane α-helical peptides could be stabilized, and their embedding inside the hydrophobic layer was prevented. This will be described more in detail in the next section.

3.9 Transmembrane proteins

With the lipid-protein interaction parameterized, the behavior of proteins interacting with the membrane was tested. For this purpose, CGMD simulations of three classes of proteins were performed: transmembrane proteins, water-soluble proteins, and peripheral/others

proteins. This section will describe the first class, transmembrane proteins, composed of three α-helix containing proteins, WALP, Rhodopsin, and ZIP, and two β-barrel containing proteins, intimin's transmembrane domain and the outer membrane protein OprG.

WALP is a small peptide with only one transmembrane α-helix. It presents two tryptophan residues at each end, an alternating sequence of alanine and leucine residues in the middle section, and is a model peptide for representing the transmembrane α helix **(117**). In the developmental stage, when CGMD simulations were performed without the special treatment for the N- and C- terminal residues, it was found that the whole peptide was getting embedded into the central hydrophobic layer of the membrane despite its initial conformation (**Fig. A9**). The two terminal-residues, Gly1 and Ala19, are hydrophobic in the parameterization (Figure 3.2). Thus, if their corresponding parameters are assigned, all the amino acids in the peptide (except Pro10) are hydrophobic. Not surprisingly, the most favorable position ended up being the center of the membrane. This motivated the introduction of the special parameters for N- and C-terminal residues. In principle, adding this special treatment to the termini would require 40 extra parameters. However, in order to avoid increasing the complexity of the model, values for the parameters similar to the ones for positively and negatively charged residues were set. Equipped with these parameters for the termini, it was found that both ends of WALP stayed in the hydrophilic layer of the membrane, consistent with the previous knowledge. This phenomenon was not observed for other proteins with multiple transmembrane α helices, probably because the contribution of the terminals was smaller than the overall energy contribution from hydrophilic residues in the loop regions between helices.

Then, WALP positioning in the membrane (**Fig. 3.3a, Table 3.2**) was measured. In the CG MD ensemble, WALP's tilt angle and insertion depth were $15 \pm 7.3°$ and 1 ± 1.7Å,

respectively, compared to 13.9° and 2.9Å obtained from the reference structure. From these values, it was concluded that the WALP positioning in the CGMD simulations agrees with the suggested orientation from the OPM database.

	Reference		Simulation	
Protein	θ (°)	z(Å)	θ(°)	z(Å)
WALP	13.9	3.6	15 ± 7.3	1 ± 1.7
Rhodopsin	13.8	0.6	13 ± 1.6	1 ± 1.3
ZIP	25.4	-1.1	45 ± 5.5	-4 ± 1.5
ZIP-charged*			$24 \pm 2.7°$	-4 ± 1.4Å
Intimin	38.9	2.2	38 ± 0.6	5 ± 1.7
OprG	38.9	11.5	39 ± 1.3	14 ± 1.7

Table 3.2. Summary of the tilt angle θ (°) and insertion depth z(Å) for the transmembrane proteins. The reference values were calculated from the structures obtained in the OPM database, whereas the simulation values were calculated from the CGMD samples. (*) ZIP-charged corresponds to the results of simulations that mimicked the divalent metal cation binding by a double mutation to positively charged residues.

In **Figure 3.3a**, a time series of the simulation for WALP is shown. A sudden transition was observed in the configuration of WALP at around 10^7 MD steps, where the C-terminus portion of the protein went inside the membrane, adopting a kinked conformation that rapidly returned into the original configuration (**Figure 3.4**). This transition is consistent with what was reported in a previous all-atom MD simulation by Ward et al. (**117**), where they observed a kinked conformation during WALP folding inside the membrane. Notably, this type of rare transition was not observed for the N-terminus. While both N- and C-terminal residues are unstable in the hydrophobic layer of the membrane, their interactions with the hydrophilic layer are distinct. The C-terminal residue has a negative charge, resulting in a repulsive interaction with the lipid head particle H2 and making the transmembrane placement only marginally

stable. On the other hand, the positively charged N-terminal residue is attracted to the H2 particle and tends to keeps its position in the hydrophilic layer.

The second protein is the G-protein-coupled-receptor, rhodopsin, which consists of seven transmembrane α-helices, making them good candidates after testing a single membrane-spanning α-helical peptide. For rhodopsin, a tilt angle of $13 \pm 1.6°$ and an insertion depth of 1 ± 1.3Å were calculated. These values have good agreement with the reference configuration (**Fig. 3.3b, Table 3.2**).

Third, a zinc-regulated, iron-regulated transporter-like protein (ZIP) was examined. With the same setup, markedly tilted configurations were observed, with the tilt angle $45 \pm 5.5°$ and the insertion depth -4 ± 1.5Å, which deviate considerably from $25.4°$ and -1.1Å in the reference configuration (**Figure 3.3c, Table 3.2**). By examining ZIP's reference structure, two points that could have affected the result were noticed. First, the reference structure lacks three flexible loops: one joining α-helices 3 and 4, one joining α-helices 7 and 8, and one at the N-terminal position. These loops contain strongly hydrophilic Arg, Gln, and His residues near their beginning and their end. The simulations did not include these residues in the missing loops. If they were explicitly included, they would have preferred to be positioned either at the membrane surface or outside the membrane. Furthermore, ZIP has multiple binding sites for heavy metal ions (Zn^{2+} or Cd^{2+}), including one between α-helices 2 and 3, which was embedded inside the membrane in the above simulation (**118**). The lack of these loops and a chelated metal ion in the CGMD simulation might explain why the deviation is greater than in the case of rhodopsin, in which all the loops joining the transmembrane α-helices are present, and there are no heavy metal-binding sites. To test this hypothesis, the two missing flexible loops mentioned above and the last 5 residues of the missing N-terminal flexible loop were

added using Modeller v10.1 (**119**). Additionally, a bound divalent ion was mimicked via an *in silico* double mutation to the residues in the coordination site, G120K and G122K (**Fig. A10**). With this modified setup, the tilt angle of $24 \pm 2.7°$ and the insertion depth of -4 ± 1.4Å were obtained, agreeing with the reference configuration calculated from the OPM database. This might indicate some role for the flexible loops joining transmembrane helices and the bound ions contributing to the overall orientation of this protein.

The last two transmembrane proteins are the β-barrel containing proteins, intimin's transmembrane domain, and the outer membrane protein OprG. It was found that despite using a parameterization suited to α helical transmembrane proteins, the tilt angle and the insertion depth only differ slightly from the values calculated from the reference structures (**Figure 3.3(d)(e), Table 3.2**). Since the SASA of each residue is different depending on the secondary structure, β-sheet forming residues have different contributions to the stabilizing effect in the lipid-protein interaction. In an implicit solvent model, the lipid-protein interaction becomes the main driving force acting on the orientation. Thus, having a more precise contribution for each residue considering its local spatial configuration might help reduce the difference observed for this class of proteins.

3.10 Water-soluble and peripheral/other proteins

Next, CGMD simulations of five proteins classified either as water-soluble globular proteins or peripheral proteins were performed. These were important targets because a parameter set that not only stabilizes membrane proteins inside the bilayer but also destabilizes non-membrane proteins in the membrane relative to an aqueous solution is desired.

The first sample of a water-soluble protein is myoglobin, an oxygen-binding protein found in skeletal muscle tissue (**120**). It was found that this protein presented a very weak interaction with the membrane (**Figure 3.5a**) but did not stay on the membrane surface despite having hydrophilic residues on its surface, thus showing the expected behavior. The second water-soluble protein examined is pepsin, an endopeptidase that digests proteins into smaller peptides. Compared to myoglobin, pepsin presented more hydrophobic residues on its surface, making it more suitable for interacting with lipid tails. However, it also presented several negatively charged residues distributed over the surface. These residues created a high energy barrier that prevented its insertion into the membrane (**Figure 3.5b**). The final of the water-soluble proteins was the calcium-modulated protein, calmodulin. This protein presents many α-helices in its structure; however, they were formed mainly from hydrophilic residues. It was observed that it weakly interacted with the surface of the membrane (**Figure 3.5c**). Again, the distribution of these hydrophilic residues over the protein structure destabilizes a possible membrane-bound state. For these three proteins, it was also observed that when weakly interacting with the membrane, they explore different orientations. However, their topology and amino acid sequence composition were not designed to interact favorably with the lipid membrane, as expected.

The next protein that was tested is acutohaemolysin, a phospholipase from the venom of the snake *Agkistrodum acutus,* which is categorized as a peripheral protein, suggested to be tethered on the membrane surface. In the CGMD simulations, it was observed that this protein rapidly approached the membrane surface, where it remained interacting weakly (**Figure 3.5d**). While interacting, it explored different orientations for $\sim 0.2 \times 10^6$ MD steps until it found its preferred orientation and bound to the membrane. It is important to notice that the protein surface in contact with the membrane mainly presented either hydrophobic or positively

charged residues. Positively charged residues have a favorable interaction with the lipid phosphates, represented in the H2 beads in the model. The hydrophobic residues interact with tail beads. These interactions together stabilized the binding on the membrane surface. Notably, it was also observed transient dissociations of the protein from the membrane after several million simulation steps. However, after each dissociation, acutohaemolysin quickly recovered its preferred orientation ($\sim 60°$ in the current definition) and bound back to the membrane.

The last protein of this group is crambin, a small protein of *Abyssinian cabbage* that presents a structure that has been well studied (**121**). It was found in the simulations that crambin, similar to acutohaemolysin, quickly interacted with the membrane, exploring different conformations before adopting its membrane-bound state after less than one million MD steps (**Figure 3.5e**). However, contrary to what was expected for a peripheral protein, this surface-bound state was followed by an insertion into the membrane core. That is, crambin positioned itself inside the membrane, adopting an orientation of around $\sim 50°$ with the flexible loop joining its two alpha-helices facing outwards (**Figure 3.6**). The insertion of crambin into the membrane with the loop facing outwards is consistent with a previous study by Ahn et al. **(122**). Lastly, in the simulations, no transition of crambin going outside the membrane could be found during $\sim 40 \times 10^6$ MD steps, suggesting high stability for the inserted conformation.

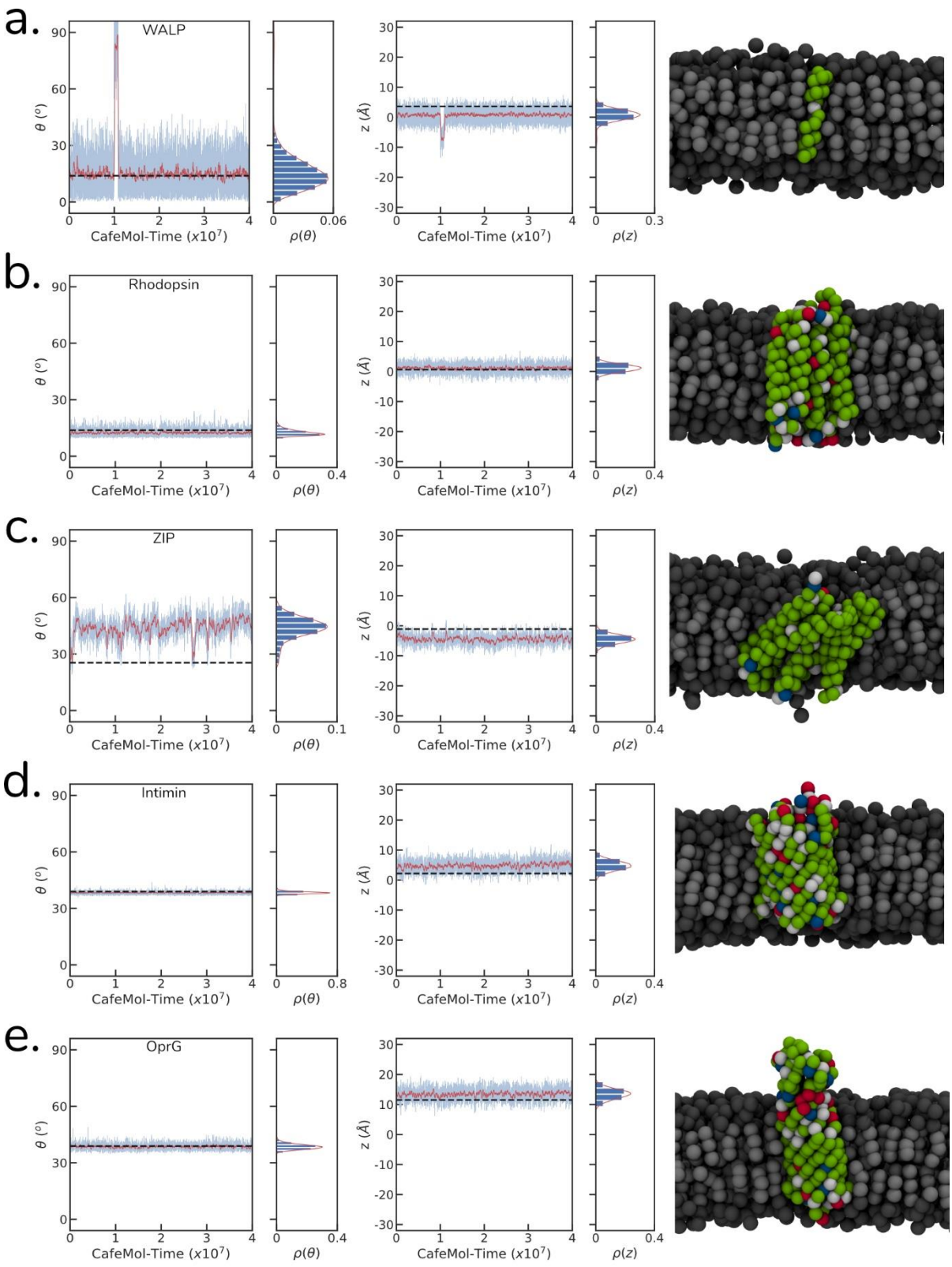

Figure 3.3. Time series for the CGMD simulation of five transmembrane proteins inside the lipid bilayer. The title angles (left column), the insertion depth (central column), and a representative snapshot (right column) are shown for (a) WALP, (b), rhodopsin, (c) zinc-regulated iron-regulated transporter-like protein (ZIP), (d) intimin's transmembrane domain, and (e) the outer membrane protein OprG. In the left and central columns, light blue curves represent the raw data, red curves the moving average over 200 data points, and dashed straight

lines the corresponding values for the configuration calculated from the reference structures from the OPM database. The bar graph on the right side is the histogram of the raw data. In the right column, amino acids of proteins are colored by the scheme defined in Figure 2. The membrane is located roughly in the range of $-25\text{Å} \leq z \leq 25\text{Å}$.

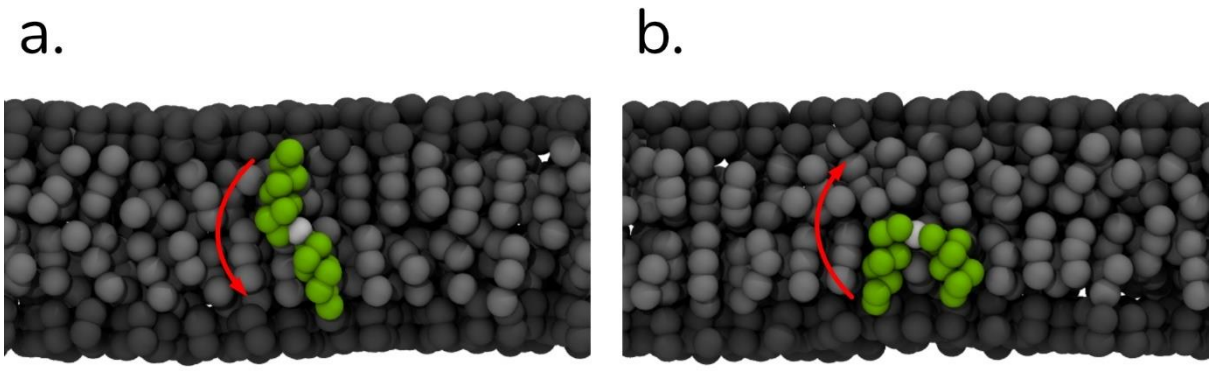

Figure 3.4. The conformational transition of WALP in CGMD simulations. It was observed a rare large-scale transition exemplified in Figure 3(a) after $\sim 1.0 \times 10^7$ CafeMol-time units. (a) A typical configuration before the transition. The C-terminus is at the top here. (b) A kinked configuration after the transition. The C-terminus jumped from the top to the bottom of the hydrophilic layers.

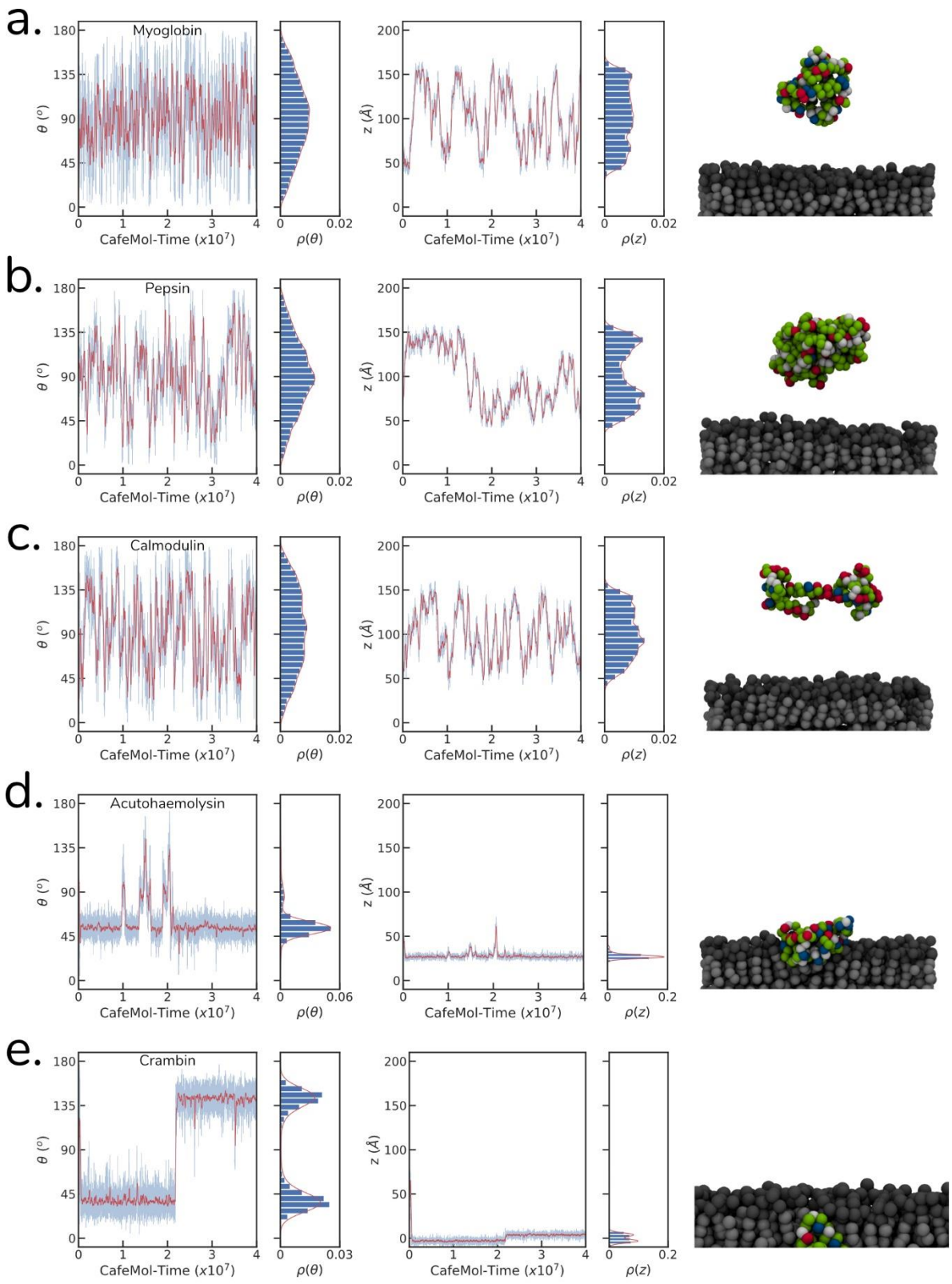

Figure 3.5. Placement of water-soluble, peripheral, and other proteins relative to the lipid bilayer plane in CGMD simulations. The title angles (left column), the insertion depth (central column), and a representative snapshot (right column) are shown for three water-soluble proteins; (a) myoglobin, (b), pepsin, and (c) calmodulin, and two peripheral and other proteins; (d) acutohaemolysin and (e) crambin. In the left and central columns, light blue curves represent the raw data, red curves the moving average over 200 data points. The bar graph on

the right side is the histogram of the raw data. In the right column, amino acids of proteins are colored by the scheme defined in Figure 2. The angle θ (°) is merely defined by the angle between a representative vector that connects two distant amino acids and the z-axis. The membrane is located roughly in the range of $-25\text{Å} \leq z \leq 25\text{Å}$

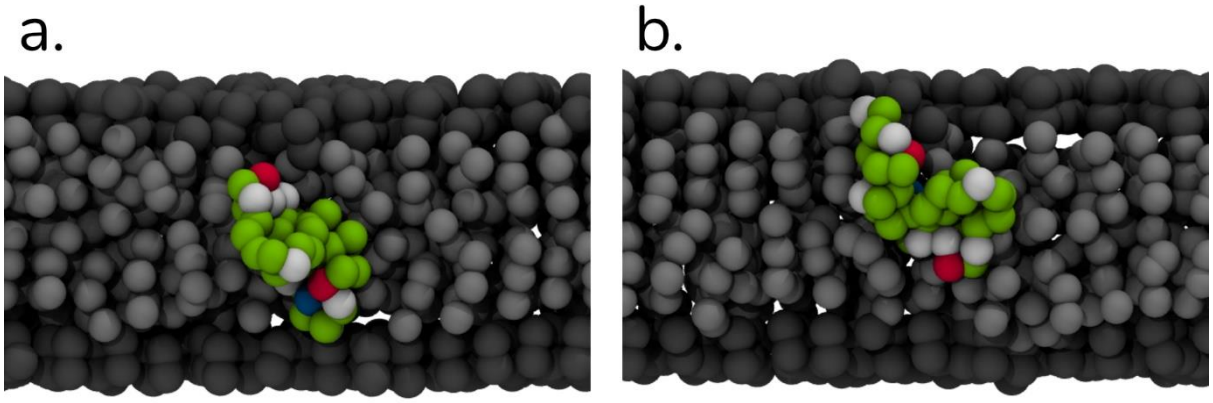

Figure 3.6. The two placements of crambin in CGMD simulations. Two oppositely-oriented placements were observed. The characteristic solvent-exposed residues Gly20 and Thr21 are facing to the bottom (a) and the top (b) of the lipid membrane.

3.11 Discussion and conclusions

In this study, based on two previously developed CG force fields, iSoLF for lipids and AICG2+ for proteins, lipid-protein interaction potential using a modified Lennard-Jones energy function was developed and parameterized. The middle-range hydrophobic-hydrophilic interaction energy scale parameters were tunned by using an experiment-based hydrophobicity scale and an all-atom MD-based free energy profile along the normal to the membrane surface. Additionally, a pair of parameters for the N- and C- terminal residues was also defined. The parametrized lipid-protein interactions were tested for 10 proteins, including five transmembrane proteins, three water-soluble proteins, and two peripheral/other proteins. Overall, the lipid-protein interaction could reproduce the expected behavior for different classes of proteins inside/outside lipid environments.

However, some points need to be further improved. Firstly, the parameterization of the lipid-protein interaction was performed with a hydrophobicity scale optimized for α helical structures. Although the test simulations for β barrel proteins were successful, refining parameters against transmembrane β proteins may increase the accuracy. Secondly, the interface between the lipid heads and the lipid tails is recognized by some amino acids, such as tryptophan, due to their amphipathic nature (**123**). Probably, refining the parameter for the tail residue T1 could mimic these interactions. Lastly, the model would greatly increase its applicability if a specific interaction with the lipid head beads H1 is defined. It was found that the expected behavior for peripheral and globular proteins could be reproduced despite not having a specific interaction between lipid beads H1 and amino acids. However, for example, there are known proteins that recognize sphingomyelin (SM) or lipids containing phosphatidylserine (PS), in which electrostatic interactions play a major role. Representing

these interactions is not trivial due to the nature of the system and the CG representation. The dielectric constant changes rapidly across the membrane and represents a major drawback when calculating the electrostatic interactions among particles (**124**). Additionally, in this model, water is treated implicitly, representing the net effect of the electrostatic interactions in the different energy functions of the force fields. Therefore, electrostatic interactions need to be considered carefully in order to avoid artifacts in the model. All these limitations can be overcome in the next round of the force field development.

Chapter 4

Conclusions

In this work, a new implicit solvent coarse-grained model for the simulation of biological membranes has been developed by extending a previous work from Cooke, Kremer, and Deserno (**52**). This force field was created in need of a lipid model with a resolution suitable for combination with Cα protein models. Thus, this required mapping lipids into five-bead CG molecules. For the parameterization, a hybrid approach was employed. Intramolecular interactions were fitted against statistical distributions from all-atom simulations. In contrast, intermolecular interactions were tuned to reproduce experimental measurements. Despite representing two-tailed phospholipids as single-tailed linear molecules, the lipid model could capture key properties of biological membranes. Namely, the spontaneous formation for membranes and the stabilization vesicle conformations could be observed. On top of that, the correct area-per-lipid and hydrophobic thickness for POPC and DPPC at 30 °C were reproduced, as well as the correct phase behavior.

In the next stage of this work, the developed lipid model was combined with the AICG2+ protein model. For this purpose, a lipid-protein interaction was parameterized in order to represent membrane proteins interacting inside lipidic environments. Since hydrophobicity scales capture well the chemical nature of amino acids inside different environments, they were used as target data for tuning the lipid-protein interaction. Additionally, free energy profiles for the insertion of amino acids into lipid bilayers were also used as complementary data.

Finally, to test the lipid-protein interaction, simulations of various proteins were performed. It was found that in all the cases, the proteins adopted the expected configuration calculated from the OPM database.

Regardless of the ability of the model to successfully simulate biological membrane environments by capturing the hydrophobic effect into its interaction potentials, there are still improvements to be made. First, the phase behavior of lipid membranes at temperatures different from 30 °C is still not well represented. This can be addressed by fine-tuning the lipid-lipid interaction in a more detailed way, assigning specific parameters for the interaction of each pair of beads. In fact, this might also help reproduce the phase separation of multicomponent membranes, an important feature of lipid bilayers. Second, there are no specific interactions between the different amino acids and the lipid H1 head beads at this stage. Representing these interactions is important because there are many proteins for which it is known they recognize specific lipids on the membrane through their characteristic head groups. Even though it is possible to calibrate this interaction to reproduce experimentally calculated protein-membrane binding coefficients, a general parameterization will favor its applicability for cases where such a binding coefficient is not accessible yet. Finally, the last point of improvement is related to the available parameters. Currently, the iSoLF force field has parameters for performing simulations with POPC or DPPC lipids. However, lipid membranes are heterogeneous systems that contain several types of lipids. In order to have a more faithful representation of these systems, more lipids need to be parameterized.

In conclusion, this new force field, despite its current shortcomings, opens the possibility to apply structure-based CG protein models to membrane protein systems, contributing to the MD field.

References

1. M. Karplus and G.A. Petsko, Nature 347, 631 (1990).

2. M. Karplus and R. Lavery, Isr. J. Chem. 54, 1042 (2014).

3. K.Y. Sanbonmatsu and C.S. Tung, J. Struct. Biol. 157, 470 (2007).

4. D.E. Shaw, J.P. Grossman, J.A. Bank, B. Batson, J.A. Butts, J.C. Chao, M.M. Deneroff, R.O. Dror, A. Even, C.H. Fenton, A. Forte, J. Gagliardo, G. Gill, B. Greskamp, C.R. Ho, D.J. Ierardi, L. Iserovich, J.S. Kuskin, R.H. Larson, T. Layman, L.S. Lee, A.K. Lerer, C. Li, D. Killebrew, K.M. Mackenzie, S.Y.H. Mok, M.A. Moraes, R. Mueller, L.J. Nociolo, J.L. Peticolas, T. Quan, D. Ramot, J.K. Salmon, D.P. Scarpazza, U. Ben Schafer, N. Siddique, C.W. Snyder, J. Spengler, P.T.P. Tang, M. Theobald, H. Toma, B. Towles, B. Vitale, S.C. Wang, and C. Young, Int. Conf. High Perform. Comput. Networking, Storage Anal. SC **2015-January**, 41 (2014).

5. J. Jung, C. Kobayashi, K. Kasahara, C. Tan, A. Kuroda, K. Minami, S. Ishiduki, T. Nishiki, H. Inoue, Y. Ishikawa, M. Feig, and Y. Sugita, J. Comput. Chem. **42**, 231 (2021).

6. J.D. Durrant and J.A. McCammon, J. Biol. 9, 23 (2011).

7. R.E. Rudd and J.Q. Broughton, Phys. Rev. B - Condens. Matter Mater. Phys. **58**, R5893 (1998).

8. R.C. Bernardi, M.C.R. Melo, and K. Schulten, Biochim. Biophys. Acta - Gen. Subj. **1850**, 872 (2015).

9. S. Izvekov, A. Violi, and G.A. Voth, J. Phys. Chem. B **109**, 17019 (2005).

10. H. Noguchi, J. Chem. Phys. **134**, (2011).

11. A.J. Sodt and T. Head-Gordon, J. Chem. Phys. **132**, (2010).

12. S.J. Marrink, A.H. De Vries, and A.E. Mark, J. Phys. Chem. B **108**, 750 (2004).

13. I.R. Cooke, K. Kremer, and M. Deserno, Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. 72, 2 (2005).

14. H. Noguchi and M. Takasu, J. Chem. Phys. **115**, 9547 (2001).

15. J.L. Klepeis, K. Lindorff-Larsen, R.O. Dror, and D.E. Shaw, Curr. Opin. Struct. Biol. 19, 120 (2009).

16. M.C. Zwier and L.T. Chong, Curr. Opin. Pharmacol. 10, 745 (2010).

17. D.J. Huggins, P.C. Biggin, M.A. Dämgen, J.W. Essex, S.A. Harris, R.H. Henchman, S. Khalid, A. Kuzmanic, C.A. Laughton, J. Michel, A.J. Mulholland, E. Rosta, M.S.P. Sansom, and MW van der Kamp, Wiley Interdiscip. Rev. Comput. Mol. Sci. 9, 1 (2019).

18. S. Takada, Curr. Opin. Struct. Biol. 22, 130 (2012).

19. S. V. Bennun, M.I. Hoopes, C. Xing, and R. Faller, Chem. Phys. Lipids 159, 59 (2009).

20. S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A.E. Dawid, and A. Kolinski, Chem. Rev. 116, 7898 (2016).

21. J. Kleinjung and F. Fraternali, Curr. Opin. Struct. Biol. 25, 126 (2014).

22. R. Goetz and R. Lipowsky, J. Chem. Phys. 108, 7397 (1998).

23. H. Noguchi and M. Takasu, J. Chem. Phys. 115, 9547 (2001).

24. I.R. Cooke, K. Kremer, and M. Deserno, Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. 72, 2 (2005).

25. S.J. Marrink, A.H. De Vries, and A.E. Mark, J. Phys. Chem. B 108, 750 (2004).

26. W. Shinoda, R. Devane, and M.L. Klein, Mol. Simul. 33, 27 (2007).

27. L. Lu and G.A. Voth, J. Phys. Chem. B 113, 1501 (2009).

28. A.P. Lyubartsev, Eur. Biophys. J. 35, 53 (2005).

29. A.J. Sodt and T. Head-Gordon, J. Chem. Phys. 132, (2010).

30. J.C. Shelley, M.Y. Shelley, R.C. Reeder, S. Bandyopadhyay, P.B. Moore, and M.L. Klein, J. Phys. Chem. B 105, 9785 (2001).

31. E.M. Curtis and C.K. Hall, J. Phys. Chem. B 117, 5019 (2013).

32. E.E. Barrera, E.N. Frigini, R.D. Porasso, and S. Pantano, J. Mol. Model. 23, 2 (2017).

33. H. Koldsø, D. Shorthouse, J. Hélie, and M.S.P. Sansom, PLoS Comput. Biol. 10, (2014).

34. K. Koshiyama and S. Wada, Sci. Rep. 6, 1 (2016).

35. V. Corradi, E. Mendez-Villuendas, H.I. Ingólfsson, R.X. Gu, I. Siuda, M.N. Melo, A. Moussatova, L.J. Degagné, B.I. Sejdiu, G. Singh, T.A. Wassenaar, K. Delgado Magnero, S.J. Marrink, and D.P. Tieleman, ACS Cent. Sci. 4, 709 (2018).

36. M. Xue, L. Cheng, I. Faustino, W. Guo, and S.J. Marrink, Biophys. J. 115, 494 (2018).

37. C. Arnarez, J.J. Uusitalo, M.F. Masman, H.I. Ingólfsson, D.H. De Jong, M.N. Melo, X. Periole, A.H. De Vries, and S.J. Marrink, J. Chem. Theory Comput. 11, 260 (2015).

38. C. Clementi, H. Nymeyer, and J.N. Onuchic, J. Mol. Biol. 298, 937 (2000).

39. T.X. Hoang and M. Cieplak, J. Chem. Phys. 112, 6851 (2000).

40. N. Koga and S. Takada, J. Mol. Biol. 313, 171 (2001).

41. A.R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, and I. Bahar, Biophys. J. 80, 505 (2001).

42. J. Karanicolas and C.L. Brooks, Protein Sci. 11, 2351 (2009).

43. G. Brannigan, P.F. Philips, and F.L.H. Brown, Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. 72, 4 (2005).

44. T. Sintes and A. Baumgärtner, J. Chem. Phys. 106, 5744 (1997).

45. F. Schmid, D. Düchs, O. Lenz, and B. West, Comput. Phys. Commun. 177, 168 (2007).

46. M. Kranenburg, M. Venturoli, and B. Smit, J. Phys. Chem. B 107, 11491 (2003).

47. A. Srivastava and G.A. Voth, J. Chem. Theory Comput. 9, 750 (2013).

48. A.J. Pak, T. Dannenhoffer-Lafage, J.J. Madsen, and G.A. Voth, J. Chem. Theory Comput. 15, 2087 (2019).

49. J.D. Revalee, M. Laradji, and PB Sunil Kumar, J. Chem. Phys. 128, (2008).

50. S.J. Attwood, Y. Choi, and Z. Leonenko, Int. J. Mol. Sci. 14, 3514 (2013).

51. S. Leekumjorn and A.K. Sum, J. Phys. Chem. B 111, 6026 (2007).

52. I.R. Cooke and M. Deserno, J. Chem. Phys. 123, (2005).

53. H. Kenzaki, N. Koga, N. Hori, R. Kanada, W. Li, K.I. Okazaki, X.Q. Yao, and S. Takada, J. Chem. Theory Comput. 7, 1979 (2011).

54. X. Gao, J. Fang, and H. Wang, J. Chem. Phys. 144, (2016).

55. Extreme learning. http://extremelearning.com.au/evenly-distributing-points-on-a-sphere/ (accessed Oct 1, 2019)

56. M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, and E. Lindah, SoftwareX 1–2, 19 (2015).

57. J.P.M. Jämbeck and A.P. Lyubartsev, J. Phys. Chem. B 116, 3164 (2012).

58. J.P.M. Jämbeck and A.P. Lyubartsev, J. Chem. Theory Comput. 8, 2938 (2012).

59. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, J. Chem. Phys. 79, 926 (1983).

60. SLipids. http://www.fos.su.se/~sasha/SLipids/Downloads.html (accessed Jun 1, 2019)

61. W.J. Allen, J.A. Lemkul, and D.R. Bevan, J. Comput. Chem. 32, 174 (2012).

62. G. Pranami and M.H. Lamm, J. Chem. Theory Comput. 11, 4586 (2015).

63. V. Agrawal, G. Arya, and J. Oswald, Macromolecules 47, 3378 (2014).

64. N. Kučerka, M.P. Nieh, and J. Katsaras, Biochim. Biophys. Acta - Biomembr. 1808, 2761 (2011).

65. J.A. Nelder and R. Mead, Comput. J. 7, 308 (1965).

66. T. Bereau, M. Hu, P. Diggins, and M. Deserno, 771, 1 (2011).

67. M. Kranenburg and B. Smit, J. Phys. Chem. B 109, 6553 (2005).

68. Q. Waheed, R. Tjörnhammar, and O. Edholm, Biophys. J. 103, 2125 (2012).

69. Y. Wang, P. Gkeka, J.E. Fuchs, K.R. Liedl, and Z. Cournia, Biochim. Biophys. Acta - Biomembr. 1858, 2846 (2016).

70. X. Cheng and J.C. Smith, Chem. Rev. **119**, 5849 (2019).

71. L. Fagerberg, K. Jonasson, G. Von Heijne, M. Uhlén, and L. Berglund, Proteomics **10**, 1141 (2010).

72. X. Cui and Z. Xie, Molecules **22**, (2017).

73. J.T. Groves and J. Kuriyan, Nat. Struct. Mol. Biol. **17**, 659 (2010).

74. N. Samart, D. Althumairy, D. Zhang, D.A. Roess, and D.C. Crans, Coord. Chem. Rev. **416**, 213286 (2020).

75. K. Murata, K. Mitsuoka, T. Hiral, T. Walz, P. Agre, J.B. Heymann, A. Engel, and Y. Fujiyoshi, Nature **407**, 599 (2000).

76. N. Okamoto and N. Yamanaka, Curr. Biol. **30**, 359 (2020).

77. A. Hahn, J. Vonck, D.J. Mills, T. Meier, and W. Kühlbrandt, Science (80-.). **360**, (2018).

78. J. He, H.C. Ford, J. Carroll, C. Douglas, E. Gonzales, S. Ding, I.M. Fearnley, and J.E. Walker, Proc. Natl. Acad. Sci. U. S. A. **115**, 2988 (2018).

79. Y. Arinaminpathy, E. Khurana, D.M. Engelman, and M.B. Gerstein, Drug Discov. Today **14**, 1130 (2009).

80. I. Shimada, T. Ueda, Y. Kofuku, M.T. Eddy, and K. Wüthrich, Nat. Rev. Drug Discov. **18**, 59 (2018).

81. Y. Cheng, Curr. Opin. Struct. Biol. **52**, 58 (2018).

82. F. Li, P.F. Egea, A.J. Vecchio, I. Asial, M. Gupta, J. Paulino, R. Bajaj, M.S. Dickinson, S. Ferguson-Miller, B.C. Monk, and R.M. Stroud, J. Biol. Chem. **296**, 100557 (2021).

83. A. Kabedev, S. Hossain, M. Hubert, P. Larsson, and C.A.S. Bergström, J. Pharm. Sci. **110**, 176 (2021).

84. J.P. Ulmschneider and M.B. Ulmschneider, Acc. Chem. Res. **51**, 1106 (2018).

85. T. Mori, J. Jung, and Y. Sugita, J. Chem. Theory Comput. **9**, 5629 (2013).

86. C. Kandt, W.L. Ash, and D. Peter Tieleman, Methods **41**, 475 (2007).

87. P.J. Bond, J. Holyoake, A. Ivetac, S. Khalid, and M.S.P. Sansom, J. Struct. Biol. **157**, 593 (2007).

88. WG Noid, J.W. Chu, G.S. Ayton, V. Krishna, S. Izvekov, G.A. Voth, A. Das, and H.C. Andersen, J. Chem. Phys. **128**, (2008).

89. S. Seo and W. Shinoda, J. Chem. Theory Comput. **15**, 762 (2019).

90. M.I. Mahmood, A.B. Poma, and K.I. Okazak, Front. Mol. Biosci. **8**, 1 (2021).

91. D. Ugarte La Torre and S. Takada, J. Chem. Phys. **153**, (2020).

92. W. Li, W. Wang, and S. Takada, Proc. Natl. Acad. Sci. U. S. A. **111**, 10550 (2014).

93. J.A. Killian, Biochim. Biophys. Acta - Rev. Biomembr. **1376**, 401 (1998).

94. G. Nawrocki, W. Im, Y. Sugita, and M. Feig, Proc. Natl. Acad. Sci. U. S. A. **116**, 24562 (2019).

95. G. Singh and D.P. Tieleman, J. Chem. Theory Comput. **7**, 2316 (2011).

96. D.C. Marx and K.G. Fleming, J. Am. Chem. Soc. **143**, 764 (2021).

97. G. Babbi, C. Savojardo, P.L. Martelli, and R. Casadio, Int. J. Mol. Sci. **22**, 1 (2021).

98. L. Monticelli, SK. Kandasamy, X. Periole, R.G. Larson, D.P. Tieleman, and S.J. Marrink, J. Chem. Theory Comput. **4**, 819 (2008).

99. P.C.T. Souza, R. Alessandri, J. Barnoud, S. Thallmair, I. Faustino, F. Grünewald, I. Patmanidis, H. Abdizadeh, B.M.H. Bruininks, T.A. Wassenaar, P.C. Kroon, J. Melcr, V. Nieto, V. Corradi, H.M. Khan, J. Domański, M. Javanainen, H. Martinez-Seara, N. Reuter, R.B. Best, I. Vattulainen, L. Monticelli, X. Periole, D.P. Tieleman, A.H. de Vries, and S.J. Marrink, Nat. Methods **18**, 382 (2021).

100. A. Majumder and J.E. Straub, J. Chem. Theory Comput. (2021).

101. K. Tanaka, J.M.M. Caaveiro, K. Morante, J.M. González-Manãs, and K. Tsumoto, Nat. Commun. **6**, 4 (2015).

102. B.I. Sejdiu and D.P. Tieleman, Biophys. J. **118**, 1887 (2020).

103. B.C. Goh, H. Wu, M.J. Rynkiewicz, K. Schulten, B.A. Seaton, and F.X. McCormack, Biochemistry **55**, 3692 (2016).

104. F. Fernandes, A. Coutinho, M. Prieto, and L.M.S. Loura, Biochim. Biophys. Acta - Biomembr. **1848**, 1837 (2015).

105. Y. Gao, E. Cao, D. Julius, and Y. Cheng, Nature **534**, 347 (2016).

106. M.R. Elkins, J.K. Williams, MD. Gelenter, P. Dai, B. Kwon, I. V. Sergeyev, B.L. Pentelute, and M. Hong, Proc. Natl. Acad. Sci. U. S. A. **114**, 12946 (2017).

107. M.A. Lomize, I.D. Pogozheva, H. Joo, H.I. Mosberg, and A.L. Lomize, Nucleic Acids Res. **40**, 370 (2012).

108. YC. Kim and G. Hummer, J. Mol. Biol. 375, 1416 (2008).

109. M. Waldman and A.T. Hagler, J. Comput. Chem. 14, 1077 (1993).

110. See http://membrane.urmc.rochester.edu/wordpress/?page_id=126 for WHAM: the weighted histogram analysis method, accessed Feb 1, 2021.

111. S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, and P.A. Kollman, J. Comput. Chem. **13**, 1011 (1992).

112. S. Özdirekcan, C. Etchebest, J.A. Killian, and P.F.J. Fuchs, J. Am. Chem. Soc. **129**, 15174 (2007).

113. A. Ivetac and M.S.P. Sansom, Eur. Biophys. J. **37**, 403 (2008).

114. S. Simm, J. Einloft, O. Mirus, and E. Schleiff, Biol. Res. **49**, 1 (2016).

115. D.M. Engelman, T.A. Steitz, and A. Goldman, Struct. Insights into Gene Expr. Protein Synth. 147 (1986).

116. J.L. MacCallum, W.F. Drew Bennett, and D. Peter Tieleman, Biophys. J. **94**, 3393 (2008).

117. M.D. Ward, S. Nangia, and E.R. May, J. Comput. Chem. **38**, 1462 (2017).

118. T. Zhang, J. Liu, M. Fellner, C. Zhang, D. Sui, and J. Hu, Sci. Adv. **3**, 1 (2017).

119. A. Šali and T.L. Blundell, J. Mol. Biol. **234**, 779 (1993).

120. B.A. Wittenberg, J.B. Wittenberg, and P.R.B. Caldwell, J. Biol. Chem. **250**, 9038 (1975).

121. M.M. Teeter and W.A. Hendrickson, J. Mol. Biol. **127**, 219 (1979).

122. HC Ahn, N. Juranić, S. Macura, and J.L. Markley, J. Am. Chem. Soc. **128**, 4398 (2006).

123. A.J. De Jesus and T.W. Allen, Biochim. Biophys. Acta - Biomembr. **1828**, 864 (2013).

124. W. Huang and D.G. Levitt, Biophys. J. **17**, 111 (1977).

Appendix

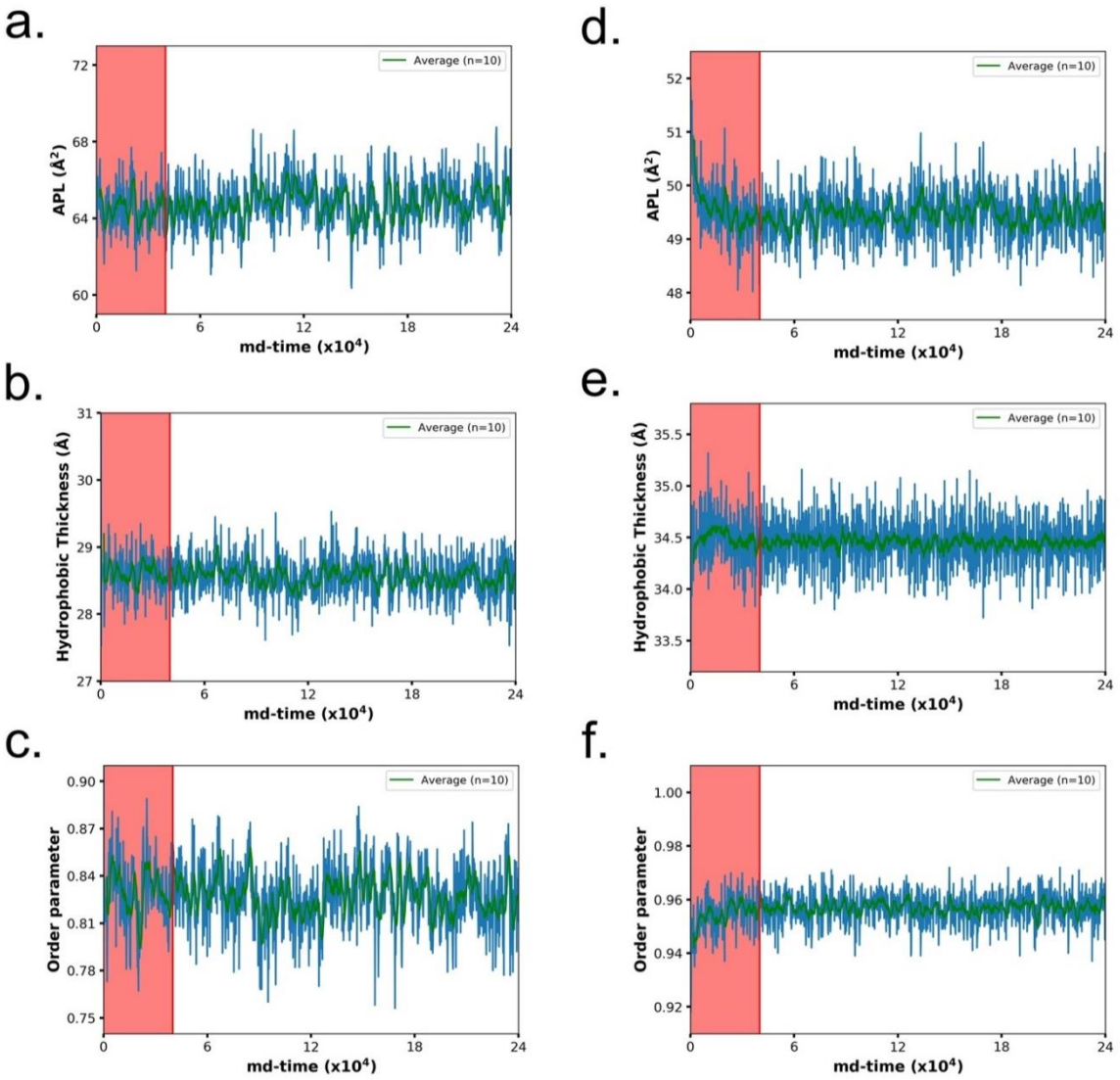

Figure A1. Time series for POPC and DPPC lipids. The area-per-lipid (APL), hydrophobic thickness, and order parameter are shown in the left column for POPC (**a-c**) and in the right column for DPPC (**d-f**). In each plot, the red zone indicates the section for the time series that was not used, the blue shadow represents the raw data, and the green line shows the running average over the last 10 data points.

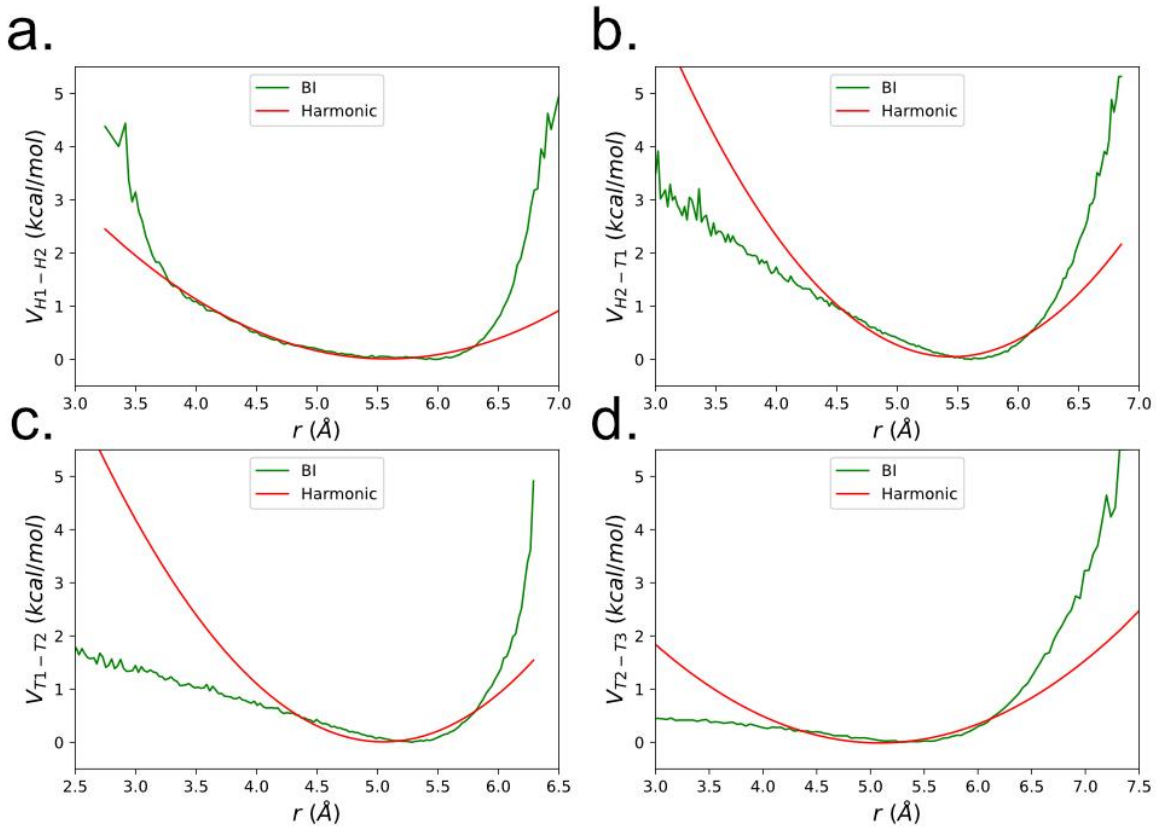

Figure A2. Boltzmann Inversion for POPC virtual bonds. (**a-d**) In each plot, the green lines represent the inverted potential obtained using the Boltzmann-inversion method for the four POPC virtual bonds. The red lines represent a fitting around the minima using a quadratic potential.

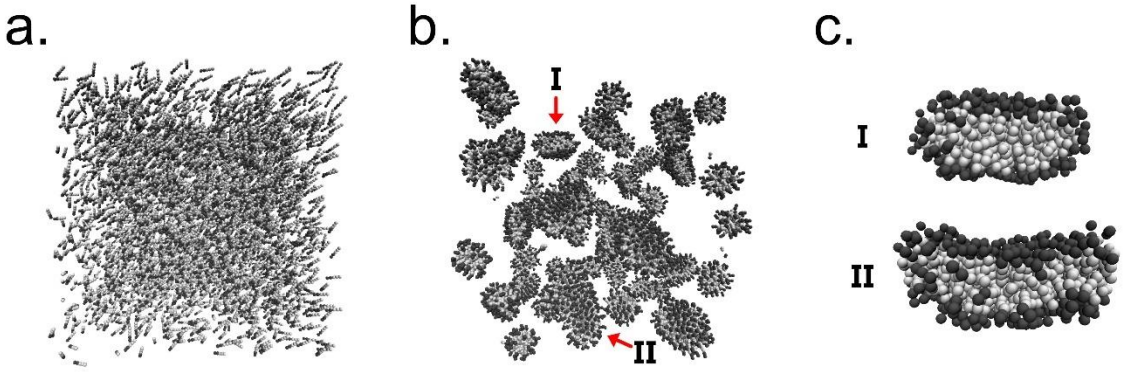

Figure A3. POPC self-assembly at 30°C. (a) The initial configuration of 5000 POPC lipids randomly placed in a box of $500\text{Å} \times 500\text{Å} \times 500\text{Å}$ without periodic boundary conditions. Each wall has an associated repulsive potential to prevent lipids from diffusing away. (b) Configuration obtained after 1.6×10^5 MD-time. (c) Clusters formed during the simulation (I and II in b).

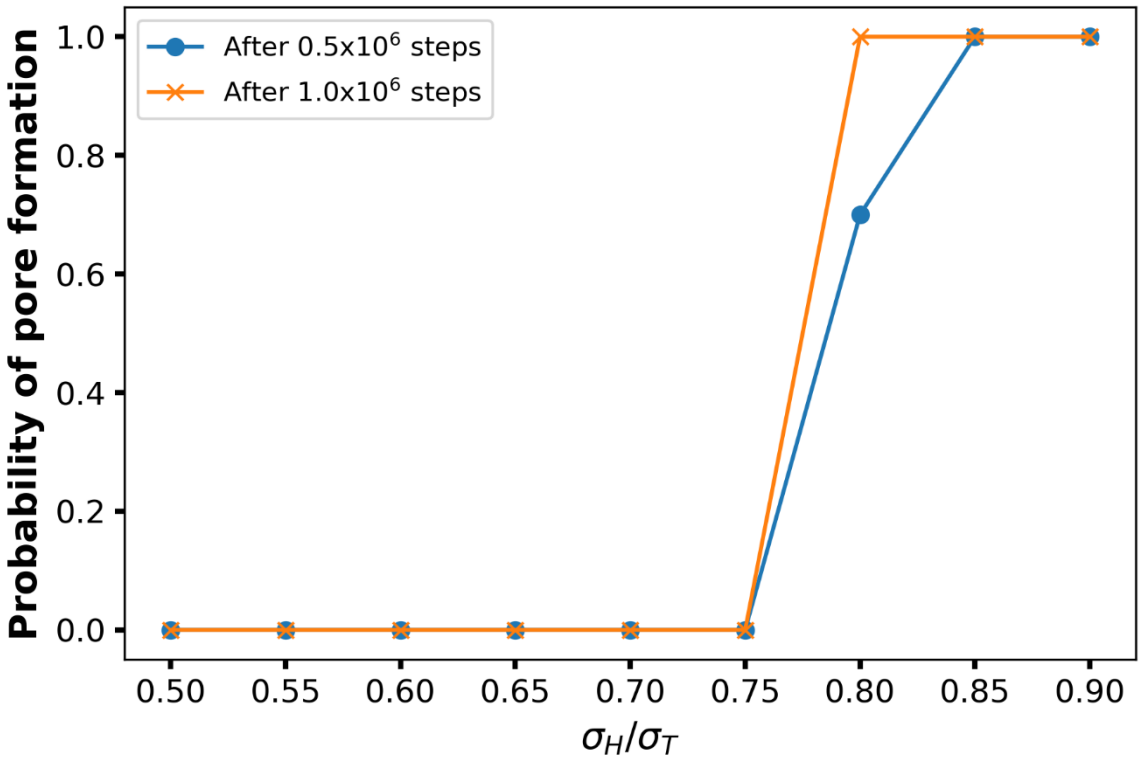

Figure A4. Pore formation probability for POPC. Each point was calculated as the ratio of membranes containing a hole over 20 trajectories. For a sigma ratio, σ_H/σ_T, lower or equal to 0.75, no pore was observed. However, for ratios bigger than 0.75, pores started to form.

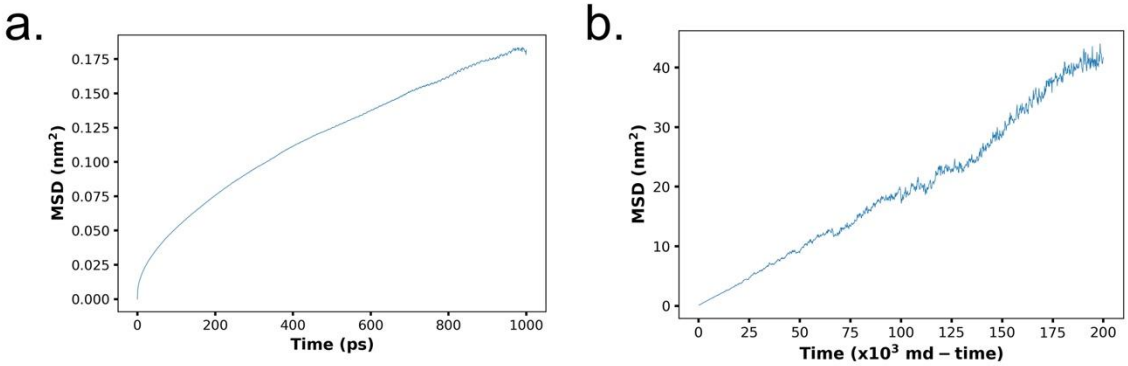

Figure A5. Comparison of lateral diffusion for POPC. (**a**) All-atom simulation of the MSD, using the SLipids force field. (**b**) Coarse-grained simulation of the MSD, using the developed lipid model, iSoLF.

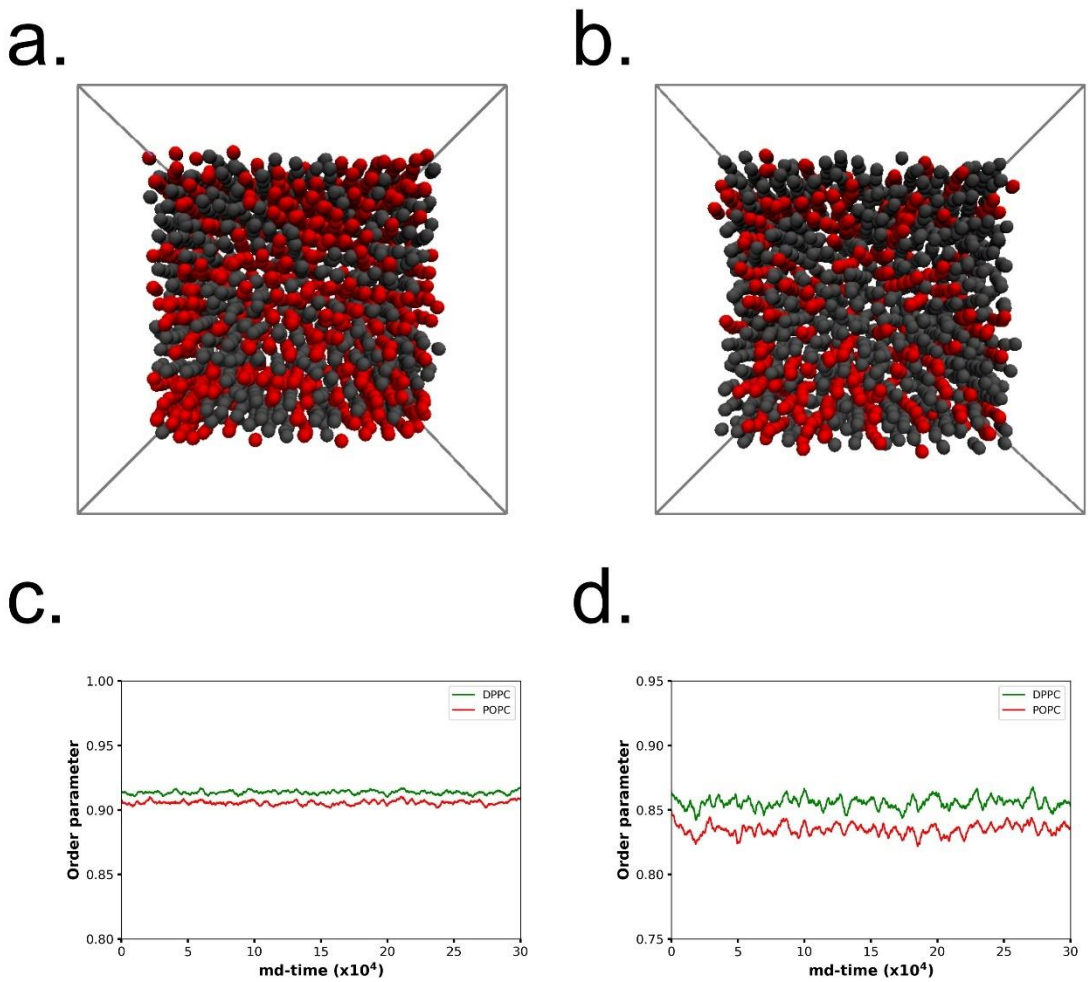

Figure A6. Two-component membrane simulation. (a) Last snapshot of the 1:1 POPC:DPPC ratio membrane. (b) Last snapshot of the 2:1 POPC:DPPC ratio membrane. (c) Order parameter time series for the 1:1 ratio system. It stayed in the gel phase. (d) Order parameter time series for the 2:1 ratio system. POPC presented a fluid phase while DPPC stayed in a gel phase. In (a) and (b), POPC and DPPC lipids are assigned the colors grey and red, respectively.

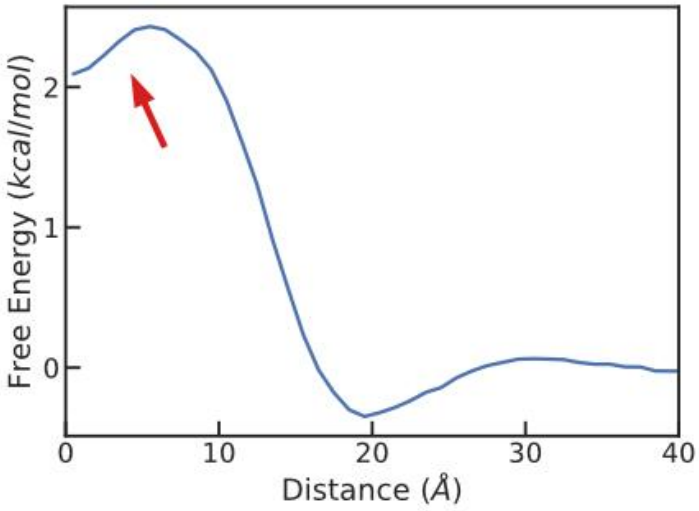

Figure A7. Free energy profile with local minima at the center of the membrane. By using as combination rule the arithmetic mean, $\sigma_{ij} = (\sigma_i + \sigma_j)/2$, a local minimum appeared at the center of the bilayer. This happened because the difference in size between amino acids CG beads and the lipid tail beads T3 permitted the formation of a small cavity that stabilized hydrophilic residues.

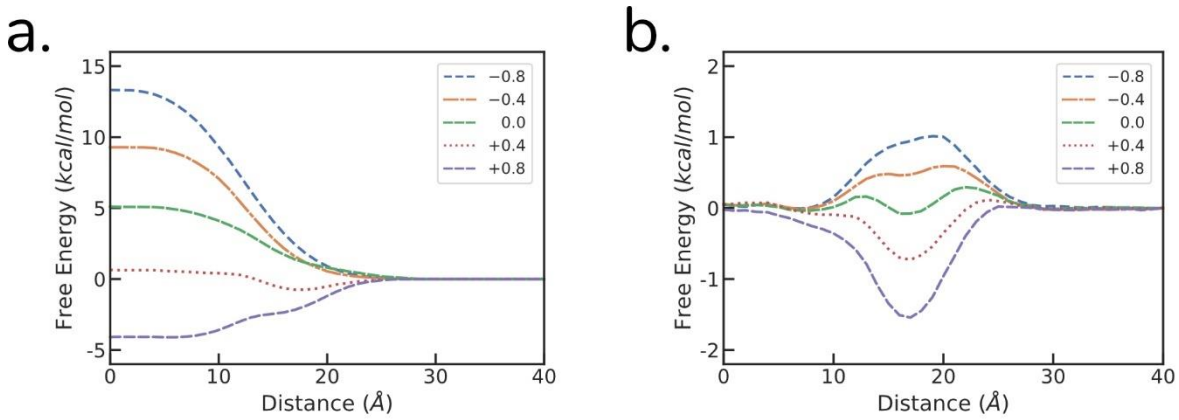

Figure A8. Free energy as a function of ε_{HP}. (a) Free energy profiles obtained by varying the force coefficient for tail beads (T1, T2, and T3) while setting the force coefficient for the head beads (H1 and H2) equal to $0\ kcal/mol$. (b) Free energy profiles obtained by varying the force coefficient for head beads (H1 and H2) while setting the force coefficient for the tail beads (T1, T2, and T3) equal to $0.448\ kcal/mol$. This show that the parameterization of amino acids for the interactions with the lipid head and tail beads can be performed almost independently.

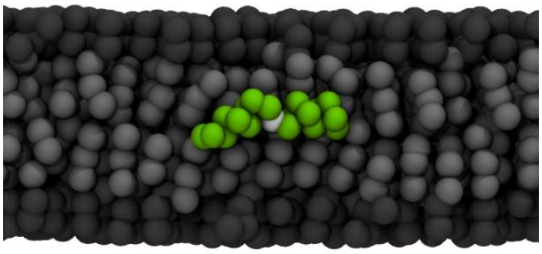

Figure A9. WALP embedded into the membrane. For small peptides, the terminal residues play a major role since they contribute to the overall stabilization through electrostatic interactions with the hydrophilic region of the lipid membrane. By including special treatment for the N- and C- terminal, the embedded configuration of WALP is destabilized.

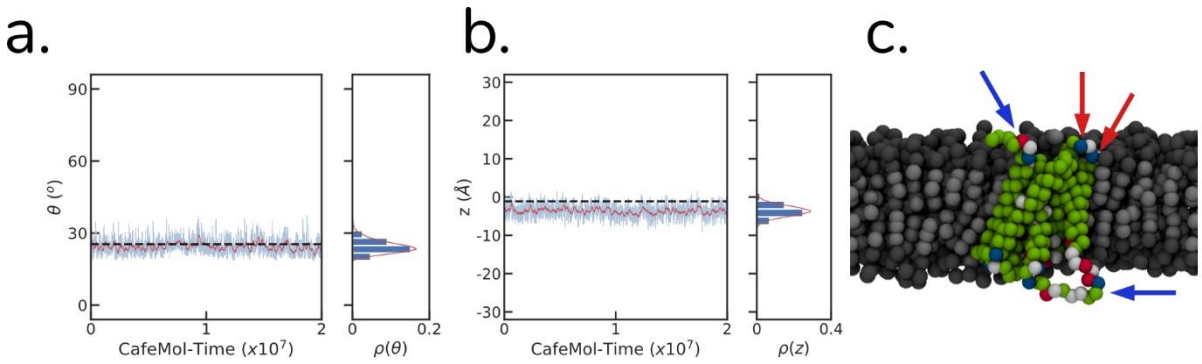

Figure A10. Orientation for the mutated ZIP. The figures show the (a) tilt angle and (b) insertion depth for the artificially mutated ZIP. By mimicking a divalent cation bound to the protein and filling the missing flexible loops, the correct orientation of ZIP was obtained. In the snapshot (c), blue arrows indicate the filled flexible loops, and red arrows indicate the mutated residues.