Machine Learning for Metallic Corrosion Modeling: A Computational Exploration

Kiran

First Printing, 2024

Table of Contents

Chapter 1: Introduction 11
Molecular Dynamics (MD) Simulations 11
Machine Learning (ML) Algorithms 12
Chapter 2: Relationship Between Cholesterol Distribution and Ordering of Lipids 17
Introduction 17
Methodology 19
Force Field 19
Simulation Details 20
Results and Discussion 22
Cholesterol Distribution in the Leaflets of the Bilayer 22
Cholesterol Flip-Flop Rate 27
Arrangement of Cholesterol in the Plane of the Bilayer 28
Cholesterol Localization as a Function of lipids Structural Properties 30
Conclusion 39
Chapter 3: Relationship Between Cholesterol Distribution and Phase of Bilayer 41
Introduction 41
Computational Methods 43
Systems Setup 43
Simulations Parameters 44
Lipid Bilayers Properties 45
Results and Discussion 46
Comparison of Dry Martini and Wet Martini force fields − Study I 46
Cholesterol Distribution as a Function of Lipid Bilayers Domains – study II 49
Conclusion 61
Chapter 4: Estimation of Phase Diagrams for Three-Component Lipid Mixtures 63

Introduction 63
Material and Methods 66
Datasets 66
Design of Elements 71
Evaluation Metrics 74
Results and Discussion 75
Prediction of Phospholipid Melting Temperature 75
Prediction of Phase Diagrams 78
Conclusion 84
Chapter 5: Modeling and Predicting Performance of Corrosion Inhibitors 87
Introduction 87
Methodology 89
Description of the Experimental Data 89
Employed ML Methods 92
ML Implementation Methodology 92
Results and Discussion 95
Identifying the Best ML Model 95
Predictions of Time-dependent Corrosion Rates Using RF 99
Sensitivity to Input Variables 106
Conclusions 109
Chapter 6: Conclusions 111
References 115
Appendix A: Correspond to the Materials Presented in Chapter 3 128
Appendix B: Correspond to the Materials Presented in Chapter 4 136
Appendix C: Correspond to the Materials Presented in Chapter 5 146

Chapter 1: Introduction

This Ph.D. research has two main objectives. First, studying the nature of plasma membrane of eukaryotic cells, known as lipid bilayer, in aqueous environment using molecular dynamics (MD) simulations and machine learning (ML) methods. Second, investigating the application of ML for the purpose of modeling metallic corrosion. The rest of this document can be summarized as follows. In chapter 1, a brief introduction to MD and ML techniques is presented. Chapters 2 to 4 cover research objective I, while chapter 5 contains the details of research objective II. Finally, in chapter 6 the main conclusions of the research are reported.

Molecular Dynamics (MD) Simulations

MD is a technique to simulate molecular motion on a computer. Forces between atoms are electrostatic in nature because of interactions between electron clouds and nuclei. In classical MD, a representation of the physical system of interest is generated by placing N atoms/molecules in a volume V. The forces between the atoms are represented by mathematical functions of distance collectively called a force field. The nature of interactions between the atoms depends on their characteristics. In non-polar species, the dipole-dipole interactions between their electron clouds dominate, which are called van der Waals interactions. Van der Waals interactions are commonly represented by Lennard-Jones force field. Charged and polarized species interact via Coulombic interactions. Bond and angular interactions are represented by harmonic potentials. In addition, potential functions are also defined for dihedral interactions within a molecule. **Equation 1** shows the force field or potential that defines the energies between particles.[1]

$$V(r) = \sum_{bonds} K_r\left(r - r_{eq}\right)^2 + \sum_{angles} K_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedral} \frac{V_n}{2}[1 + cos(n\varphi - \gamma)] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right] + \sum_{i<j} \frac{q_i q_j}{\in R_{ij}} \quad \text{Equation 1}$$

The first term represents a harmonic bond between two bonded particles with a spring constant K_r and equilibrium distance of r_{eq}. The second term is the harmonic angle between three bonded particles with a constant $K_\theta$and equilibrium angle of θ_{eq}. The third term is the dihedral angle that are between four particles and represent a potential of rotation about the central bond. The fourth and fifth terms are space energies of Van der Waals and electrostatics between non-bonded particles. The negative gradient of the potential energy is the force. Once the forces acting on all the particles are calculated, the motion of the particles is generated by numerically integrating the classical Newton's equation of motion. By doing this for long enough time until reaching to the equilibrium we can analyze behavior of different systems under different conditions. Often, it is useful to represent molecules using a coarse-grained description wherein all a group of atoms are represented by a united bead. A popular coarse-grain model to study lipids is called the MARTINI force field.[2]

Machine Learning (ML) Algorithms

Machine learning methods are suitable for developing predictive models in the cases where a large dataset is available, the outcome to be predicted depends on several variables, and when a mechanistic model of the relationship between the input variables and the outcome is not well established. Before we embark on describing our work in

detail, it is useful to provide a brief introduction to machine learning methods used here for the readers who are experts in the corrosion field but not necessarily in ML. Machine learning refers to a class of algorithms that have the ability to learn and perform a task when they are provided with some data.[3,4] Thus, ML algorithms get to learn how to perform a task, such as predicting the outcome of an experiment, when they are trained on the data obtained from some previously performed experiments. Specifically, the ML algorithms that learns from data wherein the outcome of the experiment is specified are called supervised learning.[4] Some popular supervised ML algorithms are Random Forest (RF), Artificial Neural Network (ANN), Supported Vector machines Regression (SVR), and K Nearest Neighbors (KNN).

To understand RF algorithm, one needs to first understand what is meant by a decision tree. In a decision tree, a dataset is split around input variables into smaller subsets. The split is performed so that the subsets that are formed have smaller variances in the outcome values. Each split can be thought of as a branch of the tree and each data subset as a leaf. The data are progressively split until some terminal condition is met. The terminal condition can be either that the maximum number of splits has been performed or the standard deviation of a subset has fallen below a cutoff value. The average value of the outcome in the terminal leaf, that is the leaf, which is not split any further, is the predicted value of the outcome for those set of input variables. As a simple example, imagine a decision tree which is used to predict the corrosion rate as a function of two key variables, pH and temperature, is created with a tree split with branches $pH < 7$ and $pH \geq 7$. These branches are each further split for temperature < 350 K and temperature $\geq$

350 K. Then, to predict the corrosion rate for a condition, say with pH = 4.5 and temperature = 300 K, the average value of the corrosion rate for the leaf: pH < 7 → temperature < 350 K is the predicted value of this decision tree. A RF is comprised of a multitude of decision trees. The decision trees are formed based on random subsets of the training data. RF algorithm reports the mean value of the predictions from all decision trees. In general, using the outcome of many ML models to make the final prediction is called ensemble learning, and has been shown to significantly improve the prediction performance.[5] Therefore, RF is an ensemble learning method based on numerous decision trees. So, in the example, the corrosion rate is the output variable (label) and pH and temperature are the input variables (features).[6]

ANNs are networks of interconnected nodes that act as universal approximators, that is, they can approximate any continuous function to an arbitrary level of accuracy with a finite number of nodes.[7] The architecture of ANN is as follows: In the first layer (named input layer), each input variable, x_i is fed to a node. Each node in the input layer is connected to the nodes in the next layer (called first hidden layer). The connections between the nodes are assigned some weights, $w_{ij.}$ At each node in the first hidden layer, a weighted sum of the inputs from the input layer is calculated, $F_j = \sum_i w_{ij} x_i$. The F_j is transformed via an activation function, such as a sigmoidal function, which is the output from each node, $x_j = \frac{1}{1+e^{-F_j}}$, which becomes the input from node j in the first hidden layer to all the nodes in the second hidden layer. This process is repeated for all the layers until the output layer is reached, which in regression problems is a single node that predicts the outcome. Training an ANN architecture entails adjusting the weights

connecting the nodes so as to minimize the mean squared error between the predicted and the actual outcome values/classes.[8,9]

SVR is another powerful supervised learning algorithm. In general, the relationship between the input variables and the label is non-linear. In SVR, the input variables are transformed to higher dimensions where the relationship is linear. In a higher dimension, a linear regression line is fitted to the data. For example, a polynomial relationship, such as $y = a_0 + a_{11}x_1 + a_{12}x_2 + .. + a_{n1}x_1^n + a_{n2}x_2^n$ can be represented as a *2n* dimensional hyperplane. There are special functions called kernels that help in determining the hyperplane in the higher dimension without increasing the computational cost.[10]

KNN is a simple algorithm in which, for each data point, *K* nearest-neighbors are identified in the training set (where *K* is an integer), and the average value of their labels is reported as the outcome of that data point.[11] The *K* nearest-neighbors are identified by defining a *distance*. For numerical input variables, Euclidean distance is commonly used, whereas, for categorical variables, Hamming distance is used.[12]

As discussed above, every ML algorithm has a specific architecture. For instance, an ANN is characterized by the number of hidden layers and the number of nodes per hidden layer; a RF is characterized by the number of trees and the maximum number of features that can be split. These parameters that define the architecture of a ML algorithm are called hyperparameters.[13] The performance of a ML algorithm on a dataset varies as one changes the hyperparameters. Therefore, values of the hyperparameters need to be tuned/adjusted to optimize performance. To tune the hyperparameters, first, the data

should be split into a training set and a testing set. The training set is often comprised of 70-80% of the entire dataset. The optimum values of the hyperparameters are found by evaluating the performance of all formed ML models on the training set. Once the hyperparameters are fixed, the founded best ML algorithm is trained on the training set, and then its performance is evaluated on the testing set.

Chapter 2: Relationship Between Cholesterol Distribution and Ordering of Lipids

The material presented in this chapter is put together as a paper: "Aghaaminiha, M.; Farnoud, A. M.; Sharma, S. Quantitative Relationship between Cholesterol Distribution and Ordering of Lipids in Asymmetric Lipid Bilayers. *Soft Matter* **2021**, *17* (10), 2742–2752".[14]

Introduction

The plasma membrane is a lipid bilayer that separates the cell from the outside environment.[15–17] While different mammalian cells are known to have a wide variety of lipids in their plasma membrane, it is now well-understood that the phospholipid composition of the inner (cytofacial) leaflet in all mammalian cells is different from that of the outer (exofacial) leaflet.[18–20] Verkleij et al. pioneered studies on phospholipid asymmetry, showing that sphingomyelin (SM) and phosphatidylcholine (PC) are abundant in the outer leaflet, while phosphatidylethanolamine (PE) and phosphatidylserine (PS) are primarily localized in the inner leaflet.[21] These findings were later verified by the recent studies of Lorent et al.[19] and Vahedi et al.[22] The phospholipid asymmetry in membrane leaflets has important functional implications for the cell. For instance, in healthy red blood cells, PS is known to be almost exclusively localized in the inner leaflet, the presence of this lipid in the outer leaflet is a sign of eryptosis and a signal to macrophages to phagocytose the old red blood cells and maintain homeostasis.[18,23]

While the asymmetry of phospholipids in the plasma membrane is well-established, there is little consensus regarding cholesterol distribution in different leaflets of the plasma membrane.[18,20,24] Unlike phospholipids, there are no specific enzymes for

cholesterol transport across the bilayer, and cholesterol distribution is believed to be primarily dictated by its affinity towards other lipids.[18,24] However, experimental results in this regard have been contradictory. It has been reported that cholesterol molecules are enriched in the outer leaflet because of their strong interactions with the saturated acyl chain phospholipids, specifically SM.[25–27] Contrarily, it has also been suggested that the inner leaflet is favored by cholesterol due to (i) the abundance of PE phospholipids in this leaflet which provides higher negative curvature,[28,29] and (ii) the abundance of C24 sphingolipids in the outer leaflet which pushes cholesterol to the inner leaflet.[30] There are also other reports that mention that cholesterol is symmetrically distributed in the cell membrane.[31,32]

Experimental studies of cholesterol distribution are complicated by several confounding factors. First, different mammalian cells have different membrane lipid compositions.[33,34] Secondly, experimental methods such as quenching of dehydroergosterol fluorescence, filipin staining, noninvasive neutron scattering, and enzymatic degradation by cholesterol oxidase, while helpful, each suffer from various drawbacks.[35,36] Finally, cholesterol is known to flip-flop between the two leaflets at a rapid rate, [37–40] making it difficult for experimental probes to trace it with high accuracy at any time-point.[41] A mechanistic study on how changes in lipid structure and chemistry affect cholesterol distribution helps to provide a general understanding of cholesterol asymmetry in the plasma membrane.

In the current work, we have studied the spatial distribution of cholesterol within asymmetric lipid bilayers via coarse-grained (CG) molecular dynamics (MD)

simulations. We have performed two sets of simulations, with the first focusing on characterizing cholesterol distribution in a lipid bilayer mimicking the known membrane lipid composition of red blood cells and the second focusing on elucidating how the structure and chemistry of the lipids in the bilayer affect cholesterol distribution. Results reveal that the ordering of lipids in membrane leaflets is the primary regulator of cholesterol spatial distribution in the asymmetric bilayers.

Methodology

Force Field

We have employed the coarse-grained (CG) Wet Martini force field with polarizable water to model all lipids, cholesterol, ions, and water molecules in the system. The polarizable water model for the Martini force field, developed by Yesylevskyy et al.,[42,43] is a three-bead model to represent four water molecules and can capture the orientational polarizability and the dielectric screening effect of bulk water. The topologies of lipids are described in **Figures A1-C** and **A2-B** (Appendix A) alongside the name and type of each bead. The main types of particles in the Martini CG force field are P, polar; N, intermediate; C, apolar; and Q, charged. Subscripts *0*, *a*, *d*, and *da* indicate if the particle has hydrogen-bond (H-bond) forming capacities. The particle with "*0*" has no H-bond capacity, "*a*" is a H-bond acceptor, "*d*" is a H-bond donor, and "*da*" is both H-bond acceptor and donor.[44] Other subscripts are numbers: 1, 2, 3, 4, and 5 that display the degree of polarity, where 5 is the most and 1 is the least polar particle.[44]

Periodic boundary conditions were applied in all three directions of the simulation box. In the first set of simulations (with the sample snapshot in **Figure A1-B**; Appendix A), the

simulation box size was 250 × 250 × 125 (Å^3), the total number of lipids in the systems was 2,043, and the total number of water molecules was 41,487. In the second set of simulations (with the sample snapshot presented in **Figure A2-A**; Appendix A), the simulation box size was 125 Å in all directions, the total number of lipids in the systems was 510, and the total number of water molecules was 10,834. Initial configuration of lipid bilayers was generated using the *Insane.py* script developed by Marrink et al.[45] All simulations were performed on the GROMACS/5.1.2 molecular simulation package.[46]

Simulation Details

The initial configuration of the system was energy minimized by employing steepest descent algorithm with the maximum force tolerance of 1000.0 kJ.mol^{-1}.nm^{-1}. This was followed by a 25 ns canonical ensemble (constant number of particles *N*, volume *V*, and temperature *T*) simulation and a 25 ns isothermal-isobaric (constant *N*, pressure *P*, *T*) ensemble simulation for pre-equilibration. The time-step chosen for both the ensembles was 10 fs. Then, three replicas of 10 μs NPT simulation with a timestep of 20 fs were performed, as in previous studies [17], and the data was collected for the last 6 μs for all the simulations.

The parameters of the simulation system parameters were chosen to be the same as in the previous work done to parameterize the polarizable water model on the Martini force field.[43] Briefly, a spherical cut-off of 1.1 nm was chosen for the Lennard Jones (LJ) interactions. For Coulombic interactions, Reaction-Field-zero (RF-zero) was employed with the spherical cut-off of 1.1 nm.[43] The RF-zero and PME (Particle Mesh Ewald) electrostatics have been systematically compared previously and no differences in the

bilayer properties were identified.[43] The relative dielectric constant was taken to be 2.5.[42,43] Both LJ and Coulombic interactions were smoothly shifted to be zero beyond the cut-off. In the short pre-equilibration runs, Berendsen thermostat and barostat with the time constants of 1 ps and 3 ps, respectively were employed for quick equilibration. For the NPT equilibration and production runs, the velocity-rescaling thermostat developed by Bussi et al.[47] was used. Semi-isotropic pressure coupling was accomplished using Parrinello-Rahman barostat with the coupling time constant of 12 ps.[17,43] Along the axis of the bilayer plane and the z-direction, the reference pressure and isothermal compressibility were set to be 1.0 bar and 3×10^{-4} bar^{-1}, respectively. All simulations are performed at 37°C.

Radial distribution functions, g(r), were calculated to examine the proximity of the hydroxyl headgroup of cholesterol (ROH CG bead) to various chemical moieties in phospholipids. Pioneering work by Huang has described the importance of the proximity of the –OH group of sterols, in the β configuration, to the carbonyl oxygen of the fatty acyl groups in phospholipids for strong cholesterol-lipid interactions.[48]

We divide the bilayer into three regions, namely, inner leaflet, midplane, and outer leaflet. The threshold for separating these regions was set to be 0.8 nm from the center of the bilayer as previously suggested.[16,17] Cholesterol flip-flop rate was defined as the average frequency by which the cholesterol molecules undergo intramembrane exchange, in other words, leave the inner/outer leaflet, reach the outer/inner leaflet, and then return to the inner/outer leaflet.[37,38,49] Orientation of a cholesterol molecule within a bilayer was defined as the angle between vector, $\bar{a}$ (the vector joining the ROH bead to

the C1 bead of cholesterol, see **Figure A1-C**, Appendix A) and vector $\bar{z}$ (the bilayer normal). To measure the ordering of lipids in each leaflet, ensemble-averaged order parameter of the acyl chains of lipids[2,50] was calculated , as in **Equation 2**:

$$S_{Chain} = \langle \frac{1}{2}(3\langle cos^2\theta \rangle - 1) \rangle \quad \text{Equation 2}$$

Where θ is the angle between each bond vector of the acyl chain beads and the bilayer normal. A high value of the S_{Chain} (~ 1) indicates that the bonds are aligned with the bilayer normal, whereas a low value (~ 0) indicates that the bonds are randomly oriented respect to the bilayer normal.[2,50] The area per lipid, *a(x)* for each leaflet was calculated based on the ensemble-averaged area of the plane in the simulation box, *A(x)* divided by the total number of two-chain lipids + cholesterol molecules in that leaflet according to **Equation 3** (*x* is the mole fraction of cholesterol in the system).[51]

$$a(x) = \frac{A(x)}{N_{lipids} + N_{chol}} \quad \text{Equation 3}$$

The thickness of the bilayer was estimated as the ensemble-averaged distance between the phosphate head group of the lipids (PO4 CG bead shown in **Figure A1-C**, Appendix A) in the two leaflets.[51,52]

Results and Discussion

Cholesterol Distribution in the Leaflets of the Bilayer

The asymmetric distribution of lipids in the plasma membrane is expected to affect the localization of cholesterol in membrane leaflets. To examine this phenomenon, an asymmetric lipid bilayer, mimicking the lipid composition of the plasma membrane of red blood cells was studied. The asymmetric lipid bilayer comprised of seven different

lipids. The outer leaflet lipids were selected to be POPC/16:0SM/POPE/CHOL with the molar ratio of 31/29/7/33 while the inner leaflet lipids were DOPC/16:1SM/DOPE/DOPS/CHOL with the molar ratio of 10/7/31/19/33. This composition of phospholipids matches the predominant lipids in each leaflet of the plasma membrane of red blood cells.[19,21,22] Cholesterol was the only sterol included in the system with a molar ratio of 1:2 with respect to the total number of lipids in the two leaflets. Cholesterol content has been reported to be in the range of 30 mol% to 50 mol% in mammalian cell plasma membranes.[53,54] Two symmetric model lipid bilayers were also studied. In the first bilayer, labeled cyto-symmetric, the two leaflets had the same lipid composition as that of the inner (cytofacial) leaflet of the asymmetric bilayer noted above. In the second symmetric bilayer, labeled exo-symmetric, the composition of both leaflets was the same as the outer (exofacial) leaflet of the asymmetric bilayer. **Table 1** listed the lipid compositions of the three systems.

Table 1

Composition of the inner and outer leaflets of the studied lipid bilayers
Cholesterol was included in all the bilayers with the molar ratio of 1:2 with respect to the total number of lipids.

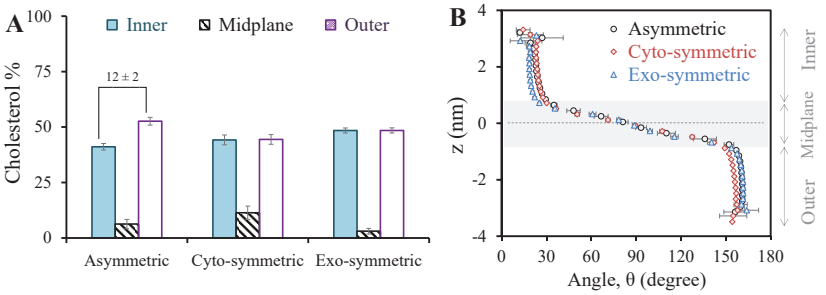
System	Inner leaflet	Outer leaflet
Asymmetric	DOPC/16:1SM/DOPE/DOPS/CHOL: 10/7/31/19/33	POPC/16:0SM/POPE/CHOL: 31/29/7/33
Cyto-symmetric	DOPC/16:1SM/DOPE/DOPS/CHOL: 10/7/31/19/33	DOPC/16:1SM/DOPE/DOPS/CHOL: 10/7/31/19/33
Exo-symmetric	POPC/16:0SM/POPE/CHOL: 31/29/7/33	POPC/16:0SM/POPE/CHOL: 31/29/7/33

In the initial configuration, cholesterol molecules were equally distributed between the two leaflets. Upon equilibration of the simulation, cholesterol distribution became asymmetric in the bilayer with asymmetric lipid composition, and symmetric in case of the two symmetric bilayers, confirming the direct role of phospholipid chemistry in cholesterol localization (**Figure 1A**). In the asymmetric bilayer, the fraction of cholesterol molecules in the inner and outer leaflets were 41±1% and 53±2%, respectively, while 6±2% of cholesterol resided in the midplane of the bilayer. In the cyto-symmetric bilayer, mimicking the inner leaflet, a larger fraction of cholesterol was found in the midplane (11±3% compared to 6±2% for the asymmetric bilayer), while the exo-symmetric bilayer only had 3±1% of cholesterol molecules in the midplane. This result is consistent with the experiments showing that cholesterol have a preference for the bilayer midplane in the case of polyunsaturated bilayers.[55]

Figure 1B shows the orientation of cholesterol molecules as a function of location within the bilayer for the three bilayers. For this calculation, the lipid bilayer was divided into ~30 slabs parallel to the plane of the bilayer. In each slab, the ensemble-averaged angle of the cholesterol molecules with respect to the bilayer normal was found. The reported angle with respect to the bilayer normal corresponds to the average over the configurations in the last 6 μs of three independent replicas of each system. Within 0.8 nm from the center of the bilayer, cholesterol orientation changes almost linearly with the distance from the center for all the three studied systems. At the distance of 0.8 nm from the center, a sharp change in the orientation of cholesterol is observed and so this region is defined as the midplane.[16]

Figure 1

Percentage and orientation of cholesterol in different regions of the bilayers
A) *Percentage of cholesterol molecules located in the inner leaflet (cyan), midplane (black striped), and the outer leaflet (purple) of the asymmetric, cyto-symmetric, and exo-symmetric lipid bilayers.* ***B)*** *Ensemble-averaged angle of cholesterol molecule with the bilayer normal as a function of the location of the hydroxyl headgroup of cholesterol. Shaded area represents the midplane of the bilayer. Beyond the midplane, the orientation is almost independent of the location. Error bars in both the figures were estimated as standard error of the mean when the last 6 μs of the three replicas of each system were split into four equally sized blocks of 1.5 μs each.*



It has been suggested that the alignment of cholesterol in the leaflet favors the cholesterol-lipid interaction in the PM.[48,56] Therefore, it is hypothesized that the ordering of lipids in a leaflet should favor the presence of cholesterol. Indeed, the outer leaflet of the asymmetric bilayer has a larger S_{Chain} (0.48±0.02) as compared to that of the inner leaflet (0.33±0.02). S_{Chain}, as defined in **Equation 2**, is a measure of ordering of the lipids in a leaflet; a high value of S_{Chain} corresponds to more order. The exo-symmetric and the cyto-symmetric bilayers have the S_{Chain} values of 0.48±0.02 and 0.32±0.02, respectively, for both leaflets, which explains the higher concentration of cholesterol in the exo-symmetric leaflets as compared to in the cyto-symmetric leaflets. These findings suggest

that the ordering of the lipids in the leaflets is an important factor in determining the localization of cholesterol within the leaflets of the bilayer. Other physical characteristics of these bilayers, such as area per lipid and the thickness, were also compared and are listed in **Table A1** (Appendix A). The cyto-symmetric is the thinnest and exo-symmetric is the thickest bilayer, as expected, but the difference in the thickness is only ~ 2 Å. The asymmetric bilayer has a slightly smaller area per lipid (0.52±0.01 nm^2) at the inner leaflet as compared to the cyto-symmetric bilayer (0.52±0.01 nm^2) and has slightly larger area per lipid (0.48±0.01 nm^2) at the outer leaflet as compared to the exo-symmetric bilayer (0.47±0.01 nm^2). This difference confirms that the asymmetry influences the properties of the leaflets.[57] The values of area per lipid are smaller than those usually reported for the bilayers where no cholesterol is present (~ 0.5 to 0.7 nm^2).[58]

It should be noted that the composition of lipids in different leaflets of the asymmetric bilayer did not get altered during the simulations. The density profiles of different lipids in the equilibrated bilayer confirm this point (**Figure A3**; Appendix A). On average, in the asymmetric bilayer only 2 out of 103 DOPC, 1 out of 74 SM, 7 out of 311 DOPE, and 6 out of 192 DOPS lipids moved from the inner leaflet to the outer leaflet, while no lipid molecule from the outer leaflet moved to the inner leaflet.