Identification of domain phases in selected lipid membrane compositions

Mateusz Rzycki[1][0000−0003−3655−3048], Karolina Wasyluk[1], and Dominik Drabik[1][0000−0003−4568−4066]

Department of Biomedical Engineering, Wroclaw University of Science and Technology Wroclaw, 50-370, Poland
mateusz.rzycki@pwr.edu.pl

Abstract. Lipid microdomains are specialized structures that play crucial roles in various physiological and pathological processes, such as modulating immune responses, facilitating pathogen entry, and forming signaling platforms. In this study, we explored the dynamics and organization of lipid membranes using a combination of molecular dynamics simulations and a suite of machine learning (ML) techniques. Using ML algorithms, we accurately classified membrane regions into liquid order, liquid-disordered, or interfacial states, demonstrating the potential of computational methods to predict complex biological organizations. Our investigation was mainly focused on two lipid systems: POPC/PSM/CHOL, and DPPC/DLIPC/CHOL. The study underscores the dynamic interaction between ordered and disordered phases within cellular membranes, with a pivotal role of cholesterol in inducing domain formation.

Keywords: microdomains · lipid membranes · machine learning · molecular dynamics · gel domains

1 Introduction

The cell membrane is a complex and dynamic structure that plays a pivotal role in maintaining integrity and regulating various cellular processes [3]. The distribution of lipids within these membranes is heterogeneous and certain lipids aggregate to form distinct domains [12]. These include structured gel phases (S_o) and less structured liquid disordered phases (L_d), with a special case of structured liquid ordered phases (L_o) induced by sterols like cholesterol [15]. Lipid rafts, a notable L_o domain, contain sphingomyelin and cholesterol, affecting the membrane's structure and function due to its composition and interactions. Lipid microdomains are highly ordered regions within membranes, crucial for signal transduction, endocytosis, and membrane trafficking, due to their ability to stabilize and organize membrane proteins [5, 16]. In contrast, the S_o domains exhibit a different narrative and are rare in living organisms. The L_d regions, composed of disordered lipids, provide a dynamic matrix. This matrix surrounds the ordered domains and is essential for protein diffusion and cellular responsiveness

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_13

[6]. Lipid rafts, specialized microdomains, play crucial roles in various physiological and pathological processes such as modulating immune responses, facilitating pathogen entry, and forming signaling platforms [7]. Therefore, the role of lipid domains in cell signaling and organization emphasizes the importance of understanding their impact on cell homeostasis.

Accurate identification of microdomains is particularly crucial, given their unique mechanical properties, which may render them potential targets for specific antimicrobial agents [14]. A promising method for identifying microdomains involves the use of machine learning (ML) techniques on molecular dynamics (MD) membrane trajectories, although research in this area remains limited. This study synthesizes methodologies from previous research [11, 4], employing supervised learning in modeled and spontaneously formed microdomains in ternary lipid mixtures in MD, thus improving the current knowledge base. The primary advantage of this approach is its precision in predicting molecular organization, domain, nondomain, and interface, accurately reflecting experimental observations.

2 Methods

2.1 System preparation and simulation

Using the CHARMM-GUI membrane builder [18], we constructed several membrane systems with various lipid compositions, each designed to represent different molecular organizations. Six training systems were created, each containing 150 lipid molecules, representing different domain stages. Lipid compositions included phosphatidylcholine (POPC), sphingomyelin (PSM), cholesterol (CHOL), as well as other phosphatidylcholines (DPPC and DLIPC), mixed in specific ratios to mimic nondomain, interface, and domain-like phases. Lipid ratios were set as follows: 8:1:1 for nonraft (POPC/PSM/CHOL), 2:1:1 and 4:3:3 for interface stages (POPC/CHOL/PSM), and varying ratios such as 1:2:2, 2:1, and 1:1 for raft-like stages (CHOL/PSM). Analogously, for the second approach (spontaneous domain formation), we replaced PSM with DPPC and POPC with DOPC in train systems, following established literature protocols [19, 12, 4].

Additionally, two specialized testing systems were devised, containing 1140 and 900 lipid molecules, respectively, with equal proportions of POPC/PSM/CHOL and DPPC/DLIPC/CHOL. The first system was structured with a central circular configuration of PSM and CHOL surrounded by POPC, designed to replicate an idealized raft-like domain [4, 15, 12]. The second system featured a stochastic distribution of DPPC, DLIPC, and CHOL, aimed at exploring lipid behavior in a nonordered environment. These models facilitate the study of lipids dynamics and the structural properties of different membrane configurations.

Molecular dynamics simulations were performed with Gromacs software (v. 2022) [1] and the CHARMM36 force field [18]. The simulation protocol involved energy minimization, NVT and NPT equilibration (constant number of particles, volume/pressure, and temperature), ending with a production run. All

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_13

simulations were performed at 295.15 K, using standard CHARMM protocol. A detailed description of the simulation procedure is described in our previous work [14]. The production run involved at least 500ns.

2.2 Lipid features and Machine learning techniques

In our molecular dynamics simulations, we analyzed the last 50 ns of each trajectory to extract ten different lipid features that characterize the local membrane environment. These features included lipid type (1), area per lipid (APL, 1), total lipid length (2), the number of surrounding lipids by each type, first shell (3), and composition of the second shell (3).

APL was determined using a custom MATLAB script. For each lipid, a position point was identified at the midpoint between the phosphorus atom (P) and the second carbon atom (C2) in the phospholipids, and between the C3 and O3 atoms for CHOL. The local environment of each lipid was further quantified by Voronoi tessellation to determine the composition of the first and second lipid shells. Additionally, the order parameter (S_{cd}) for acyl chains was calculated using the MEMBPLUGIN [8] S_{cd} addon in Visual Molecular Dynamics (VMD) software, averaging the values for both acyl chains of each lipid. The lengths of the lipids were measured as the geometric distance between the phosphorus and last carbon atoms in the acyl chain.

Statistical analysis was performed using the Python Scipy library, employing a one-way ANOVA test with a significance threshold of 0.05 and the Tukey test for post hoc analysis to ensure the reliability of our findings. We employed a suite

Table 1. The architectures and accuracy of tested ML models

ML model	Architecture	Best accuracy
KNN	5 closest neighbors	91% ± 3%
DNN	10x128x128x32x8x4x3, ReLu activation, 20% dropout, softmax without dropout (last layer)	87% ± 5%
RF	entropy criterion, 1,000 estimators, max. depth 15	88% ± 2%
SVC	polynomial kernel	85% ± 3%

of ML techniques to unravel the complex organization of microdomains and lipid membranes, using different algorithms to categorize membrane compositions as raft/domain, nondomain, or interface. Initially, we normalized all input data and applied one-hot encoding to categorical variables, ensuring that our dataset was optimized for ML processing. Subsequently, several selected ML algorithms from the scikit-learn package were adopted: Support Vector Classifier (SVC), Random Forest (RF), K-Nearest Neighbors (KNN), and Deep Neural Network (DNN). The models were trained on a random selection of 80% of data, and the remaining 20% of data was used to determine an accuracy metric for the method. The choice of an equal lipid composition (1:1:1) ensures a balanced distribution of features in the test dataset. In this article, we have selected and presented one

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_13

ML technique for each approach with the highest accuracy of all the methods listed. These were KNN and RF for the first and second approaches, respectively. All architectures are presented in Table 1.

3 Results and discussion

In this study, we explored two scenarios of lipid microdomain formation using molecular dynamics simulations. The first scenario involved an idealized domain with a core of CHOL and PSM, surrounded by unsaturated POPC lipids, showing slow lipid migration and interfacial mixing [12, 4]. The second scenario explores the spontaneous segregation of saturated DPPC and unsaturated DLIPC lipids facilitated by the addition of CHOL. We used the L_o and L_d nomenclature for these domains, observing how DPPC's high phase-transition temperature influenced behavior differently from the POPC/PSM/CHOL mixture. Our simulations, which lasted approximately 2 μs, highlighted the complex and nuanced domain formation process, involving a more probabilistic gel domain (S_o) formation compared to the previous model. Using ML techniques, the selected parameters of all lipids were used to classify them into three distinct clusters representing raft-like / domain, nonraft / nondomain, and interface regions. The results of selected ML approaches are shown in Figure 1.

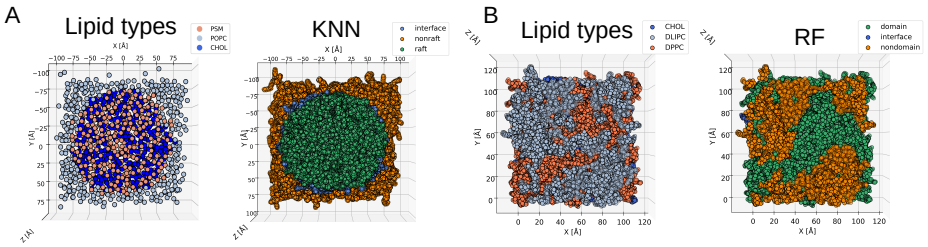

Fig. 1. Classification of (A) POPC/PSM/CHOL and (B) DPPC/DLIPC/CHOL membrane regions using KNN and RF methods, respectively based on selected lipid parameters (top view). Each left panel illustrates the structural composition of the lipid membrane and the right panel depicts the ML lipid classification.

The organization of the idealized raft system is presented in Figure 2A. The raft, nonraft, and interfacial areas constitute 55%, 31%, and 13% of the system, respectively. This distribution is consistent with experimental studies on POPC/PSM/CHOL vesicles [13]. The formation of these domains is influenced by temperature variations, which often leads to a predominance of L_o regions due to the affinity of CHOL for PSM [20]. The interface mainly comprises PSM, CHOL (30%), and POPC (22%), while nonraft areas are dominated by CHOL and PSM with minimal POPC presence (see Figure 2C). The nonraft region is predominantly composed of POPC (92%), with minor contributions from PSM

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_13

(5%) and CHOL (3%). These distributions deviate slightly from the literature, where typically the L_d phase contains about 71% DOPC, 24% PSM, and 5% CHOL [3]. Our findings for the L_o phase align with the trends for CHOL and PSM but with a slightly lower PC representation. The discrepancies may be attributed to variations in lipid ratios (43:32:25) in both our study and the presence of more unsaturated DOPC, which supports phase separation [3]. It is worth noting that the experimental models hardly quantify the transition phase, thus a binary classification in ML might mirror the experimental outcomes closely.

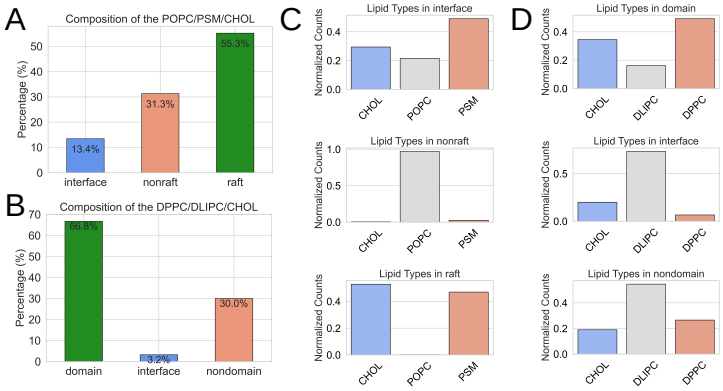

Fig. 2. Composition of (A) POPC/PSM/CHOL and (B) DPPC/DLIPC/CHOL with applied ML classifiers, lipid species distribution in recognized domains in (C) POPC/PSM/CHOL and (D) DPPC/DLIPC/CHOL systems.

In a DPPC/DLIPC/CHOL mixture, the distribution of lipids and their phases are less distinct (see Figure 1B). The domains are dispersed in a lattice pattern, highlighting a more complex and heterogeneous assembly of lipid phases. Interestingly, the location of CHOL molecules can be seen mainly at the interface between DPPC and DLIPC (see Figure 2B). Here, about 67% of the system is L_o, 30% is L_d, and the interfacial area comprises 3.2%, indicating the challenges in identifying transition zones. Domain compositions show a balanced distribution, with the nondomain phase mainly consisting of DLIPC, supplemented by about 30% DPPC and 20% CHOL. In contrast, the L_o phase predominantly includes DPPC and CHOL, with 15% of DLIPC (see Figure 2D). This distribution and composition of these domains align with other experimental studies [2, 15].

Additionally, we used ML predictions to analyze specific lipids based on their molecular organization, focusing on the area per lipid (APL) and order parameters. In the rafts, typically enriched with PSM and CHOL, we found a decrease in APL. This suggests a tighter lipid packing and a consequent increase in the order parameter, reflecting the dense lipid packing. The L_d phases, which contain more unsaturated lipids such as POPC, tend to indicate higher APL values and lower order parameters due to looser acyl chain packing and increased dis-

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_13

order [17]. Our findings, illustrated in Figure 3A, confirm these patterns. Lipids in the raft phase consistently exhibited the lowest APL values, whereas those in the nondomain regions showed the highest. Interestingly, CHOL in the nonraft phase displayed unusually low APL values, probably distorted due to its limited occurrence and not entirely indicative of the overall behavior. Intermediate APL values in the interface region support its role as a transitional area between the L_o and L_d phases, reflecting the characteristics of both phases.

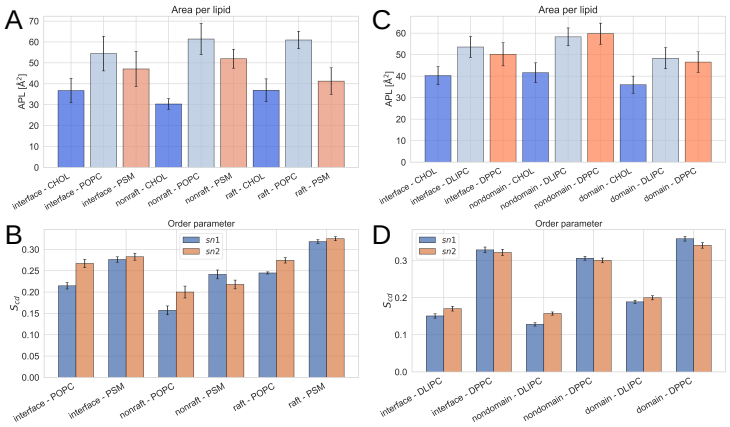

Fig. 3. Quantitative analysis of lipid organization in (A, B) POPC/PSM/CHOL and (C, D) DPPC/DLIPC/CHOL. Changes in APL (A, C) and order parameters (B, D) are displayed for all identified regions.

DPPC/DLIPC/CHOL observations are consistent with previous studies, indicating the highest APL values in nondomain areas and the lowest in domain areas [15]. The slight effect of the APL values of the interface region on the ML predictions suggests a diminished role for this parameter in the phase identification process.

The order parameter reflects the orientation of the lipid acyl chains in the membrane, with higher values indicating rigid, straight chains typical of the L_o phases. Our results show the highest order parameters for PSM and POPC in the raft domains and the lowest in the nondomain areas, with values in the interfacial zones falling between, showing a smooth transition from ordered to disordered regions (see Figure 3B). Within the raft phase, there is a notable reduction in APL for POPC. This pattern is mirrored by the PSM, where the decrease in APL corresponds to an increase in the order of the lipid tails, highlighting the effect of CHOL on the behavior of associated lipids. This supports previous findings of the limited effect of PC on disorder in the L_o phase [4]. The interface serves as a transition zone with reduced POPC and PSM order parameters. It should be noted that the $sn1$ tail of POPC shows a significant deviation in order from its unsaturated $sn2$ tail. The APL values for POPC in nondomain areas align with

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_13

those of pure PC systems, whereas an increase is observed in PSM [10]. These variations imply a significant influence of PC on neighboring lipids, accompanied by a decrease in PSM order and a shift between the $sn1$ and $sn2$ tails. The role of CHOL in these organizational changes appears to be minimal. In the second system, we observe a consistent trend in lipid tail ordering and APL with the highest order of DPPC in the domain phase and the lowest in the L_d phase (see Figure 3 C, D). This trend is also apparent with DLIPC. These differences are more nuanced than previously noted, particularly in distinguishing between the interface and nondomain phases. Integration of DLIPC into the domain phase showed minimal disruption, indicating good compatibility with DPPC and CHOL [9]. The interface is heavily affected by DLIPC, which makes it akin to a nondomain because of its liquid fraction. In nondomain areas, the DPPC-CHOL combination does not significantly alter the DLIPC tail ordering. These findings highlight CHOL's pronounced effect on ordering DPPC over DLIPC, suggesting that the domain more closely resembles the L_o phase, contrary to initial expectations of a S_o phase.

4 Conclusions

In this study, we investigate the dynamics and organization of lipid membranes, focusing on the formation and characterization of lipid microdomains. We used molecular dynamics simulations and a suite of machine learning techniques to analyze and classify membrane regions into ordered, disordered, or interfacial states. Our findings highlight the dynamic interaction between the L_o and L_d phases, with cholesterol playing a crucial role in the formation of the L_o domains. We examine the organization of the membrane in different mixtures: POPC/PSM/CHOL and DPPC/DLIPC/CHOL to understand the behavior in varying compositional contexts. Idealized circular rafts were identified better than spontaneously induced ones. However, it is worth noting that in both systems the domain compositions were consistent with literature reports. The dependence on the initial training system's composition is a main limitation that could be reduced through experimentally supported training preparations. Further development of more sophisticated machine learning models capable of integrating multiscale data will enhance the ability to predict membrane organization and dynamics under various physiological conditions.

Acknowledgement

M.R. thanks the National Science Centre, Poland for the financial support (grant number 2022/45/N/NZ9/02130).

References

1. Abraham, M.J., Murtola, T., Schulz, R.: Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX **1-2**, 19–25 (2015)

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_13

2. Barnoud, J., Rossi, G., Marrink, S.J., Monticelli, L.: Hydrophobic Compounds Reshape Membrane Domains. PLoS Comput. Biol. **10**(10), e1003873 (2014)
3. Bezlyepkina, N., Gracià, R.S., Shchelokovskyy, P., Lipowsky, R., Dimova, R.: Phase diagram and tie-line determination for the ternary mixture DOPC/eSM/Cholesterol. Biophys J **104**(7), 1456–1464 (2013)
4. Canner, S.W., Feller, S.E., Wassall, S.R.: Molecular Organization of a Raft-like Domain in a Polyunsaturated Phospholipid Bilayer: A Supervised Machine Learning Analysis of Molecular Dynamics Simulations. J. Phys. Chem. B **125**(48), 13158–13167 (2021)
5. Chichili, G.R., Rodgers, W.: Cytoskeleton–membrane interactions in membrane raft structure. Cell. Mol. Life Sci. 2009 66:14 **66**(14), 2319–2328 (2009)
6. Cournia, Z., Allen, T.W., Andricioaei, I., Antonny, B., Bondar, A.N.: Membrane Protein Structure, Function, and Dynamics: a Perspective from Experiments and Theory. The J. Membr. Biol.2015 248:4 **248**(4), 611–640 (2015)
7. Drabik, D., Drab, M., Penič, S., Iglič, A., Czogalla, A.: Investigation of nano- and microdomains formed by ceramide 1 phosphate in lipid bilayers. Sci. Rep. 2023 13:1 **13**(1), 1–14 (2023)
8. Guixà-González, R., et al.: MEMBPLUGIN: studying membrane complexity in VMD. Bioinformatics **30**(10), 1478–1480 (2014)
9. Keller, F., Heuer, A.: Chain ordering of phospholipids in membranes containing cholesterol: what matters? Soft Matter **17**(25), 6098–6108 (2021)
10. Leftin, A., Molugu, T.R., Job, C., Beyer, K., Brown, M.F.: Area per Lipid and Cholesterol Interactions in Membranes from Separated Local-Field 13C NMR Spectroscopy. Biophys J **107**(10), 2274 (2014)
11. Löpez, C.A., Vesselinov, V.V., Gnanakaran, S., Alexandrov, B.S.: Unsupervised Machine Learning for Analysis of Phase Separation in Ternary Lipid Mixture. J. Chem. Theory Comput. **15**(11), 6343–6357 (2019)
12. Peter, C., Kremer, K., Carbone, P., Niemel, P.: Concerted diffusion of lipids in raft-like membranes. Faraday Discuss. **144**(0), 411–430 (2009)
13. Pokorny, A., Yandek, L.E., Elegbede, A.I., Hinderliter, A., Almeida, P.F.F.: Temperature and Composition Dependence of the Interaction of d-Lysin with Ternary Mixtures of Sphingomyelin/Cholesterol/POPC. Biophys J **91**, 2184–2197 (2006)
14. Rzycki, M., Drabik, D., Szostak-Paluch, K., Hanus-Lorenz, B., Kraszewski, S.: Unraveling the mechanism of octenidine and chlorhexidine on membranes: Does electrostatics matter? Biophys. J. **120**(16), 3392–3408 (2021)
15. Sodt, A.J., Sandar, M.L., Gawrisch, K., Pastor, R.W., Lyman, E.: The molecular structure of the liquid-ordered phase of lipid bilayers. J. Am. Chem. Soc. **136**(2), 725–732 (2014)
16. Staubach, S., Hanisch, F.G.: Lipid rafts: signaling and sorting platforms of cells and their roles in cancer. Expert Rev. Proteomics **8**(2), 263–277 (2011)
17. Veatch, S.L., Keller, S.L.: Miscibility phase diagrams of giant vesicles containing sphingomyelin. Phys. Rev. Lett. **94**(14) (2005)
18. Wu, E.L., et al.: CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. J. Comput. Chem. **35**(27), 1997–2004 (2014)
19. Yang, J., Martí, J., Calero, C.: Pair interactions among ternary DPPC/POPC/cholesterol mixtures in liquid-ordered and liquid-disordered phases. Soft Matter **12**(20), 4557–4561 (2016)
20. Yasuda, T., Tsuchikawa, H., Murata, M., Matsumori, N.: Deuterium NMR of Raft Model Membranes Reveals Domain-Specific Order Profiles and Compositional Distribution. Biophys J **108**(10), 2502–2506 (2015)

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_13