

THE UNIVERSITY OF CHICAGO

PROTEIN-LIPID INTERACTIONS IN LIPID DROPLETS

A DISSERTATION SUBMITTED TO  
THE FACULTY OF THE PRITZKER SCHOOL OF MOLECULAR ENGINEERING  
IN CANDIDACY FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

BY

SIYOUNG KIM

CHICAGO, ILLINOIS

DECEMBER 2021

Copyright © 2021 by Siyoung Kim

All rights reserved

To my parents

## TABLE OF CONTENTS

<b>LIST OF FIGURES .....</b>	<b>viii</b>
<b>LIST OF TABLES .....</b>	<b>xii</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>xiii</b>
<b>ABSTRACT.....</b>	<b>xv</b>
<b>Chapter 1 Introduction .....</b>	<b>1</b>
<b>Chapter 2 STR2GMX: A Python package for the creation of GROMACS-formatted CHARMM force field.....</b>	<b>12</b>
<b>ABSTRACT.....</b>	<b>12</b>
<b>INTRODUCTION .....</b>	<b>13</b>
<b>METHODS .....</b>	<b>16</b>
<b>RESULTS .....</b>	<b>19</b>
<b>DISCUSSION.....</b>	<b>24</b>
<b>CONCLUSIONS .....</b>	<b>25</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>26</b>
<b>SUPPORTING INFORMATION.....</b>	<b>27</b>
<b>Chapter 3 The Surface and Hydration Properties of Lipid Droplets .....</b>	<b>29</b>
<b>ABSTRACT.....</b>	<b>29</b>
<b>STATEMENT OF SIGNIFICANCE.....</b>	<b>30</b>

INTRODUCTION .....	30
METHODS .....	33
RESULTS AND DISCUSSION.....	42
CONCLUSIONS .....	49
ACKNOWLEDGEMENTS.....	53
<b>Chapter 4 Determinants of endoplasmic reticulum-to-lipid droplet protein targeting .....</b>	<b>66</b>
ABSTRACT.....	66
INTRODUCTION .....	67
RESULTS .....	68
DISCUSSION.....	79
METHODS .....	82
ACKNOWLEDGEMENTS.....	87
<b>Chapter 5 Stressed Lipid Droplets: How Neutral Lipids Relieve Surface Tension and Membrane Expansion Drives Protein Association .....</b>	<b>104</b>
ABSTRACT.....	104
INTRODUCTION .....	105
METHODS .....	110
RESULTS AND DISCUSSION.....	118
CONCLUSIONS .....	138
ACKNOWLEDGEMENTS.....	141

## **Chapter 6 Physical Characterization of Triolein and Implications for Its Role in Lipid**

<b>Droplet Biogenesis.....</b>	<b>146</b>
ABSTRACT.....	146
INTRODUCTION .....	147
METHODS .....	150
RESULTS .....	157
DISCUSSION.....	168
CONCLUSIONS .....	171
ACKNOWLEDGEMENTS.....	171

## **Chapter 7 Membrane rigidity as a key factor governing initial stages of lipid droplet**

<b>formation .....</b>	<b>187</b>
ABSTRACT.....	187
INTRODUCTION .....	188
METHODS .....	189
RESULTS .....	194
DISCUSSION.....	201
CONCLUSIONS .....	203
ACKNOWLEDGEMENTS.....	203
SUPPORTING INFORMATION.....	209

<b>Chapter 8 Seipin transmembrane segments function in triglyceride nucleation and droplet budding from the membrane .....</b>	<b>212</b>
ABSTRACT.....	212
SIGNIFICANCE STATEMENT .....	213
INTRODUCTION .....	213
RESULTS .....	215
DISCUSSION.....	222
METHODS .....	225
ACKNOWLEDGEMENTS.....	229
<b>FUTURE DIRECTIONS.....</b>	<b>241</b>
<b>REFERENCES.....</b>	<b>246</b>

## LIST OF FIGURES

<b>Figure 1-1</b> Illustration of cellular organelles of interest (left). While the other organelles are bilayer-bounded, lipid droplets are monolayer-bounded. Metabolic pathway of triacylglycerol (TG) and phosphatidylcholine (PC) synthesis with key proteins (right). .....	4
<b>Figure 2-1</b> Illustration of the workflow to create GROMACS-formatted CHARMM force field. ....	15
<b>Figure 2-2</b> Solvation of trialanine. ....	20
<b>Figure 2-3</b> System setup and the creation of GROMACS inputs of a LD system, containing <i>LiveDrop</i> and CCT. ....	22
<b>Figure 2-4</b> Verification of STR2GMX. ....	23
<b>Figure 2-5</b> Workflow to create a lipoprotein using STR2GMX. ....	27
<b>Figure 3-1</b> LD structure and density profile. ....	54
<b>Figure 3-2</b> Packing defects of ERbil and LDs. ....	55
<b>Figure 3-3</b> The order parameters of POPC, DOPE, and TG molecules in POPCbil, DOPEbil, ERbil, and 8nm LD. ....	56
<b>Figure 3-4</b> The area per phospholipid (black line) and the amount of interdigitation of PLs by SURF-TG (green dashed line) for 8nm LD. For comparison, the APLs of POPCbil, DOPEbil, and ERbil were 0.645 nm <sup>2</sup> , 0.615 nm <sup>2</sup> , and 0.628 nm <sup>2</sup> , respectively. ....	57
<b>Figure 3-5</b> The water density profiles at simulation times. ....	57
<b>Figure 3-6</b> Water permeation into LD. ....	58
<b>Figure 3-7</b> The molecular structures of POPC, DOPE, and TG. ....	59
<b>Figure 3-8</b> The illustration of the packing defect analysis. ....	60
<b>Figure 3-9</b> The structures of the last snapshots of 8nm LD and 16nm LD. ....	61
<b>Figure 3-10</b> The time-averaged density and overlap parameter profiles for (a) 8nm LD and (b) 16nm LD. ....	62
<b>Figure 3-11</b> The amount of interdigitation of PLs by TG (purple line), CORE-TG (orange line), and SURF-TG (green line) and the SURF-TG % (red dashed line) with simulation time for 8nm LD. ....	63
<b>Figure 3-12</b> The density of PL for ERbil (black solid line) and 8nm LD (red dashed line) and PL + TG for 8nm LD (red solid line), averaged from 8 $\mu$ s to 10 $\mu$ s. The average Z position of the phosphorus atoms of the upper leaflet was zeroed. Water is toward (+) direction and the membrane center toward (-) direction. ....	64
<b>Figure 3-13</b> The density of water (blue) and TG oxygen atoms (red) for 8nm LD (top) and 16nm LD (bottom), averaged from 8 $\mu$ s to 10 $\mu$ s. The center of the phosphorous atoms (similarly the center of a LD) was zeroed. ....	65
<b>Figure 4-1</b> The Hydrophobic, Membrane-Embedded Motif <i>LiveDrop</i> Accumulates on Nascent LDs. ....	89



<b>Figure 4-2</b> LD Accumulation of <i>LiveDrop</i> Requires Specific Sequence Features .....	91
<b>Figure 4-3</b> Simulations of <i>LiveDrop</i> Suggest Conformational Changes and Energetic Contributions of Specific Sequence Features to LD Targeting .....	92
<b>Figure 4-4</b> Positively Charged Residues and Tryptophan Residues Are Required for Efficient LD Targeting of <i>LiveDrop</i> in Cells .....	94
<b>Figure 4-5</b> The ALG14 Hairpin Is Sufficient to Target LDs and It Relies on Similar LD Targeting Features as <i>LiveDrop</i> .....	96
<b>Figure 4-6</b> The Distribution of Tryptophan and Positively Charged Residues Differentiates LD Targeting from Non-LD Targeting Hairpin Motifs .....	98
<b>Figure 4-7</b> Umbrella sampling simulations of <i>LiveDrop</i> and free energy profiles of single amino acids .....	100
<b>Figure 4-8</b> Thermodynamic cycle for the membrane insertion of a single amino acid .....	102
<b>Figure 5-1</b> Bilayer and LD under the constant surface tension.....	119
<b>Figure 5-2</b> Packing defect constants ( $\pi$ ) of PL acyl, TG acyl, and TG glycerol defects for BI0 (orange), B1e (red), LD0 (light blue), and LDe (dark blue).....	120
<b>Figure 5-3</b> NVT simulation of the stressed LD (LDe).....	121
<b>Figure 5-4</b> Transition PMFs for PLs (continuous lines) and TG molecules (lines with squares) in the unstressed LD (gray) and stressed LD (black) from unbiased NVT simulations. ....	122
<b>Figure 5-5</b> Order parameters of POPC (top) and DOPE (bottom) from BI0 (orange), B1e (red), LD0 (light blue), and LDe (dark blue). Gray and black squares indicate the order parameters of SURF-TG from LD0 and LDe, respectively.....	123
<b>Figure 5-6</b> CCT $\alpha$ and the AI helix. ....	125
<b>Figure 5-7</b> Binding success (occurrence) of the AI helix (wildtype in black, mutant in crosshatch) to four different surfaces. For each membrane, the simulations were replicated 13-14 times. ..	127
<b>Figure 5-8</b> Unsuccessful binding of the AI helix in BI0.....	129
<b>Figure 5-9</b> Successful binding of the AI helix in B1e. ....	130
<b>Figure 5-10</b> Successful binding of the AI helix in LDe.....	131
<b>Figure 5-11</b> Molecular groupings (MG) of all-atom trajectories to analyze the normalized coordination number. ....	135
<b>Figure 5-12</b> PMF (blue) of the AI helix binding to the LDe surface, obtained with the umbrella sampling simulations. ....	136
<b>Figure 5-13</b> Normalized coordination number between Val278 and each membrane group, averaged over the mutant trajectories (W278V) that have complete binding with the stressed LD. ....	136
<b>Figure 5-14</b> Normalized joint probability distribution of the defect areas, $a_{PL}$ and $a_{TG}$ that Trp278 (a) or Phe285 (b) associates with, given that each residue is completely associated. ....	137

<b>Figure 5-15</b> System setup for the unstressed bilayer (BI0), stressed bilayer (BIe), unstressed LD (LD0), and stressed LD (LDe).....	142
<b>Figure 5-16</b> Snapshots and defects of the unstressed bilayer and LD. ....	143
<b>Figure 5-17</b> Density profile of the bilayer membranes (top) and LDs (bottom).....	144
<b>Figure 5-18</b> Normalized coordination number between tryptophan and phenylalanine residues and each membrane group in stressed bilayers (left) and stressed LDs (right), averaged over the trajectories that have the complete binding. ....	145
<b>Figure 6-1</b> The glycerol moiety of three TG models with partial charges. The revised TG model (C36/LJ-PME-r) has a significantly reduced charge distribution in the ester group (shown in red). .....	173
<b>Figure 6-2</b> Physical properties of the bulk TG evaluated with three different TG models.....	174
<b>Figure 6-3</b> PMF of a TG molecule in a POPC bilayer as a function of the Z distance from the bilayer center (TG flip-flop), calculated from the REUS simulations.....	174
<b>Figure 6-4</b> Height field, $z(x, y)$ , of the phosphorus atoms of the upper leaflet respective to the glycerol moiety of the SURF-TG molecule (circle at the origin). ....	175
<b>Figure 6-5</b> Order parameters of <i>sn-1</i> (top) and <i>sn-2</i> (bottom) chains. ....	176
<b>Figure 6-6</b> The RDFs of TG oxygen atoms (O11, O21, O31) with the oxygen atom of water. ....	176
<b>Figure 6-7</b> Bending modulus of five POPC bilayer membranes, containing different amounts of DAG or TG. ....	177
<b>Figure 6-8</b> Buckled POPC+DAG membrane (DAG 30%). ....	178
<b>Figure 6-9</b> CG simulations of a symmetric bilayer and TG nucleation. ....	179
<b>Figure 6-10</b> CG simulation of an asymmetric bilayer and membrane bending toward the lower leaflet.....	180
<b>Figure 6-11</b> Initial (top) and 28.3 ns (bottom) snapshots of a bilayer membrane.....	181
<b>Figure 6-12</b> Description of the CG model. ....	182
<b>Figure 6-13</b> Comparison of the 2-dimensional radial distribution functions computed from the CG (red) and CG mapped atomistic (black) trajectories for DOPC.....	183
<b>Figure 6-14</b> Comparison of the radial distribution functions computed from the CG (red) and CG mapped atomistic (black) trajectories for TG. ....	184
<b>Figure 6-15</b> Comparison of the radial distribution functions of the TG glycerol moiety computed from CG mapped atomistic trajectories of three different TG models. ....	184
<b>Figure 6-16</b> PMF obtained with REUS (orange) and TT-MetaD (black).....	185
<b>Figure 6-17</b> Orientation and position of TG molecules in a bilayer membrane. ....	185
<b>Figure 6-18</b> Height field (left) and DAG distribution (right) of the lower leaflet of the DAG 30% membrane.....	186
<b>Figure 6-19</b> TG flip-flop PMF comparison.....	186

<b>Figure 7-1</b> TG concentration-dependent nucleation. ....	204
<b>Figure 7-2</b> Ostwald ripening and PL rigidity-dependent LD shape. ....	205
<b>Figure 7-3</b> PL density-dependent budding. ....	206
<b>Figure 7-4</b> TG nucleation PMF. ....	207
<b>Figure 7-5</b> CG mapping of a POPC molecule into a 4-bead PL model to calculate order parameters (Table 7-2) and distribution (Fig. 7-7). ....	209
<b>Figure 7-6</b> A sharp switching function to compute the coordination number. ....	210
<b>Figure 7-7</b> Distribution (number) of CG beads in the bilayer normal. ....	210
<b>Figure 7-8</b> Vesicular system consisting of a 1:1 ratio mixture between soft and stiff PLs. ....	211
<b>Figure 8-1</b> Structural model of human seipin. ....	230
<b>Figure 8-2</b> Seipin hydrophobic helix and transmembrane segments attract TG. ....	231
<b>Figure 8-3</b> Flexibility of seipin TM segments increases lipid permeability. ....	232
<b>Figure 8-4</b> Position-dependent diffusion coefficient. The center of the mass of the luminal domain of seipin is at origin. ....	233
<b>Figure 8-5</b> CG models of seipin and lipids. ....	233
<b>Figure 8-6</b> Seipin lowers the critical concentration of TG nucleation. ....	234
<b>Figure 8-7</b> Cage-like geometry of seipin TM segments is key to modulating the morphology of a forming oil lens. ....	235
<b>Figure 8-8</b> CG-MD shows LD growth in a large bilayer with its curvature comparable to the actual ER tubule. ....	236
<b>Figure 8-9</b> Cryoelectron microscopy of human seipin. ....	237
<b>Figure 8-10</b> Normalized coordination number by atom. ....	238
<b>Figure 8-11</b> Root-mean-square distance (RMSD) of the $\alpha$ -carbon atoms of the luminal domain (black) or the whole subunit (red). The error bars represent the standard deviation of the RMSDs of 11 subunits. ....	238
<b>Figure 8-12</b> Root-mean-square fluctuation (RMSF) of the $\alpha$ -carbon atoms. The error bars represent the standard error of the RMSFs of three blocks. ....	239
<b>Figure 8-13</b> Mean squared distance of the 20 luminal PLs, trapped inside the hydrophobic helix. ....	239
<b>Figure 8-14</b> CG simulation of the bilayer containing 6% TG with a diameter of 40 nm. The ENM of seipin used a spring constant of 0.2 kcal/mol/ $\text{\AA}^2$ . The diameter of an oil lens is shown with simulation steps. ....	239
<b>Figure 8-15</b> CG simulation of the bilayer containing 6% TG with a diameter of 40 nm. ....	240

## LIST OF TABLES

<b>Table 3-1</b> A detailed description of the systems used in this work.....	35
<b>Table 4-1</b> Summary of simulations and lipid compositions used .....	103
<b>Table 5-1</b> Description of simulations. ....	113
<b>Table 6-1</b> Description of AA simulations. ....	151
<b>Table 6-2</b> Description of CG simulations. ....	156
<b>Table 6-3</b> Physical properties of DOPC bilayer and TG bulk from experiments, AA simulations, and CG simulations. Standard errors are given for the simulations that we performed. ....	165
<b>Table 7-1</b> Description of CG simulations. ....	192
<b>Table 7-2</b> Physical properties of the POPC bilayer from the AA trajectory and the PL bilayers from the CG trajectories with two different angle potential parameters. Standard errors are given in parentheses. ....	196

## ACKNOWLEDGEMENTS

Most importantly, I would like to thank my advisor, Gregory A. Voth, for providing motivation and resources for my Ph.D. research. His pursuit of scientific excellence has always inspired me, and I have benefited greatly from his generous support. Past five years in his group has been a delightful and rewarding experience both personally and academically. I would also like to express my gratitude to Andrew L. Ferguson and Allison H. Squires for serving on my dissertation committee.

I am also thankful to Jessica M. J. Swanson. She is an extremely insightful, encouraging, supportive, and talented scientist and mentor. It has been a great joy to work with her and her group members. Moreover, I am grateful to Myong In Oh for the collaboration, fruitful discussions, and critical inputs on the research topics.

I am incredibly appreciative of Robert V. Farese, Jr. and Tobias C. Walther and their group members. It has been a wonderful learning opportunity to collaborate with such talented experimentalists. Their feedbacks and inputs on computational simulations have been critical to my research. Especially, I am sincerely grateful to Jeeyun Chung, a meticulous and brilliant scientist, for her extensive help in my understanding of the experiments and their implications. I have enjoyed my time working with them, which has become a very productive collaboration that aims to understand the biology and biophysics of lipid droplets.

I want to express my sincere gratitude to many great scientists who have helped me during my graduate school years. I cannot thank Chenghan Li and He Ma enough for their help in every aspect of my scientific research: from writing scientific scripts to theoretical discussions. I am also grateful to Marton Voros and Federico Giberti for their mentorship. Although we interacted together only for a short amount of time, their guidance still pushes me to become a better scientist.

I am also thankful to Richard W. Pastor who was kind enough to explain the details of lipid force fields and the analysis of simulations. I would like to thank Wooje Cho for the previous collaboration. I am very appreciative of Harry Ryu, a good friend of mine, for thoroughly reviewing many of my writings. Moreover, I appreciate the critical discussions with the other Voth group members, both previous or current: Morris E. Sharp, Patrick G. Sahrman, Sriramvignesh Mani, Alexander J. Pak, Viviana Monje-Galvan, Timothy Loose, Zack Jarin, Joshua Zuchniarz, Jeri Beiter, Daniel Beckett, Vilmos Zsolnay, Fikret Aydin, and Yuxing Peng.

My Chicago life would have been colder without the Howorth family. I am very grateful to David, Mary, James, and Hannah. Their extraordinary hospitality warms my heart, burns away the fatigue, and leaves behind lingering gratitude. It has been such a pleasure living with them in Forest Park. I am also humbly grateful to the music team of the Fourth Presbyterian Church. It was an unforgettable and pleasant experience singing in the Chancel Choir. I especially thank John W. W. Sherer, the Organist and Director of Music at the Church, for his stunning performances and musical leadership.

Finally, I would like to thank my parents, Jungho Kim and Younga Shin. I am surprised every day by how deep their love is. I am grateful for their unconditional love and unlimited support. I would not be here today without them.

## ABSTRACT

Lipid droplets (LDs) are intracellular organelles whose primary function is energy storage. Known to emerge from the endoplasmic reticulum (ER) bilayer, LDs have a unique structure in which their core consists of neutral lipids, triacylglycerol (TG) or sterol ester for example, surrounded by numerous proteins and a phospholipid (PL) monolayer. Despite their importance in metabolism and metabolic diseases such as obesity, type II diabetes, or lipodystrophy, LDs have not garnered the attention that they deserve. I have used both the all-atom (AA) and coarse-grained (CG) simulations of LDs and the relevant proteins along with experimental collaboration to address a broad range of questions. The thesis will consist of exploration of five topics related to LD biology and biophysics: characterization of LD surfaces and comparison to bilayers, ER-to-LD targeting, Cytosol-to-LD targeting, biophysics of LD biogenesis, and LD biogenesis orchestrated by the ER protein seipin. The key finding of the first three parts is that LD surfaces can be distinguished from the ER bilayers due to the TG glycerol moiety exposed to the cytosol. The TG glycerol moiety exhibited at the surface can form hydrogen bonds with some protein residues (e.g., tryptophan), working as a peptide targeting mediator. I have drawn a comparable conclusion in the ER bilayer where some protein residues embedded in the hydrophobic phase (e.g., serine) preferentially attract TG, working as a TG tethering site. Finally, the CG simulations have shown TG nucleation and LD growth by increasing accessible time and length scales. The central conclusion from the CG simulations is that the transmembrane segments of seipin, critical for its function, surround an oil lens and create a unique ER-LD neck structure.

## Chapter 1 Introduction

Lipids are biomolecules characterized by high solubility in nonpolar solvents. They play critical cellular roles in forming barriers between subcellular compartments, hosting proteins, and storing excess energy. Importantly, lipids define the boundary of cells, protecting cells from the surrounding environment and regulating the flux of substances.

A broad range of lipids is present in cells, including but not limited to sterols, fatty acids, phospholipids (PLs), monoacylglycerols, diacylglycerols (DAGs), and triacylglycerols (TGs). PL is the primary lipid class that provides structural support for membranes and forms the matrix of membranes. PLs usually have a polar head group, with either a net charge or a dipole and two acyl chains. Hence, PLs can be classified as an amphiphile. Therefore, in cells, PLs form a bilayer structure in which their head groups are exposed to water (the main constituent of cells or any living systems), and their acyl chains extend towards the bilayer center. Such a bilayer membrane structure can efficiently regulate the passive diffusion of substances. For instance, the energy barrier of permeating a water molecule to the membrane center is approximately 6 kcal/mol, approximately ten times larger than thermal energy at room temperature. In contrast, hydrophobic substances are more likely to be trapped in the hydrophobic phase. Therefore, it is of pharmaceutical interest to develop drug delivery systems that efficiently transport therapeutic materials to the interior of targeted cells.

Cells consist of diverse types and distributions of lipids. Organelles or cell membranes usually have more than hundreds of different types of lipids. Not only that, different organelles, cell types, species, and even positions within the bilayer membrane display different lipid compositions. This highly diverse chemical structure and distribution of lipids enable cells to have physically distinct lipid domains and respond to highly selective biological signals. For instance,



the cell membranes (or plasma membranes) contain lipid microdomains enriched with sphingomyelin and cholesterol, referred to as lipid rafts. They are characterized by higher orders and packing of lipids and longer hydrophobic thickness than the surroundings. They also have a high population of negatively charged PLs such as phosphatidylserine (PS) and phosphoinositide phosphatidylinositol 4,5-bisphosphate (PIP<sub>2</sub>) in the inner leaflet of the cell membrane. In contrast, the ER contains lower amounts of those lipids, which increases its membrane flexibility and reduces the net charge of the membrane.

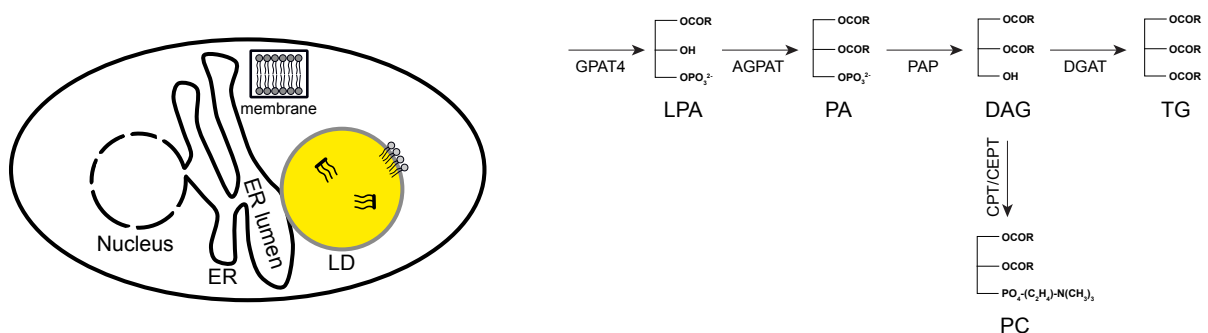
These distinct physical properties of the bilayers determine which cytosolic proteins interact with the bilayer. For instance, the amphipathic lipid packing sensor (ALPS) of ADP-ribosylation factor GTPase-activating protein 1 (ArfGAP1) senses packing defects-rich bilayers such as the ER bilayer or Golgi apparatus. In particular, the ALPS motif is known to bind to a highly curved surface with a radius of 50 nm or lower. The curvature sensitivity stems from the charge difference between the polar face and the other amphipathic helices in which the polar face has more uncharged residues such as serine and threonine. Consequently, the weaker interaction between the membrane head group and the polar face of the helix increases the importance of the hydrophobic interaction between the hydrophobic face of the helix and the membrane tail region. Therefore, if a membrane is curved, it is more likely for the hydrophobic phase to be more exposed to the surface, facilitating the ALPS targeting. The other example of curvature-sensing protein domains is Bin/amphiphysin/Rvs (BAR) domains. Even though the BAR domains preferentially target curved surfaces, they do not bind to the ER tubule, which is typically has a high curvature with a radius of ~30 nm. A propensity of the BAR domains for the cell membrane over the ER bilayer stems from the high positive charge of the BAR domains. The plasma membrane has

abundant negatively charged PLs, and the strong electrostatic interaction between the two is the primary driving force of the BAR domains' preference towards the cell membrane.

Finally, some proteins can sense the states of membranes. CTP:Phosphocholine Cytidyltransferase (CCT) catalyzes the rate-limiting step of the phosphatidylcholine (PC) synthesis. Its M domain, responsible for membrane association, targets expanding ER bilayers, enriched with PC's precursors such as DAG, phosphatidate (PA), or phosphatidylethanolamine (PE). Biophysically, this is related to the shape of lipids. PC has a cylindrical shape, which creates few packing defects, while its precursors feature a conical shape, inducing packing discontinuity. Therefore, bilayers that contain DAG, PA, or PE have high packing defects, promoting the binding of CCT. Although many aspects of peptide targeting require further research and analysis (e.g., how packing defects induced by curvature are different from those by conical lipids?), it is clear from previous experimental and computational studies that cytosolic proteins target different bilayers with distinct physical properties and lipid compositions. Interested readers should consult these several excellent review articles on the biophysics and biochemistry of membranes (1-8) and protein targeting to bilayers (9-14).

The focus of my doctorate research has been the computational study of the lipid droplet (LD). LDs are a spherical intracellular organelle whose size can range from 500 nm to 2  $\mu$ m in diameter. They have a unique structure in which their core is composed of neutral lipids such as TG or sterol esters, surrounded by a PL monolayer. This is striking because almost all the other organelles, except for lipoproteins, are surrounded by a PL bilayer. This unique topology arises from the hydrophobicity of the LD contents, which are TGs and sterol esters (Fig. 1-1). They are universal organelles but are most often found in adipose tissues/cells because their principal function is to store excessive energy into highly reduced TG and mobilize fatty acids for energy

generation or membrane biosynthesis (Fig. 1-1). The abnormal synthesis of LDs is closely related to metabolic diseases. For instance, overaccumulation of LDs, a typical symptom of obesity, leads to an increased rate of TG hydrolysis, which in turn releases a larger amount of fatty acids into the bloodstream. The elevated level of plasma free fatty acid will then be taken to the skeletal muscle, liver, and endothelial cells and synthesized into DAG or TG. DAG activates protein kinase C, and serine/threonine kinases can decrease tyrosine phosphorylation of insulin receptor substrates and hence interrupt insulin signaling. Also, TG accumulation in liver cells leads to hepatic steatosis (or fatty liver disease). However, the connection between LD biology in cells and conditions is not entirely clarified and needs further study to understand clinical implications. Intrigued readers are advised to consult several articles on the health effects of LDs (15-19).



**Figure 1-1** Illustration of cellular organelles of interest (left). While the other organelles are bilayer-bounded, lipid droplets are monolayer-bounded. Metabolic pathway of triacylglycerol (TG) and phosphatidylcholine (PC) synthesis with key proteins (right).

LDs are dynamical organelles whose sizes and numbers sensitively react to the metabolic needs of the cell. If the cell has excessive energy, fatty acids will be synthesized into TGs and stored in LDs. On the other hand, if the cell calls for energy, the TGs will be hydrolyzed to provide energy. Such lipogenesis and lipolysis are controlled by coating proteins located on the LD surface.

Therefore, a change in the metabolic need implies a change in the protein composition on the LD surface. For instance, proteins that synthesize TG (lipogenesis proteins) are usually located on the ER bilayer and re-localize to LD via ER-LD membrane bridges that form during LD growth. LD proteins that have the ER-to-LD targeting pathway mostly have hydrophobic membrane-embedded hairpin structures shown to be sufficient for LD localization in a sequence-sensitive manner. Proteins that break TGs into fatty acids (e.g., CGI-58, ATGL), protect LDs (e.g., Perilipin family), or synthesize phospholipids (e.g., CCT) target LDs from the cytosol via their amphipathic membrane targeting motifs. The reviews provide excellent overview of LD proteins (20-27) and LD targeting (28-31).

How proteins distinguish LDs from bilayer surfaces is not clearly known. A predecessor in the Voth group, Morris Sharp, has developed the general packing defect model in which neutral lipids induce larger and more persistent packing defects at LD surfaces and cytosolic proteins preferentially target LD surfaces over bilayers. However, this model does not explain differential targeting. For instance, each perilipin family member targets different LDs with distinct neutral lipid compositions: TG-abundant LDs host perilipins 1a, 1b, and 5 while sterol esters-abundant LDs do perilipins 1c and 4. Moreover, this model cannot explain why perilipins only target LDs because the ER bilayer enriched with DAG can have large and persistent packing defects. To address these questions, Jessica Swanson and I have compared the physical properties of both LDs and bilayers. Afterward, we have simulated a system of two model peptides, *LiveDrop* and the autoinhibitory motif of CTP:phosphocholine cytidyltransferase (CCT), in which the former migrates from the ER bilayer and the latter is targeted from the cytosol. We found that the specific residues (e.g., tryptophan) form hydrogen bonds with the glycerol moiety of TG that is exposed to the LD surface or buried at the LD core.

Before we move on further, it should be noted that our endeavors in studying LD targeting pathways are limited by the quality of the force fields. More specifically, the problem arises from the static charges in additive force fields. TG has a polarizable glycerol moiety that can have various charge distributions or nonbonded interaction parameters depending on the environment. For instance, bulk TG is hydrophobic. However, at the water interface, the glycerol moiety will have a high charge distribution because of its high polarizability and hydrogen bonds with water molecules. Unfortunately, additive force fields (e.g., CHARMM36m, which has been used throughout this dissertation) cannot capture this amphibious property of TG. Of course, all the other molecules also suffer from this limitation to some extent. For example, the free energy barrier for permeating a water molecule into the bilayer center is approximately 7 kcal/mol with the CHARMM36 force field. This value is an overestimation by  $\sim 1$  kcal/mol compared to the experimental value or the computational value calculated using the polarizable force fields (32). This discrepancy could become a serious issue in LD simulations because TG is frequently found at both locations: LD core and PL monolayer.

There are three important quantities one should remember when developing a TG model for LD. The first is the interfacial tension of a bulk TG at the water interface, experimentally measured to be 32 mN/m (33). The second quantity is the area per phospholipid of LDs, which is experimentally measured to be  $>14$  % higher than bilayers (34). In other words, the surface area of the LD is larger than the bilayer if the number and composition of PLs in each leaflet are the same. Exposure of TG molecules to the LD surface likely explains less PL packing at the LD surface than the bilayer. Finally, the experimental value of the LD surface tension is  $\sim 1-4$  mN/m (34, 35). This value is significant compared to the surface tension of bilayers which is typically

zero. The LD surface's high surface tension is hypothesized to stem from exposure of TG molecules, which is a worse amphiphile than PL, to the surface.

Unfortunately, none of the current force fields could reproduce all those properties. And even if the model reproduces the first quantity, the calculated second and third quantities deviate greatly from the reference values and vice versa for the second and third quantities. The poor description of water with the TIP3 model could contribute to this problem as well. I have made the first all-atom TG model (called CHARMM36 model) with the parameters taken from PL. This way, TG can be considered PL with its head group replaced by one of its tail chains. However, since the PL glycerol moiety is immersed in the polar environment, this approach gives TG a high charge distribution. Therefore, this model results in significantly higher hydration (approximately ten times larger than the experimental value) in the bulk TG or LD core and has low interfacial tension at the water interface. However, it faithfully represents the surface properties of LDs by having a significant amount of the TG glycerol moiety exposed to water. Those TG molecules work as a membrane component and increase the area per phospholipid. I also have developed the second TG model (called CHARMM36/LJPMER; Details are in Chapter 6.) that reproduces the interfacial tension against water by reducing the charge distribution of the TG glycerol moiety from the CHARMM36 model. However, the area per phospholipid of LDs compared to bilayers is small, and not enough TG molecules are present at the LD surface.

Interestingly, the coarse-grained force fields also suffer from the same issue and have very similar results to those of the all-atom models. For instance, the MARTINI force field for TG can be constructed in a similar manner by replacing the PL head group with the PL chain. Analogous to the AA model, this CG model has low interfacial tension at the TG/water interface. Also, this model does not show TG nucleation in a bilayer membrane even if the TG concentration is higher

than the critical concentration. The “secret” way to remedy this situation is to include one more hydrophobic tail atom in each acyl chain. Many MARTINI papers have not indicated this, but it has become evident after obtaining one of the structures. With this modification, the interfacial tension at the TG/water interface increases, and TG nucleates at the critical concentration. However, the area per phospholipid of the LD is nearly the same as that of the bilayer, and the LD surface has no difference from the bilayer. Also, other coarse-grained model based on Shinoda-DeVane-Klein (SDK) shares the same problem (36). The TG model made by replacing a PL head group with a PL acyl chain has low interfacial tension at the water interface. Stefano Vanni has parameterized the nonbonded parameters of the pair between TG glycerol moiety and water to increase the interfacial tension. Again, this model barely shows the difference between LD surfaces and bilayers. Because of those opposing quantities, it is doubtful whether a faithful TG model can be constructed using additive force fields. The authors should keep this limitation in mind when reading the following chapters.

The final topic of my thesis is on how LDs emerge from the ER bilayer. LD biogenesis can be conceptually divided into three steps: TG nucleation, LD growth, and LD budding. TG dissolved in the ER bilayer undergoes a phase transition during TG nucleation if the TG concentration is above the critical value, which is known to be 2.4% mol in PC bilayers (37). The nucleated TG forms its own distinct phase between the leaflets of the ER bilayer. If the cells have extra energy (e.g., in experiments, cells are incubated with oleate), TG will be synthesized continuously, resulting in the growth of the TG cluster. This process accompanies membrane remodeling in the ER bilayer, in which two leaflets of the bilayer are separated, and the space in between the two is filled with TG. The last step is spontaneous budding of the enlarged TG phase from the ER bilayer.

LD formation can be understood using the classical nucleation theory (Eq. 1-1). First, TG bulk free energy, proportional to the volume of the TG cluster ( $V$ ), is the driving force of TG nucleation. Second, the surface tension of LDs ( $\gamma$ ) obstructs the LD formation, and the resulting energy term is proportional to the surface area ( $A$ ). LDs have relatively high surface tension ( $\sim 1-4$  mN/m) compared to PC bilayers typically without surface tension. Finally, since LD formation happens in the membrane, the membrane deformation term should be considered. This energy term, which is approximately constant regardless of the LD size, is proportional to the bilayer's surface area and bending modulus and the square of curvature. The interplay between two energy terms, surface tension energy and membrane deformation energy, determines how the LD shape evolves during LD growth. For instance, when a TG cluster is small, the surface tension energy is smaller than the membrane deformation energy because the LD area is small, and the bending modulus of the bilayer (15-35 mN/m) is higher than the LD surface tension ( $\sim 1-5$  mN/m). This makes the shape of the TG cluster flat to minimize membrane deformation. On the other hand, when the LD becomes large, the surface tension term becomes much larger than the membrane deformation one. Therefore, to reduce surface tension energy, the forming LD becomes spherical.

$$\Delta G = -V\Delta g_v + A[\gamma + 0.5 K_C(c_1 + c_2 - c_0)^2 + K_G c_1 c_2]$$

Equation 1-1. TG nucleation according to the classical nucleation theory.  $\Delta g_v$  is the bulk free energy per unit volume, which is the difference in free energy per unit volume between the bulk TG phase and PL phase (mostly hydrocarbon phase).  $A$  and  $\gamma$  are the surface area and the surface tension of the LD.  $K_C$  is the bending modulus,  $K_G$  is the Gaussian bending modulus,  $c_0$  is the spontaneous curvature.  $c_1$  and  $c_2$  describe local curvature, and their inverses are the radii of curvature.



LD biogenesis without protein results in the formation of defective LDs in cells. The ER protein seipin is thought to play a critical role in LD biogenesis as its lack leads to aggregated and small LDs or supersized LDs in cells as well as Bernardinelli-Seip congenital lipodystrophy type 2 (BSCL2). Recent studies suggest that seipin determines LD formation sites and catalyzes TG nucleation (38). Seipin consists of 10-12 subunits, exhibiting a cage-like structure. Each subunit has a highly conserved luminal domain, flanked by two transmembrane segments and non-conserved variable cytosolic tails. Fly and human seipin have a conserved hydrophobic helix in the luminal domain positioned toward the luminal leaflet (39, 40). However, yeast seipin lacks the hydrophobic helix, which may explain why yeast seipin alone is not sufficient to function (41). The recent coarse-grained simulations with the MARTINI or SDK force field demonstrate that the interactions between the hydrophobic helix (in particular, S166 in human seipin) and TG catalyze TG nucleation (42, 43). However, how each part of seipin functions, how seipin facilitates LD growth, and why the oligomeric structure of seipin is important remain elusive. Gregory Voth and I and our collaborators have addressed those questions and extended the molecular understanding of seipin-driven LD formation. The current models of LD biogenesis and remaining scientific questions are reviewed in these references (44-51).

The thesis will be presented in the following order. First, I will show how the CHARMM force field can be converted into the GROMACS format. The combination of the two has been very popular (and used in my research) because GROMACS outperforms other molecular dynamics engines. The CHARMM force field has various lipid types and has been continuously updated by leading scientists. However, converting the CHARMM force field into GROMACS inputs is not trivial, especially if the system contains new molecules not included in the original CHARMM release. The presented python package will show how one converts between the two

with only a few lines of a python script. Second, I will present the all-atom simulations of LDs and show how the LDs differ from the bilayers. Third, I will present the ER-to-LD targeting pathway with model peptides, *LiveDrop* and ALG14. Fourth, I will demonstrate how LD proteins are targeted from the cytosol. The model peptide, CCT autoinhibitory motif, was used in this simulation.

The following work is related to LD biogenesis. First, I will present a new set of TG parameters that reproduces the interfacial tension at the water interface. I will also show how conical neutral lipids, TG and DAG, can modulate the calculated bilayer properties such as bending modulus and TG flip-flop free energy. A new coarse-grained model that is comparable to the MARTINI resolution will also be presented. Second, I will explore the biophysics of LD formation using a new coarse-grained model. The linear 4-bead model for PL and TG is relatively simple. However, it can faithfully demonstrate the interplay between two thermodynamics terms, membrane deformation energy and TG bulk energy, concentration-dependent nucleation, Ostwald ripening, and budding control by tension. Finally, I will introduce a coarse-grained model of seipin and seipin-driven LD biogenesis. This chapter will include our recent findings of how seipin transmembrane functions and converts the shape of an oil lens.

## **Chapter 2 STR2GMX: A Python package for the creation of GROMACS-formatted CHARMM force field**

### **ABSTRACT**

STR2GMX is a Python package that converts CHARMM-formatted topology and parameters into GROMACS-compatible inputs. It takes a single frame coordinate that has no missing atoms and a list of CHARMM topology and parameter files, and outputs files for GROMACS simulations. The conversion process as well as solvation and ionization can be achieved with a few lines of Python code, combined with the MDAnalysis library. The accuracy and reliability of the package were tested by comparing the potential energy of two GROMACS simulations with the inputs created by CHARMM-GUI or STR2GMX of the identical starting configuration. System setup and creation of the GROMACS inputs of lipid droplet (LD) systems are demonstrated, for which a new topology/parameter is required to describe neutral lipids. The program and other examples are available at [git@github.com:ksy141/STR2GMX.git](https://github.com:ksy141/STR2GMX.git).

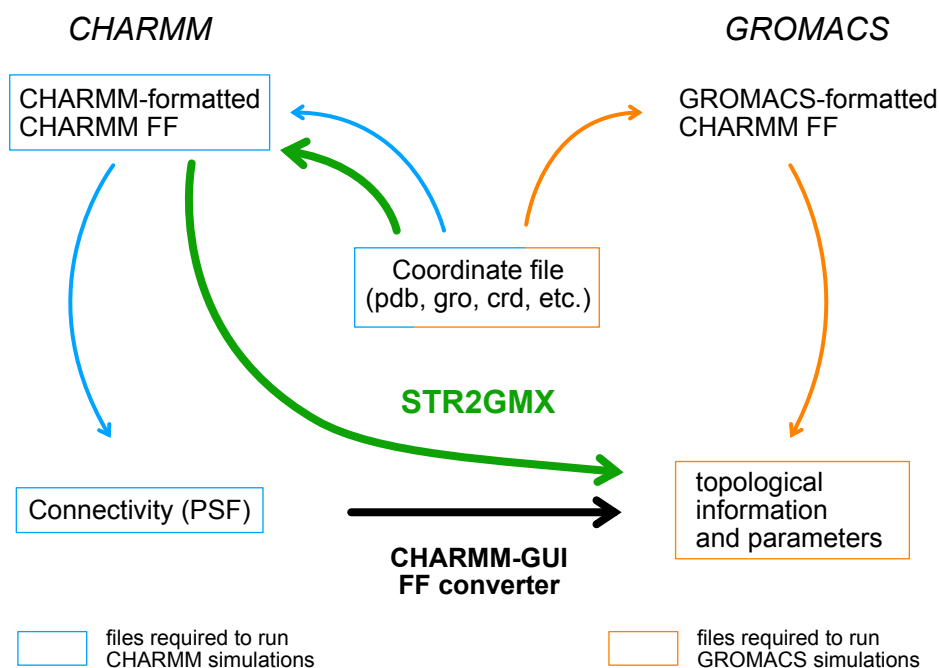
## INTRODUCTION

All-atom (AA) Molecular Dynamics (MD) simulations have proven to be highly valuable for the study of biomolecules, characterizing structural and dynamical properties at an atomic level of detail (52). Various force fields and MD engines to conduct AA-MD simulations have been continuously developed over the last 40 years, each of which having its own strength and limitation. Each MD engine requires force fields to be in its native format. For instance, CHARMM force fields are native to the CHARMM engine; However, they should be re-formatted if simulations are carried out with GROAMCS (53, 54). Such conversion is not straightforward and sometimes challenging.

CHARMM-GUI is a powerful web-based server that provides a full pipeline of the preparation of biomolecular simulations, not only providing good initial structures of complex biomolecular systems but also converting force fields into a format suitable for chosen MD engines (55, 56). It builds and equilibrates a system with CHARMM and then converts force fields at the last step. Proven to be highly valuable for most biomolecular simulations, CHARMM-GUI cannot meet all the needs in part because it is a web-based server, maximizing the convenience of users at the cost of flexibility. For example, users may want to run simulations with the Lennard-Jones cutoff-free CHARMM36 force field (C36/LJ-PME) (57, 58) or the older version of CHARMM force field (e.g. C22) (59) that CHARMM-GUI does not support at this moment. In addition, users may want to perform simulations on a system that contains molecules not included in the CHARMM force fields. Finally, users may want to build a structure that CHARMM-GUI does not support, for instance, a spherical lipoprotein or a lipid droplet (LD) system. System structures can be obtained relatively easy, if not too complicated, thanks to great tools such as PACKMOL (60). However, how to perform MD simulations with GROMACS?

We present a Python package, STR2GMX, to simplify the conversion process of CHARMM-formatted topology and parameters into GROMACS-based inputs (Fig. 2-1). Such a combination, GROMACS simulations with CHARMM force fields, has gained great popularity in biomolecular simulations because GROMACS is one of the fastest MD engines and free, and CHARMM force fields support numerous molecules (54, 61). Two sets of inputs are required for STR2GMX: a single frame coordinate file and a list of CHARMM topology and parameter files. The input coordinate file can be in any format as long as it is a supported format by the MDAnalysis library (62). The structure should be complete, requiring no missing atoms. For each residue or each protein chain, a MDAnalysis's Universe object is created, and the Atom attributes are assigned based on the CHARMM topology and parameter files. Angles, dihedrals, and 1-4 pairs are generated based on the topological information included in CHARMM topology. We note that STR2GMX benefits from MDAnalysis's structural hierarchy and powerful selection tools. Other methods to convert CHARMM force fields into GROMACS inputs will be discussed later.

In this article, we explain the layout of STR2GMX and demonstrate several system setups and input generation for GROMACS simulations, ordered from simple to complex. The reliability and accuracy of STR2GMX will be tested on a system of a protein dimer embedded in a multicomponent bilayer by comparing the potential energy.



**Figure 2-1** Illustration of the workflow to create GROMACS-formatted CHARMM force field.

The orange arrows describe the workflow using a functionality of `gmx pdb2gmx`. The blue arrows describe the workflow that first creates a protein structure file (PSF) and then converts into GROMACS format. STR2GMX (green arrows) takes a complete structure and a set of CHARMM-native topology and parameter files to create GROMACS inputs. Comparisons of three different ways will be made in the Discussion section.

## METHODS

STR2GMX was inspired by `psf2itp.py`, a python script that CHARMM-GUI used to convert a protein structure file (PSF) file into GROMACS inputs (56). Because a PSF file already contains topological information (atom types, partial charges, bonds, angles, dihedrals, impropers, CMAP, etc.), `psf2itp.py` only uses a set of CHARMM parameter files to make GROMACS inputs. However, STR2GMX should read both CHARMM topology and parameter files for GROMACS inputs. STR2GMX first builds the topological information of each residue species or each protein chain using the CHARMM topology. During this step, angles, dihedrals, and 1-4 pairs are created based on the bond information. STR2GMX then reads CHARMM parameters to build GROMACS inputs.

A care should be taken for an input coordinate file. STR2GMX does not have a functionality to guess coordinates of missing atoms. Therefore, a coordinate file should be complete, otherwise it will issue warnings. For the case of proteins, one can use CHARMM-GUI's `pdbreader` or `solution builder` to guess coordinates of missing atoms and patch N- and C- termini. If a protein chain contains disulfide bonds, they should be specified during this step. Otherwise, corresponding hydrogen atoms (atom name of HG1) of cysteine should be removed manually in a coordinate file. STR2GMX currently supports NTER, GLYP, PROP, NNEU, NGNE, ACE, and ACP for N-terminus and CTER, CNEU, CT1, CT2, and CT3 for C-terminus.

The first step of using STR2GMX is to create a `ReadToppars` object to save the CHARMM topology and parameters. If users provide a path to a CHARMM force field folder, a `ReadToppars` object will read all the files included in the folder. However, this way, the files will not be read in order as users hope, which could be an issue if there are duplicate entities. A recommended way is to make a file that specifies a list of files, which will then be read in order. A default file,

`toppar.str`, is included in the package as well as the C36 and C36/LJ-PME force fields (61, 63). The C36 force field was taken from the CHARMM-GUI output, which contains more molecules than the original release. A `ReadToppars` object can be created as follows:

```
from os.path import expanduser
from STR2GMX import *
toppar = expanduser('~'/STR2GMX/FF/C36/toppar.str')
```

A `Molecule` object for each non-protein molecular species takes a `MDAnalysis's Atomic Group` as an argument. An `Atomic Group` should only contain atoms of the species. For instance, when making a `Molecule` object for 3-palmitoyl-2-oleoyl-D-glycero-1-phosphatidylcholine (POPC), the following `Atomic Group` should be created and passed as an argument where `u` is the `Universe` object:

```
ag = u.select_atoms('resname POPC')
```

Using a passed `Atomic Group`, a `Molecule` object saves the number of residues of the same type and a residue name as attributes. It then internally creates a new `MDAnalysis's Universe` object that contains atoms of a singular molecule (the first residue). Atomic data, bonds, and improper dihedrals of this species are read from the `ReadToppars` object and then assigned to the `Atom` or `Universe` object. By default, angles and dihedrals are generated based on the bond information using `MDAnalysis's guess` functions. However, this should be turned off for small molecules such as ions and water. A `Molecule` object for POPC can be created as follows:

```
POPC = Molecule(ag, toppar)
```

A `Molecule` object should be created for each non-protein molecular species; However, this could be tedious if a system contains various species. A `MultipleMolecules` object can reduce this burden



by automatically creating a Molecule object for each residue type. In this case, an Atomic Group should contain all the atoms of the species of the interest. For most cases, two non-protein objects, one that requires the generation of angles and dihedrals and another that should not generate angles and dihedrals automatically, will be sufficient. Consider a bilayer membrane that contains POPC, 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), and phosphatidylinositol (SAPI), solvated with 0.15M KCl solution. Molecule objects should be created for POPC, DOPE, SAPI, POT, CLA, and TIP3 residues; However, this can be simplified with the following lines:

```
mols = []
lipids = u.select_atoms('resname POPC DOPE SAPI')
mols += MultipleMolecules().generate(lipids, toppar)
solv = u.select_atoms('resname POT CLA')
mols += MultipleMolecules().generate(solv, toppar, generate_angles = False,
generate_dihedrals = False)
```

A chain object is created for a protein chain that has one N-terminus and one C-terminus. It is an identical class with Molecule, however, with the addition of looping over residues to make sure the connectivity between neighboring residues and the proper CMAP. Disulfide bonds should be passed as arguments if there are. For a system that contains more than one protein chain, segname for each protein chain should be given, otherwise, a resulting topology file will be overwritten to a default segname, PROA.

Finally, a Topology object is a container of all the Molecule or Chain objects. By iterating each Molecule or Chain object, a Topology object gathers required topological information from the ReadToppars object to build GROMACS-compatible inputs.

In addition to the creation of a GROMACS-formatted CHARMM force field, STR2GMX provides a convenient solvation tool. When a Universe object is passed as an argument, a Solvation

object fills up the whole space with water. Water molecules within a cutoff distance from the atoms contained in the Universe object will be removed. The space that should be filled up with water is controlled by the dimensions of a simulation box. In the case of a membrane, one can set a forbidden zone where there should be no water molecules using Z coordinates, which will be passed as arguments. Finally, to balance the net charge, a solvation tool provides the conversion of water molecules into ions. Types of ions and concentration can be given. In the case of the purpose of adding neutralizing ions, concentration should be provided as 0. An Atomic Group that contains water molecules is passed as an argument and randomly chosen water molecules in the Atomic Group will be converted to chosen ion types. The shortest distance between an ion pair will be printed so that users can appreciate the quality of this conversion process.

## RESULTS

In this section, we demonstrate various examples of system setups and input generation for GROMACS simulations, ordered from simple to complex. We first consider a protein chain in a solution, consisting of three alanine residues with the N-terminus patched with the standard N-terminus and the C-terminus patched with the neutral C-terminus (Fig. 2-2). A trialanine chain was created with the PyMOL's build function and the patch was applied using the CHARMM-GUI's pdbreader. The protein chain has a net charge of +1 as the N-terminus is positively charged and the C-terminus is neutral. A trialanine was first solvated with water and then KCl and CaCl<sub>2</sub> ions were sequentially added. The final net charge of the system became 0.

```

import MDAAnalysis as mda
import numpy as np
from STR2GMX import *
from os.path import expanduser

### TRI-ALA PEPTIDE
u = mda.Universe('step1_pdbreader.crd')
u.atoms.positions -= u.atoms.center_of_mass()

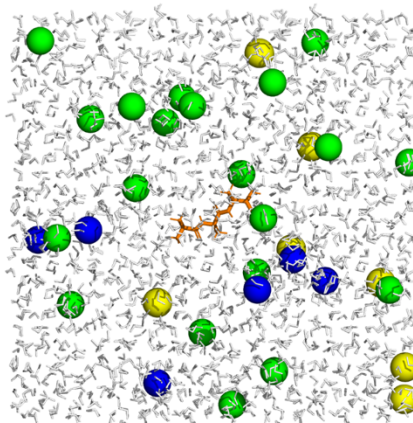
pbc = 50
u.atoms.positions += np.array([pbc, pbc, pbc]) / 2
u.dimensions = [pbc, pbc, pbc, 90, 90, 90]

### SOLVATE WITH 0.1M KCl AND 0.1M CaCl2
s = Solvate()
u2 = s.sol(u, cutoff=5)
ag = u2.select_atoms('resname TIP3')
u3 = s.ion(ag, qtot=1, conc=0.1, pos='POT', neg='CLA')
ag = u3.select_atoms('resname TIP3')
u4 = s.ion(ag, qtot=0, conc=0.1, pos='CAL', neg='CLA')
u4.atoms.write('step2.gro')

### MAKE GROMACS INPUTS
u4 = mda.Universe('step2.gro')
toppar = ReadToppars(toppars=expanduser('~/.STR2GMX/FF/C36/toppar.str'))
chain = Chain(u4.select_atoms('protein'), toppar)
SOL = MultipleMolecules().generate(u4.select_atoms('resname POT CAL CLA TIP3'), \
    toppar, generate_angles = False, generate_dihedrals = False)

top = Topology([chain, *SOL], toppar)
top.write()

```



**Figure 2-2** Solvation of trialanine.

Trialanine was patched with the standard N-terminus (NTER) and the neutral C-terminus (CNEU). A Python script sequentially adds water, 0.1 M KCl, and 0.1 M CaCl<sub>2</sub>. Before adding 0.1 M KCl, the net charge of the system was +1 because of the positively charged N-terminus. STR2GMX determines the number of positive and negative ions to neutralize `qtot` of the system, therefore, after adding 0.1 M KCl, the system did not have a net charge. CaCl<sub>2</sub> ions were added afterward, followed by the creation of GROMACS inputs. In the inlet snapshot, trialanine, water, potassium, calcium, and chloride are indicated with orange, white, yellow, blue, and green, respectively.

In the next example, we demonstrate a lipid droplet (LD) system (Fig. 2-3). LDs are universal organelles that store triolein (TG) in the core, surrounded by a PL monolayer and coating proteins (47). Two peptides are included in this example (64, 65), the auto-inhibitory (AI) helix of CTP:phosphocholine cytidyltransferase (CCT) and the hydrophobic hairpin motif of Glycerol-3-phosphate acyltransferase 4 (GPAT4), referred to as *LiveDrop*. LD targeting of those peptides, using two distinct pathways, has been studied previously (66, 67). We first took an equilibrated LD structure from the previous work and removed water and ions (68). Due to their cellular

positions, we placed the AI helix and *LiveDrop* at the water level and PL level, respectively. Lipids within 0.8 Å of *LiveDrop* were removed. The system then became solvated with 0.15 M NaCl solution, followed by the creation of GROMACS inputs. Because TG is not included in the original C36 release, we included a TG topology in the force field folder and indicated this in the `toppar.str` file, which is available at [git@github.com:ksy141/SMDAnalysis.git](https://github.com/ksy141/SMDAnalysis). In the supporting information, we also demonstrate a workflow to create a spherical LD with modified TG parameters (69).

```

import MDAAnalysis as mda
import numpy as np
from STR2GMX import *
from os.path import expanduser

### MEMBRANE
LD = mda.Universe('LD.gro')
pbc = LD.dimensions[0:3]

### PROA [CCTa]
CCTa = mda.Universe('CCT.gro')
CCTa.atoms.positions -= CCTa.atoms.center_of_mass()

### PROB [LiveDrop]
Live = mda.Universe('LiveDrop.gro')
Live.atoms.positions -= Live.atoms.center_of_mass()

### PHOSPHORUS LEVEL OF THE UPPER AND LOWER LEAFLET
UP = LD.select_atoms('name P and prop z > %f' %(pbc[2] / 2))
LP = LD.select_atoms('name P and prop z < %f' %(pbc[2] / 2))
zUP = UP.center_of_mass()[2]
zLP = LP.center_of_mass()[2]

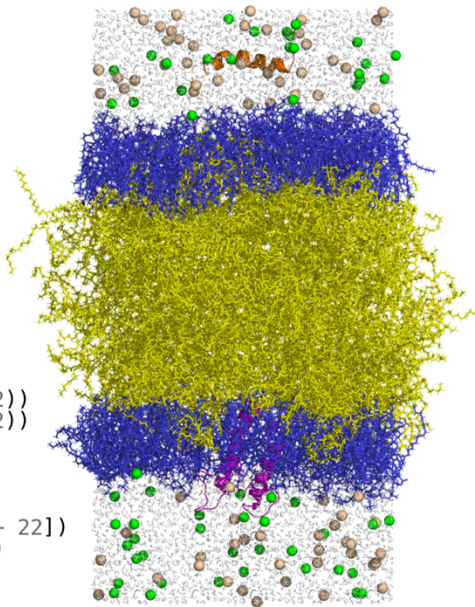
### PLACE PROTEINS
CCTa.atoms.positions += np.array([pbc[0]/2, pbc[1]/2, zUP + 22])
Live.atoms.positions += np.array([pbc[0]/2, pbc[1]/2, zLP])

### MERGE INTO ONE STRUCTURE
u = mda.Merge(CCTa.atoms, Live.atoms, \
              LD.select_atoms('not resname TIP3 SOD CLA'))
u.dimensions = LD.dimensions
ag = u.select_atoms('not byres (around 0.8 protein)')
ag.write('combined.gro')

### SOLVATE (BECAUSE OF SAPI, QTOT IS -)
s = Solvate()
u1 = mda.Universe('combined.gro')
u2 = s.sol(u1, zUP=zUP, zDW=zLP, cutoff=3) # Waters will be excluded from zLP to zUP
ag = u2.select_atoms('resname TIP3')
u3 = s.ion(ag, qtot=-15, conc=0.15, pos='SOD', neg='CLA')
u3.atoms.write('final.gro')

### MAKE GROMACS INPUTS
u = mda.Universe('final.gro')
topparpath = expanduser('~/.SMDAnalysis/FF/CHARMM/toppar.str')
toppar = ReadToppars(toppars=topparpath)
chainA = Chain(u.select_atoms('index 0-361'), toppar, segname='PROA')
chainB = Chain(u.select_atoms('index 362-1645'), toppar, segname='PROB')
LIP = MultipleMolecules().generate(u.select_atoms('resname POPC DOPE SAPI TRIO'), toppar)
SOL = MultipleMolecules().generate(u.select_atoms('resname SOD CLA TIP3'), toppar, \
                                   generate_angles=False, generate_dihedrals=False)
top = Topology([chainA, chainB, *LIP, *SOL], toppar)
top.write()

```

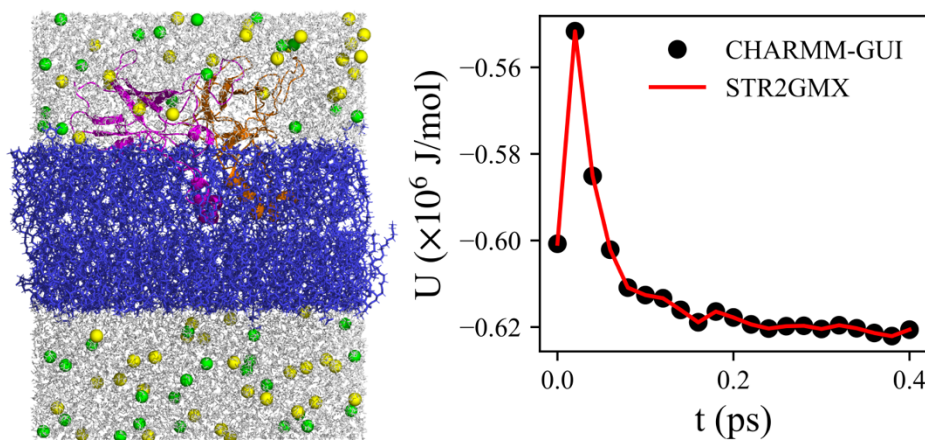


**Figure 2-3** System setup and the creation of GROMACS inputs of a LD system, containing *LiveDrop* and CCT.

An equilibrated LD structure without solution was taken as an initial structure. In the inlet, blue and yellow represent PL and TG molecules, respectively. The AI helix of CCT and *LiveDrop* are indicated with orange and pink. Sodium, chloride, and water in wheat, green, and white, respectively.

Finally, we verify the accuracy and reliability of the package by comparing the potential energy of two short GROMACS simulations of a complex system, one simulated with the

CHARMM-GUI output and another simulated with the STR2GMX output. Those two simulations had the identical initial configuration and simulation parameters. The system contains a human seipin dimer, mediating LD biogenesis, and various lipid types (38, 40). Seipin is known to form an oligomer, however for the sake of simplicity, a seipin dimer was placed in a membrane. Each seipin chain contains a disulfide bond. An initial structure was prepared with the CHARMM-GUI membrane builder. The identical potential energy between two simulations during a short period proves the accuracy and reliability of the package (Fig. 2-4). Python code to create GROMACS inputs using STR2GMX is included in the Supporting Information. Many other examples have been tested and all those examples show the perfect agreement in the potential energy between simulations prepared with CHARMM-GUI and STR2GMX.



**Figure 2-4** Verification of STR2GMX.

(left) Seipin dimer in a multi-component bilayer membrane. Blue represents lipids of 14 types and orange and pink represent each protein chain. Water, potassium, and chloride are in white, yellow, and green, respectively. (right) A potential plot for the simulation that was run with GROMACS inputs included in the CHARMM-GUI output (black dots) or with GROMACS inputs created with STR2GMX (red line). A Python script for the input generation using STR2GMX is included in the Supporting Information.

## DISCUSSION

To run GROMACS simulations with CHARMM force fields, those should be converted to GROMACS-based inputs. Such conversion can be achieved with three ways. First, the user can use GROMACS's built-in function, `gmx pdb2gmx` (Fig. 2-1 orange arrows). Topological information and parameters which are already formatted for GROMACS are read from the GROMACS folder. While this is the most efficient method to build GROMACS inputs, it bears several weaknesses. First, topological information and parameters included in the GROMACS folder are not up to date. The latest CHARMM force field included in the GROMACS release is C27 although GROMACS-formatted C36 can be downloaded from the MacKerell lab website. If a new CHARMM force field is released, the user should wait until the force field is converted into GROMACS format and uploaded. For instance, at this moment, GROMACS-formatted C36/LJ-PME is not available. Second, if the user wants to include a new molecule type in the CHARMM force field, it is more straightforward to work with the CHARMM format. For instance, the CHARMM General Force Field provides topology and parameter files of small organic molecules in the CHARMM format (70).

The second way is to make a PSF file using CHARMM (53) or NAMD (71) and then converts this to GROMACS inputs (Fig. 2-1 blue arrows) using TopoGromacs (72) or `psf2itpy.py` (56). While this approach has an advantage of enabling the user to run simulations with various MD engines, building a PSF file itself could be disadvantageous. Since a PSF file includes topological information of all molecules, the time for making a PSF file of a large system could be demanding. For instance, the time spent to create a PSF file of a POPC bilayer membrane with CHARMM, containing 1.8 M atoms, was more than an hour on a laptop with i7-10510U (69). In addition, to make a PSF file, a structure should be split into each segment,

creating intermediate files. NAMD/VMD's `psfgen` performs much faster in making a PSF file, however, a disadvantage is that it only reads PDB files. The maximum number of residues is limited to 10K and the residues beyond those are considered duplicated entities in PDB files. Therefore, in a large biomolecular system, water molecules should be saved into different PDB files for every 10K residues. In the above bilayer example, there were more than 420K water molecules and one should split up those residues into 43 different PDB files, each of which includes 10K water molecules.

An easy-to-use and user-friendly Python package, STR2GMX, converts CHARMM force fields into GROMACS-compatible inputs. While its output can be only used in GROMACS, users can enjoy the simplicity and efficiency of STR2GMX. As demonstrated before, a few lines of Python code are sufficient to build a system and make GROMACS inputs. Most users are also more familiar with a Python language (STR2GMX) than a Fortran (CHARMM) or a Tcl (NAMD/VMD) language. Also, bypassing the creation of a PSF file can reduce the time spent on a system preparation and avoid the creation of intermediate files. In the above 1.8 M-atoms example, STR2GMX creates the GROMACS inputs within 2 seconds on the same machine. Plus, STR2GMX uses MDAnalysis to read coordinates, which supports numerous file formats (62). Therefore, the user can work with the coordinate file that they are familiar with.

## CONCLUSIONS

We present STR2GMX, a Python package that converts CHARMM-based to GROMACS-based inputs. Various examples of the system setup and GROMACS input generation are demonstrated. This package will be particularly useful when a simulation should be conducted with a new CHARMM force field or a system contains a new molecule that is not included in the original

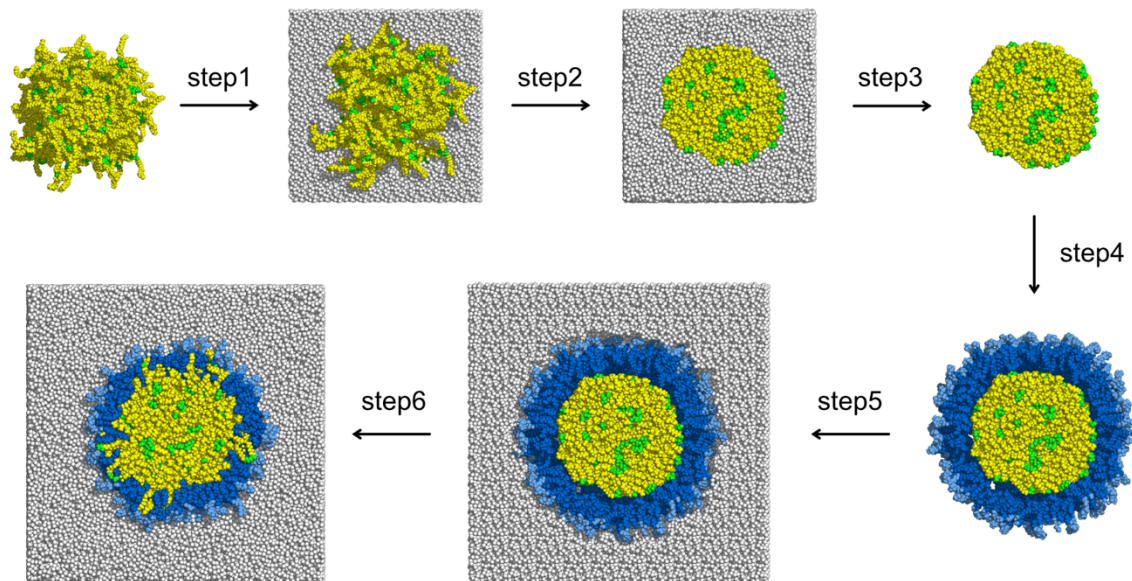


CHARMM force field release. Obtaining GROMACS-compatible inputs from a complete structure and native CHARMM force fields could be easily achieved with a few lines of Python code using this package. STR2GMX also conveniently provides solvation and ionizing tool for easier system setup.

## **ACKNOWLEDGEMENTS**

Jumin Lee and Wonpil Im wrote `psf2itp.py` that converts a PSF file into GROMACS inputs. STR2GMX uses `psf2itp.py` in reading CHARMM parameters and writing the CHARMM-GUI-style outputs. This research was supported by a grant from the National Institute of General Medical Science (NIGMS) of the National Institutes of Health (NIH), grant R01-GM063796. S.K. thanks He Ma for the valuable discussions.

## SUPPORTING INFORMATION



**Figure 2-5** Workflow to create a lipoprotein using STR2GMX

*Snapshots were clipped with the XZ plane.*

step1: Add water with STR2GMX.

step2: Build GROMACS inputs with STR2GMX and run a simulation with GROMACS.

step3: Take a bulk TG using MDAnalysis.

step4: Place POPC molecules with PACKMOL.

step5: Add water with STR2GMX.

step6: Build GROMACS inputs with STR2GMX and run a simulation with GROMACS.

## Python script for the creation of GROMACS inputs of a seipin dimer (related to Fig. 2-4)

```
import MDAnalysis as mda
from STR2GMX import *

mols = []
toppar = ReadToppars(toppars='/Users/siyoungkim/STR2GMX/FF/C36/toppar.str')

u = mda.Universe('../step6.1_equilibration.gro')
PROA = u.select_atoms('index 0-2553')
PROB = u.select_atoms('index 2554-5107')

SG1 = u.select_atoms('resname CYS and name SG and index 242')
SG2 = u.select_atoms('resname CYS and name SG and index 344')

SG3 = u.select_atoms('resname CYS and name SG and index 2795')
SG4 = u.select_atoms('resname CYS and name SG and index 2897')

chainA = Chain(PROA, toppar, segname='PROA', add_bonds = [SG1, SG2])
chainB = Chain(PROB, toppar, segname='PROB', add_bonds = [SG3, SG4])
mols.extend([chainA, chainB])

lipids = u.select_atoms('not protein and not resname TIP3 POT CLA')
lipids_mols = MultipleMolecules().generate(lipids, toppar,
generate_angles=True, generate_dihedrals=True)
mols.extend(lipids_mols)

solv = u.select_atoms('resname POT CLA TIP3')
solv_mols = MultipleMolecules().generate(solv, toppar, generate_angles=False,
generate_dihedrals=False)
mols.extend(solv_mols)

top = Topology(mols, toppar, prefix='toppar')
top.write()
```

### **Chapter 3 The Surface and Hydration Properties of Lipid Droplets**

This chapter is adapted from [Kim et al., 2020] (68).

#### **ABSTRACT**

Lipid droplets (LDs) are energy storage organelles composed of neutral lipids, such as triacylglycerol (TG) and sterol esters, surrounded by a phospholipid (PL) monolayer. Their central role in metabolism, complex life cycle, and unique lipid monolayer surface have garnered great attention over the last decade. In this paper, results from the largest and longest all-atom simulations to date suggest that 5-8% of the LD surface is occupied by TG molecules, a number that exceeds the maximum solubility reported for TGs in PL bilayers (2.8%). Two distinct classes of TG molecules that interact with the LD monolayer are found. Those at the monolayer surface (SURF-TG) are ordered like PLs with the glycerol moiety exposed to water, creating a significant amount of chemically-unique packing defects, and the acyl chains extended toward the LD center. In contrast, the TGs that intercalate just into the PL tail region (CORE-TG) are disordered and increase the amount of PL packing defects and the PL tail order. The degree of interdigitation caused by CORE-TG is stable and determines the width of the TG-PL overlap, while that caused by SURF-TG fluctuates and is highly correlated with the area per phospholipid or the expansion of the monolayer. Thus, when the supply of PLs is limited, SURF-TG may reduce surface tension by behaving as a secondary membrane component. The hydration properties of the simulated LD systems demonstrate approximately 10 times more water in the LD core than previously reported. Collectively, the reported surface and hydration properties are expected to play a direct role in the mechanisms by which proteins target and interact with LDs.

## STATEMENT OF SIGNIFICANCE

Long (10  $\mu$ s) all-atom molecular dynamics simulations herein suggest that lipid droplets (LDs) have several unique and previously unknown physical properties. First, after a long equilibration process ( $>5$   $\mu$ s), many triacylglycerol molecules (neutral fats in the LD core) transition to the LD surface where they adopt phospholipid-like structure and dynamics, thereby creating chemically distinct defect types from bilayer membranes. Second, additional triacylglycerols intercalate into the phospholipid tails but retain core-like disorder. Finally, the amount of water in the LD core region shows a slight preference for the phospholipid-core interface and is significantly larger than previously reported. Each of these properties is expected to influence the nature by which proteins target and interact with LDs.

## INTRODUCTION

The phospholipid (PL) monolayer surrounding lipid droplets (LDs) is a distinguishing feature from other organelles, which are generally surrounded by phospholipid bilayers. LDs play a crucial role in cellular metabolic processes as they store excess energy in the form of neutral lipids such as triacylglycerols (TGs) and sterol esters (21, 47, 73). The proteins that associate with LD surfaces regulate lipogenesis and lipolysis to store and use energy when necessary, modulating the size of the LD depending on the metabolic status of the cell. Therefore, investigating the mechanisms of how proteins target and interact with LD surfaces is central to understanding the biology of LDs and related metabolic diseases (28, 74, 75).

LD proteins are classified into two types based on their origin (28). The first type (class I proteins) involves proteins embedded in the ER bilayer membrane that relocate to LD via LD-ER membrane contact points (65, 76, 77). This mechanism of LD protein targeting requires

membrane continuity and occurs either when LDs bud from the ER membrane or when LDs rebind to the ER via ER-LD bridges, in which the cytosolic leaflet of the ER membrane is continuous with the monolayer surface of LDs (65, 77). Type I proteins typically contain a hydrophobic, membrane-embedded hairpin structure that has been shown to be sufficient for LD localization in a sequence-sensitive manner (28, 65, 78). Examples of type I proteins include enzymes involved in TG synthesis, such as GPAT4 (22, 65) and DGAT2 (79-81). The second type (class II proteins) targets LDs from the cytosol and is generally characterized by one or more amphipathic helices (28, 82). For example, the perilipin family binds to the surface of LDs via their 11mer-repeat amphipathic region and regulate lipolysis (83-85). In S2 cells, CTP:phosphocholine cytidyltransferase (CCT) associates with LDs via its amphipathic M domain and catalyzes the synthesis of phosphatidylcholine in order to augment the surface monolayer of growing LDs (75, 82, 86).

Many LD proteins can bind to both bilayer membranes and monolayer surfaces but have a preference for one over the other. Given that LDs have the same PL composition as the ER bilayer, it is likely that the physical properties of monolayers versus bilayers influence preferential LD targeting. For example, the hydrophobic residues in class I proteins, initially embedded in the ER membrane, were recently shown to gain thermodynamic stabilization due to interactions with the LD monolayer, and particularly with intercalated TG molecules (78). Similarly, it has been suggested that LDs have bigger, more persistent packing defects than bilayer membranes (28, 36, 82), and that class II proteins may preferentially associate with these defects via interactions with large hydrophobic residues in their amphipathic helices (82). However, this general packing defect model is only valid when describing nonspecific adsorption of amphipathic helices; it cannot explain LDs' differential recruitment. For instance, each perilipin family member targets different

LDs depending on their distinct neutral lipid composition (87). Also, the binding affinity of an amphipathic helix was recently shown to depend on the neutral lipid composition of the LD, despite having the same phospholipid composition and density (34). Yet, how the neutral lipids, thought to be below the phospholipid monolayer, are involved in protein recruitment is largely unknown. For these reasons, characterizing the physical properties of LDs is not only an intriguing question in LD biology, but is central to understanding the mechanisms of proteins targeting to LDs.

Despite their biological importance and partially due to their more recent discovery, the physical properties of LD systems have not yet been studied as much as those of bilayer membranes. The solubility of neutral lipids in bilayers has been measured experimentally (37, 88), but has yet to be measured in LD monolayers. It is challenging to characterize lipids at the molecular scale using experimental techniques. Computational methods can complement experimental findings by providing molecular-level insight. For example, coarse-grained (CG) simulations of bilayer membranes with the addition of the small amount of neutral lipids have suggested how the spontaneous accumulation of neutral lipids between two PL monolayers could occur during the early stages of the LD biogenesis (89-91). All-atom simulations have helped describe low-density lipoproteins (LDLs) and high-density lipoproteins (HDLs), which have similar structures to LDs once they have matured to their spherical shape despite being much smaller and sterol-ester abundant (92-99). Lipid droplets, in contrast, are TG-abundant and larger (100 nm to 100  $\mu$ m in diameter) (47), resulting in a surface that is close to planar as opposed to the highly curved LDL/HDL surface. Finally, a few simulation studies have focused directly on the surface properties (36, 82), physico-chemical properties (100) and the inner lipid distribution (101, 102) of LDs. However, many of these previous studies used CG models, which may not describe

hydration, dynamical properties or entropy-driven processes correctly (103-106). Those that were all-atom simulations (36, 82, 99), involved simulation durations ( $< 1 \mu\text{s}$ ) that may not be sufficient to reach equilibration for LD systems.

In this work, all-atom molecular dynamics (MD) simulations of the largest reported LD systems (8 nm and 16 nm thick TG layers) and longest reported simulation time (10  $\mu\text{s}$ ) are analyzed. Three novel facets are found: 1) 5-8% of the LD surface is occupied by TG molecules, which is substantially larger than previous suggestions for monolayers and exceeds the experimentally reported solubility of TG in a pure POPC PL bilayer (2.8 %) (37); 2) There are two types of PL-intercalated TG molecules: those closer to the core (CORE-TG) and those on the LD surface (SURF-TG), which have the same structure and order as PLs, and thereby act as a secondary membrane component creating a chemically unique defects in LD surfaces; 3) The LD core contains  $0.01 \text{ g/cm}^3$  of water, which is approximately ten-fold higher than previously reported (36). Implications for how the reported properties could influence protein-LD interactions and targeting are discussed.

## **METHODS**

### *System setup and simulation details*

Five systems were studied: three bilayer systems with various PL compositions of 3-palmitoyl-2-oleoyl-D-glycero-1-Phosphatidylcholine (POPC), 2,3-dioleoyl-D-glycero-1-phosphatidylethanolamine (DOPE), and phosphatidylinositol (SAPI) and two LD systems with 8 nm and 16 nm thick TG layers. The bilayer systems were prepared using the CHARMM-GUI membrane builder (56, 107-109), and included a homogeneous POPC bilayer (POPCbil), a



homogeneous DOPE bilayer (DOPEbil), and a heterogeneous bilayer of 88:37:10 POPC:DOPE:SAPI, representative of the ER membrane (ERbil) (7). We note that DOPE is a non-lamellar lipid that forms the inverse hexagonal phase at room temperature (110). The forcefield used for DOPE properly reproduces this behavior, but for finite simulation times starting in the lamellar phase and under periodic boundary conditions remains in a lamellar phase (111, 112). The mechanical properties of two different DOPE phases were described by Sodt and Pastor (113). The lamellar DOPE bilayer simulations were included herein only to demonstrate the influence of both phospholipid composition and environment (bilayer versus monolayer) on the degree of order of the lipid tails. We further benchmarked our POPCbil simulation against reference data. Our computed area per lipid headgroup (APL) of POPCbil ( $64.5 \pm 0.4 \text{ \AA}^2$ ) is equal within statistical uncertainty to the reported value ( $64.7 \pm 0.2 \text{ \AA}^2$ ) (112). We also confirmed that the order parameters of POPC are in good agreement with reference data (112).

Given that LDs are thought to have a comparable composition to the ER, our LD monolayers were also composed of 88:37:10 POPC:DOPE:SAPI (7). Similar to previous studies (92, 99), the LDs were modeled with trilayer structures that were constructed by separating the two leaflets of a bilayer membrane (ERbil) and then inserting a relaxed TG layer (8nm or 16nm thick) between them. We denote the resulting LD systems as 8nm LD or 16nm LD, reflective of their TG layer thickness. The experimentally measured surface tension of LDs containing TG was determined to be 1.63 mN/m (34). This low surface tension justifies approximating a LD as a trilayer structure simulated in the zero-surface tension NPT ensemble, although the influence of surface tension will be the focus of future work. The detailed procedure of building a trilayer structure is as follows. The initial bilayer membrane was constructed with the CHARMM-GUI membrane builder (56, 107-109). Independently, a 4 nm thick (in the Z dimension) TG layer was

prepared using Packmol (114) and was equilibrated for 50 ns in the NPT ensemble resulting in a final density of 0.9040 g/cm<sup>3</sup>. The final structure of the TG layer and the bilayer membrane had the same X and Y dimensions of 9.5 nm. The 4 nm thick TG layer was duplicated in the Z dimension to build the 8 nm and 16 nm thick TG layer. The two leaflets of ERbil were separated and the TG layer was inserted with extra 1 nm spacing along the Z dimension between the TG layer and each of PL leaflets. Any TG molecules that were within 2 Å of PLs were removed. The extra space was reduced with 0.1 ns of NPT simulation. All systems were solvated in TIP3P water (115) and 0.15M NaCl solution. A detailed description of the systems is provided in the Table 3-1.

**Table 3-1** A detailed description of the systems used in this work

	Bilayer membranes			Lipid droplets	
	POPCbil	DOPEbil	ERbil	8nm LD	16nm LD
POPC:DOPE:SAPI <sup>a)</sup>	40:0:0	0:40:0	88:37:10	88:37:10	88:37:10
TG <sup>b)</sup>	0	0	0	429	866
Simulation length (μs)	0.2	0.2	1	10	10

a) The number of molecules per leaflet, b) The number of TG molecules

The TG topology was obtained from DOPE by replacing its head group with its sn-1 tail (The structures of POPC, DOPE, and TG are shown in Fig. 3-7). The TG topology used in this study is available in <https://github.com/ksy141/TG>. The same TG topology has been used in the other papers (78, 82). In order to validate our TG model, we performed a bulk TG simulation (35 ns) and a water-TG-water interfacial simulation (50 ns). Our bulk TG simulation show the density to be  $0.9034 \pm 0.0005$  g/cm<sup>3</sup> at 310 K, which is very close to experimental data (0.8991 g/cm<sup>3</sup> at 313K) (116). In the interfacial simulation, the calculated interfacial tension,  $29.7 \pm 1.7$  mN/m at 310 K, is in good agreement with the experimentally measured, 32 mN/m (33) at 298K.

The MD simulations were performed using the GROMACS (version 2018) simulation engine (117) with the CHARMM36 lipid force field (112, 118). Simulations were evolved with a 2-fs timestep. The Particle Mesh Ewald algorithm (119) was used to evaluate long-range electrostatic interactions with a real space cutoff of 1.0 nm. Lennard-Jones interactions were cut-off at 1.0 nm with the potential-shift function. Long-range dispersion was corrected for energy and pressure. The pressure was maintained semi-isotropically using the Parrinello-Rahman barostat (120) at a pressure of 1.0 bar, a compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$ , and a coupling time constant of 2.0 ps. Bonds to hydrogen atoms were constrained using the LINCS algorithm (121). The temperature was maintained at 310 K using the stochastic velocity rescaling thermostat (122) with a coupling time constant of 0.1 ps. The trajectories of POPCbil and DOPEbil were extracted every 100 ps while those for the ERbil, 8nm LD, and 16nm LD were extracted every 1 ns. The total durations of simulations of POPCbil, DOPEbil, ERbil, 8nm LD, and 16nm LD were 0.2  $\mu\text{s}$ , 0.2  $\mu\text{s}$ , 1  $\mu\text{s}$ , 10  $\mu\text{s}$ , and 10  $\mu\text{s}$ , respectively. When computing the time-averaged quantities, the first 100 ns and 5  $\mu\text{s}$  were considered to be an equilibration process and discarded for the bilayer and LD simulations, respectively. For time-series results, the running average of 100 data points were shown instead of instantaneous values. All error bars were estimated by dividing the equilibrated trajectory into five blocks, computing the quantity for each block, and then determining the standard deviation of the block quantities (i.e., block averaging). Molecular images included in this work were rendered using Visual Molecular Dynamics (VMD) (123) and the trajectories were analyzed with MDAnalysis (124)

*Classification of TG*

TG molecules were categorized into two groups: SURF-TG and CORE-TG for those on the surface and in the core of the LD, respectively. The designation was based on the distance between the TG oxygen atoms and the average position of PL tails (along the Z axis) of the closer leaflet. If all six of a given TG's oxygen atoms were above the average Z position of the PL tails ( $z_{avg}$ ) of the upper leaflet or below the average Z position of the PL tails of the lower leaflet, it was classified as SURF-TG. Otherwise, it was considered CORE-TG. Although there were some occurrences of only 1-5 TG oxygen atoms above/below the average Z position of the PL tails, these were transient species that either returned to the core region or stabilized as SURF-TG.

### *Lipid-packing defects*

For the bilayer membranes, we used a Cartesian-based algorithm to evaluate lipid-packing defects (36, 125-127). For each PL located in a leaflet, the atoms whose Z positions were greater than a certain threshold value ( $z_{thr}$ ) were chosen (we use greater here in reference to the upper leaflet, but the opposite applies to the lower leaflet throughout). In this work, the Z position of a PL's central C2 atom minus a chosen value,  $d_{PL} = 1 \text{ \AA}$ , was used as the threshold ( $z_{thr} = z_{C2} - d_{PL}$ ) to be consistent with the previous work (36, 125-127). Then, we created a three-dimensional grid with a spacing of  $1 \text{ \AA}$  on the surface of the membrane with Z coordinates ranging from  $z_{thr}$  to the highest Z position of all PL atoms ( $z_{max}$ ). The scalar value of the grid point is associated with the surrounding atoms that are closer than the half of the diagonal of the grid ( $\sqrt{3}/2 \text{ \AA}$ ) plus the atom's radius. The radii were taken from the CHARMM36 parameter set (112). Different values were added to the grid point based on the types of overlapping atoms. Polar atoms (head groups or glycerol moieties) added a value of  $10^6$ , while acyl chain atoms only add  $10^3$ . After looping over

all the PL molecules, the three-dimensional grid is reduced to a two-dimensional grid by summing up the scalar values along the Z axis.

In this two-dimensional grid, a point with the scalar value of 0 has no atoms near it and is assigned to an elementary ‘deep’ defect. On the other hand, a grid point with the value equal to or greater than  $10^6$  is assigned to no defect as this grid point overlaps with at least one of the head group or glycerol atoms. A grid point with the value ranging from  $10^3$  to  $10^6$  (exclusive) is defined as an elementary ‘shallow’ defect, which will be referred to in this work as an elementary ‘PL acyl’ defect. For each defect type, neighboring elementary defects were clustered. If a clustered defect contains  $N$  elementary defects, this cluster is considered to have a defect size of  $N \text{ \AA}^2$ . We computed the packing defect distribution with 400 bins, ranging from 0 to  $150 \text{ \AA}^2$ . Then, the probability of finding a defect with the size of  $N \text{ \AA}^2$  was fit to an exponential decay function (36, 125, 127, 128):  $P(N) = c e^{-N/\pi}$ . If a defect has a size smaller than  $15 \text{ \AA}^2$  or probability is lower than  $10^{-4}$ , it was not used in the fitting, consistent with previous work (127). The packing defect constant ( $\pi$ ), as shown in Figure 3-2b, is a helpful comparative number indicative of how quickly the decay function falls off; thus, larger packing defect constants are obtained when there is a higher probability of larger packing defects. We confirmed that the locations of PL acyl defects and deep defects found of a DOPC/DOPE bilayer membrane using our analysis script and the PackMem script (127) are exactly the same except for edges (data not shown). The differences at edges resulted from different consideration of a box boundary and this will go away if a system is big enough.

This algorithm was extended to evaluate the packing defects of LD systems. First, the PL acyl defects and deep defects were calculated by only considering PL molecules (not including TGs). Therefore, the locations in which TGs were exposed to the LD surface were considered to

have a deep defect, even though TG atoms may have overlapped with a grid point ranging from  $z_{thr}$  to  $z_{max}$ . These ‘pseudo’ deep defects were then further specified into three different categories based on their overlap with TG molecules: TG acyl defects, TG glycerol defects, and ‘real’ deep defects. An additional parameter,  $z_{avg}$ , which is defined to be the average Z position of the PL tails, was introduced to set the Z range of a new three-dimensional grid. A three-dimensional grid with spacing of 1 Å, ranging from  $z_{avg}$  to  $z_{max}$  was constructed. If a grid point that was formerly a pseudo deep defect overlaps with TG acyl atoms, it is considered to be a TG acyl defect, adding a value of  $10^{-3}$ . Similarly, if an elementary pseudo deep defect grid point overlaps with a TG glycerol atom, it is considered to be a TG glycerol defect, adding a value of 1. When no TG atoms overlap, this is finally considered a deep defect. There are grid points whose values are contributed by different types of defects. The priority is ranked with a value added to a grid point. For instance, when a grid point overlaps both with a PL acyl atom (adding a value of  $10^3$ ) and a TG acyl atom (adding a value of  $10^{-3}$ ), as the sum of those two is within the range from  $10^3$  to  $10^6$  (exclusive), this grid point is considered an elementary PL acyl defect. In this way, the algorithm differentiates the packing defects caused by PLs and TGs. An illustration of the algorithm is shown in Fig. 3-8.

### *Order parameters*

Order parameters (129) were calculated using second-order Legendre polynomials of  $\cos \theta$ :  $S_{CD} = \frac{1}{2} \langle 3 \cos^2 \theta - 1 \rangle$ . The angle ( $\theta$ ) between the position vector of a carbon atom of an acyl chain to a bonded hydrogen atom with the Z axis was used. The bracket represents the ensemble average.

### *Interdigitation*

The degree of interdigitation between TGs and PLs was calculated consistently with previous studies (36, 130): the density profiles of TGs and PLs with respect the Z axis ( $\rho_{TG}(z)$  and  $\rho_{PL}(z)$  respectively) were first used to define an overlap parameter,  $\rho_{ov}(z)$ :

$$\rho_{ov}(z) = 4 \frac{\rho_{TG}(z) \times \rho_{PL}(z)}{(\rho_{TG}(z) + \rho_{PL}(z))^2}$$

The overlap parameter can range from 0 to 1: it is 0 when one of density profiles is equal to zero and 1 when  $\rho_{TG}(z) = \rho_{PL}(z)$ . In essence, it reflects how the two density profiles (TG and PL) differ at each Z position. The amount of interdigitation ( $\lambda_{ov}$ ) was then obtained by integrating the overlap parameter along the Z axis over the whole box such that,

$$\lambda_{ov} = \int_0^L \rho_{ov}(z) dz,$$

where L is the Z dimension of the simulation box. The quantity reflects the area common to the two density profiles.

### *Water permeation PMF*

Replica-exchange (131) umbrella sampling (132) simulations were run in order to compute the permeation potential of mean force (PMF) for a water molecule through the LD monolayer. The collective variable used as the reaction coordinate was the Z distance between the center of mass of one randomly selected water molecule and the center of mass of the phosphorous atoms of the upper leaflet. Harmonic restraints with a force constant of 700 kJ/mol/nm<sup>2</sup> were placed in each of

the 40 umbrella sampling windows with a 0.15 nm spacing over a range of -3.4 nm to 2.45 nm (0 nm is the average Z position of the upper phosphorous atoms, (+) moves toward water while (-) moves toward the LD core). An initial structure for each window was prepared by running steered molecular dynamics simulation, biasing the same collective variable, with a force constant of 500 kJ/mol/nm<sup>2</sup> for 40 ns. The exchange between windows was attempted every 1000 steps. The PMF was calculated using the Weighted Histogram Analysis Method (WHAM) (133, 134) with a bin spacing of 0.03 nm. The replica-exchange umbrella sampling simulations were run for 100 ns and the trajectories were divided into five blocks to calculate error bars with block averaging. The external plugin, PLUMED2, was used for the biased simulations (135).

At each window, the coordination number between the chosen water molecule and the oxygen atoms of PL or TG was computed:  $s = \sum_{i \in A} \frac{1 - (\frac{r_i}{r_0})^6}{1 - (\frac{r_i}{r_0})^{12}}$  where A is the oxygen atoms of the glycerol moiety of either PL or TG and  $r_i$  is the distance from the oxygen atom  $i$  included in A to the oxygen atom of the chosen water molecule and  $r_0$  is set to 0.25 nm. The normalized coordination number for PL was obtained by dividing the coordination number by the number of PL in the upper leaflet, which is 135. The normalized coordination number for SURF-TG was computed by first selecting the SURF-TG oxygen atoms, computing the coordination number with the biased water molecule, and dividing it by the number of SURF-TG at each timeframe. Here a broader definition of SURF-TG was used: TG molecules with at least one of six oxygen atoms above than the average Z position of the PL acyl chain of the upper leaflet are considered SURF-TG. This was done to incorporate the effects of TG glycerol moieties located at the center of the PL monolayer (e.g., deeply intercalated CORE-TG or those transitioning to/from SURF-TG), which have significant interactions with the biased water molecule.



## RESULTS AND DISCUSSION

### *Reshaping landscapes of LD surfaces*

#### LD structures and lipid-packing defects

Ten  $\mu\text{s}$  long MD simulations of the LD-mimicking trilayer systems were performed to study the structural and physical properties of LDs. After a long equilibration time (5  $\mu\text{s}$ ) the LD surfaces clearly demonstrate partial occupancy by TG molecules (SURF-TG) (Figs. 3-1, 3-9 and 3-10). The glycerol moieties of SURF-TG are predominantly aligned with the glycerol moieties of the PLs, but more inserted toward the core. Thus, our simulations demonstrate much more active intercalation of TGs than did the previous packing defect model of LDs, in which the intrusion of TG tails results in the increased packing defects of the LD surface (36, 82). The intercalation of CORE-TG in this study corresponds to the intrusion of TG tails in the previous packing defect model as SURF-TG was not reported. As shown in Figs. 3-1, 3-9 and 3-10, the dominant TG class on the LD surface is SURF-TG. Although there is significant overlap of CORE-TG and the PL tails, little of it is exposed at the surface.

It has been suggested that lipid-packing defects play a key role in protein targeting of LDs (82, 126), making them potentially an important surface property. We therefore quantified the packing defects for both the bilayer membrane and LD systems (Fig. 3-2). Each defect type was evaluated and the probability of finding a defect size was fitted to the exponential curve (details are shown in the Methods) in order to compute a packing defect constant (Fig. 3-2a). Previous studies have reported the packing defect constant ( $\pi$ ) of deep defects or PL acyl defects for a bilayer membrane to be within the range of 6 to 10  $\text{\AA}^2$  or 7 to 15  $\text{\AA}^2$ , respectively (127). Consistent with literature values, the results from our simulations show the packing defect constant of ERbil for deep defects and PL acyl defects to be  $10.0 \pm 0.2 \text{\AA}^2$  and  $12.5 \pm 0.5 \text{\AA}^2$ , respectively (Fig. 3-

2b). While the LD systems have comparable deep defects with ERbil, the PL acyl defects did increase ( $\sim 15.3 \text{ \AA}^2$ ), and TG acyl defects and TG glycerol defects were significant additions,  $\sim 11.5 \text{ \AA}$  and  $\sim 22.0 \text{ \AA}$ , respectively (Fig. 3-2b). As shown in Fig. 3-2c, each molecular group type is covered with the corresponding defect type. Overall, the LD systems contain substantially more packing defects compared to the bilayer membrane, and the packing defects caused by TGs are predominantly created by SURF-TG. Our simulation results also show that the LD systems contain bigger PL acyl packing defects than the bilayer membrane, potentially contributed to by the intrusion of CORE-TG to the PL tails of the monolayer (Fig. 3-1b and 3-2b). Lastly, it should be noted that the PL and TG packing defect constants are comparable, suggesting that the two defects could compete with each other for interacting with proteins on LD surfaces. Investigating the relationship between the two types of defects on LD surfaces and protein interactions will be pursued in future research.

### *Characterizing the molecular properties of SURF-TG*

#### Order parameters

In order to characterize the molecular properties of SURF-TG, we calculated tail order parameters of the PLs and SURF-TG (Fig. 3-3) in the bilayer and LD systems. (The molecular structures of POPC, DOPE, and TG are shown in Fig. 3-7.) As shown in Figures 3-3a and 3-3b, the addition of DOPE in the ERbil increases the order of POPC molecules, compared to the pure POPCbil. In contrast, the order of the DOPE molecules decreases in the ERbil compared to that in the pure DOPEbil trapped in the lamellar phase (see Methods) (Figs. 3-3c and 3-3d). The changes in order correlate well with the changes in lipid packing density, which can be inferred from the APL

values. For instance, the APL of POPCbil, DOPEbil, and ERbil was 0.645 nm<sup>2</sup>, 0.615 nm<sup>2</sup>, and 0.628 nm<sup>2</sup>, respectively. Thus, the average room for each POPC molecule becomes smaller in the ERbil compared to the POPCbil, resulting in higher packing and increased order. In contrast, DOPE was more highly packed in the DOPEbil (trapped in the lamellar phase), but becomes more loosely packed in the ERbil, resulting in decreased order.

Interestingly, the LD systems track the ERbil quite well, with a slight increase in order in the lower tail order parameters (c12-c16) for both POPC and DOPE. This increased ordering of POPC tails has been previously reported, though not explained, for lipoprotein trilayer systems (99). As described below, the lower PL tails are more densely packed, and thus more ordered, in the LD systems due to intercalated TG molecules.

We then calculated the order parameter of the SURF-TG molecules. All the acyl chains within a TG molecule are identical except for the glycerol moiety (The sn-1 and sn-3 chains have  $\alpha$  carbon, while the sn-2 chain has  $\beta$  carbon; see Fig. 3-7). Therefore, we expected the order parameters of each TG acyl chain to be comparable to one another. We further expected the order parameters of SURF-TG to be comparable to those for DOPE since they have identical acyl chains. Consistent with these expectations, the order parameters of all of the acyl chains on SURF-TG molecules were comparable (Fig. 3-3e). Furthermore, the order parameters of sn-1 and sn-2 chains of SURF-TG were comparable with that of DOPE except for the first several carbon atoms (Fig. 3-3c and 3-3d). It is worth noting that the order parameters of the CORE-TG molecules were zero, meaning that they do not have any order in their structure. Even those CORE-TG that are transiently intercalated with the PL tails are highly dynamic and thus have order parameters very close to zero.

Collectively, the order parameter results demonstrate that SURF-TG are ordered in an analogous manner to DOPE with their glycerol moieties exposed to water and acyl chains extended toward the LD center. This suggests that SURF-TG serve as a secondary membrane component in the LD phospholipid monolayer, a finding that could be tested experimentally by probing TG tail order as a function of depth from the LD surface.

### Influence of intercalated TG molecules

To better understand how the two types of TGs influence LD properties, we quantified their intercalation into the PL monolayer. The overlap parameter ( $\rho_{ov}(z)$ ) profiles (Figs. 3-1b and 3-10) show the expected dominance of SURF-TG closer to the surface of the monolayer, while CORE-TG dominates in the lower PL tail region. Tracking the total amount of interdigitation ( $\lambda_{ov}$ ) (the spatial integration of the overlap) over time, there are clear differences between SURF-TG and CORE-TG. As shown in Figure 3-4,  $\lambda_{ov}^{SURF}$  fluctuates over time in a manner that is highly correlated with the APL. These fluctuations are also highly correlated with the molar ratio of SURF-TG with respect to PLs, calculated as  $N_{SURF-TG}/(N_{SURF-TG} + N_{PL})$  (Fig. 3-11). This further explains why LD systems have higher APLs than bilayer membranes: SURF-TG molecules cause system expansion in the X, Y dimensions. In contrast,  $\lambda_{ov}^{CORE}$  is fairly stable over time, converging to a value of  $\sim 1.83$  nm (Fig. 3-11). Previously reported values of interdigitation varied from 0.73 nm to 1.48 nm depending on force fields and the resolution of the simulations (36). Our simulations, therefore, already suggest a greater degree of interdigitation than previous reports based on CORE-TG alone.

In order to support our findings on the intercalation of SURF-TG, we have compared the APL of our systems with experimental data and previous simulations. The results from our simulations show the APL of ERbil, 8nm LD, and 16nm LD to be  $62.8 \pm 0.1 \text{ \AA}^2$ ,  $69.4 \pm 1.4 \text{ \AA}^2$  and  $69.5 \pm 1.9 \text{ \AA}^2$ , respectively. This 11 % increase in the APL from ERbil to LDs agrees well with the recent experimentally measured 15% increase in the APL from a POPC bilayer to a POPC monolayer surrounding a LD (34). A possible explanation for the slight difference between experiment and our simulations is surface tension. The same work reported the surface tension of a TG-containing LD to be 1.63 mN/m (34), while our simulations were conducted at zero-surface tension. We expect a larger APL increase for simulations run with an applied surface tension. Thus, we argue 11% increase of the APL at zero-surface tension agrees well with 15% increase at 1.63 mN/m. A more detailed study on the influence of surface tension on APL and the amount of SURF-TG will be the focus of future work. On the other hand, a previous short LD simulation, performed for 100 ns, reported only a 2% increase in the APL, which is most likely due to early interdigitation with CORE-TG (99). In contrast, reported united-atom and MARTINI simulations reported nearly identical APL between a bilayer membrane and a LD (36), likely due to limitations in the forcefield and coarse-grained representation. Taken together, the APL analysis shows agreement between our simulations and experiments, and further validates the existence (and amount) of SURF-TG.

To better understand the impact of intercalated CORE-TG on PL ordering, we consider the four-region model of a bilayer membrane by Marrink and Berendsen (136), wherein the lower PL tail region (with the width of 0.55 nm for each leaflet) is the low-density region, with density comparable to liquid hexane. Overall, our ERbil tracks the four-region model well, as shown in Fig. 3-12. The low-density region (-1.5nm ~ -2 nm) was apparent in ERbil (black solid line). The

8nm LD system, however, shows higher density (red solid line) due to interdigitated CORE-TG. The increased density in this region results in higher PL ordering (c12-c16) for both POPC and DOPE (Fig. 3-3). Taken together, the two types of intercalated TG molecules have different physical impact on the monolayer properties, with SURF-TG directly influencing surface properties like APL, while CORE-TG consistently, though dynamically, intercalates with the PL tail region, resulting in slightly higher density and order in the lower tail region.

### *Hydration of LDs*

Finally, we focused on the hydration of our LD simulations. Recent work has noted a significant number of water molecules below the LD PL monolayer in the presence of proteins (78). We computed the water density profiles at various simulation times (Fig. 3-5a) and observed significantly more water in the LD core (~8g water/1kg oil or 0.8% kg/kg) than previously reported (~1g water/1kg oil or 0.1% kg/kg) (36). Different from alkanes, a TG molecule has a polar glycerol moiety (see TG structure in Fig. 3-7), which can stabilize water molecules via hydrogen bonding. Consistent with this explanation, the calculated water density in the LD core is clearly proportional to the density of the TG glycerol moieties (Fig. 3-13).

In order to verify convergence of the amount of water in the LD core, we first computed the average water density in the LD core by integrating the water density profile between two monolayers and then divided by the height (Z dimension) between two. The resulting average water density (Fig. 3-5b) converges at ~8  $\mu\text{s}$  for both the 8nm and 16 nm LD systems. The middle peak in 8nm LD (Fig. 3-5a top) results from a clustering of glycerol moieties and is indicative of short-range order due to the system size (Fig. 3-13 top). This residual structure mostly disappears

in the larger 16 nm LD system (Fig. 3-5a bottom and Fig. 3-13 bottom) and explains the slightly higher water content in 8nm LD (Fig. 3-5b). To further validate our hydration results, we next calculated the water permeation PMF (a free energy profile) with replica exchange umbrella sampling (see Methods). The resulting PMF is consistent with the free energy profile computed from the water density profile,  $F(z) = -k_B T \log \rho(z)/\rho_0$ , where  $\rho(z)$  is the water density with respect to  $z$  and  $\rho_0$  is the bulk water density (Fig. 3-6a). Together with the convergence of water density with simulation time (Fig. 3-5b), the enhanced sampling results (Fig. 3-6a) confirm our hydration results are converged and reliable within the limits of the forcefield. TIP3P has been shown to underestimate permeation into hydrophobic regions due to over-polarization (137) and future efforts to develop polarizable models may shift the quantification we report herein.

For comparison we include the PMF for water permeation into a POPC bilayer membrane from the work by Venable, Kramer, and Pastor (32). Both the bilayer and LD PMFs peak around -1.6 nm, where the PL density is shifting from high- to the lower-density and where the density of oxygen atoms is lowest in the LD (Figs. 3-6a, 3-6b, and 3-12). However, the energy barrier to partition from the water phase to the center of the LD is significantly less (by 2.5 kcal/mol) than that to the bilayer center. Given this discrepancy occurs around -0.8 nm, where the glycerol moieties of SURF-TG are located, we hypothesized the reduction could be due to water hydrogen bonding with SURF-TG glycerols. Unlike PL, SURF-TG can transition to CORE-TG or CORE-TG to SURF-TG, which extends the region where the water molecule can form hydrogen bonds. As discussed above, the glycerol moieties of SURF-TG are also closer to the LD core by ~0.4 nm, additionally extending the hydrogen-bonding region. We confirmed our hypothesis by computing the normalized coordination number (Fig. 3-6c) between the oxygen atom of the biased water molecule and the oxygen atoms in the glycerol moieties of PL or SURF-TG (see Methods).

Interestingly, the normalized coordination analysis shows that the water molecule preferentially interacts with the SURF-TG glycerols (gray circles in Fig. 3-6c) over the PL glycerols (black circles). We expect this is because pulling the PL glycerol moiety toward the LD core accompanies membrane deformations, while pulling CORE-TG to SURF-TG or SURF-TG to CORE-TG has a much lower energy penalty.

Another comparison that can be made is with the experimentally measured water moisture of olive oil, whose main component is TG. Although a water-in-oil emulsion is a fundamentally different from an oil-PL-water LD due to the lack of amphiphiles (PL) and different sizes, this comparison offers a rough idea of the range of expected hydration. Depending on the origin of olive oil, reports of the moisture content vary greatly: from less than 0.2% kg/kg in (138, 139) to up to 0.8% kg/kg (140). Our hydration value is within this range at ~0.8% kg/kg.

Taken together, the results presented herein demonstrate significant hydration within the LD core with a slight increase just below the PL tails due to increased concentration of TG glycerol moieties from intercalated CORE-TG. The water density at the PL/TG interfacial region is expected to be important in the stabilization of class I LD proteins that have charged residues near their hairpin kink (78). More generally, this degree of hydration may be an important factor to consider not only class I and II protein-LD interactions, but also in the association of amphipathic molecules with LD cores.

## CONCLUSIONS

LDs are distinct from other bilayer-bound organelles due to their monolayer surface of PLs and neutral lipid core. Using the largest reported LD models (8 nm and 16 nm TG layers) and the



longest reported simulation times (10  $\mu$ s) for all-atom resolution, the simulations presented herein provide new perspectives on the surface and hydration properties of LDs. Two types of TG molecules are found, those at the surface (SURF-TG) and those towards the core (CORE-TG), that interact with, and influence, the PL monolayer in distinct ways. The SURF-TG occupy 5-8% of LD surfaces, which is greater than the solubility reported for TG in a POPC PL bilayer (2.8%) (37). The glycerol moieties in SURF-TG are largely exposed to water, and the tail order parameters, which are indicative of structure and order, are very similar to those found in PLs, especially DOPE. Thus, SURF-TG seem to behave as a membrane component, substantially increasing the amount of packing defects at the monolayer-water interface and creating chemically unique defects over neutral glycerol moieties. In contrast, the CORE-TG intercalate into the PL tails, retaining the disorder expected in the LD core, and potentially contributing to the increase in PL packing defects found in our simulations. We also demonstrate that the APL, which is representative of expansion of the membrane, fluctuates in a highly correlated manner with the amount of interdigitation by SURF-TG (the number of SURF-TG), while the amount of interdigitation by CORE-TG is stable and does not influence the size of the membrane. Finally, we observe water density in the LD core to be 0.008 g/cm<sup>3</sup>, which is an order of magnitude higher than previously reported (36).

Equilibration across the LD monolayer was a slow process ( $> 5 \mu$ s) during which TG molecules diffuse into the PL monolayer while water diffuses into the LD core. Although previous atomistic studies did report some degree of intercalation of TG from the LD core (one can see the intrusion of TG tails into the PL tails), the presence of SURF-TG was not observed (36, 82, 99). This was likely the result of limited simulation time ( $< 1 \mu$ s). Consistent with this, the SURF-TG in our simulations did not become evident until after 1  $\mu$ s. In contrast, CG studies performed long

MD simulations of large LD systems or oil lenses (the initial stage of the LD formation); however, no observation of SURF-TG or significant LD hydration was reported (36, 89, 90, 101, 102). We suspect that this is due to the challenging limitations of accuracy, transferability and entropy in CG modeling (103-105).

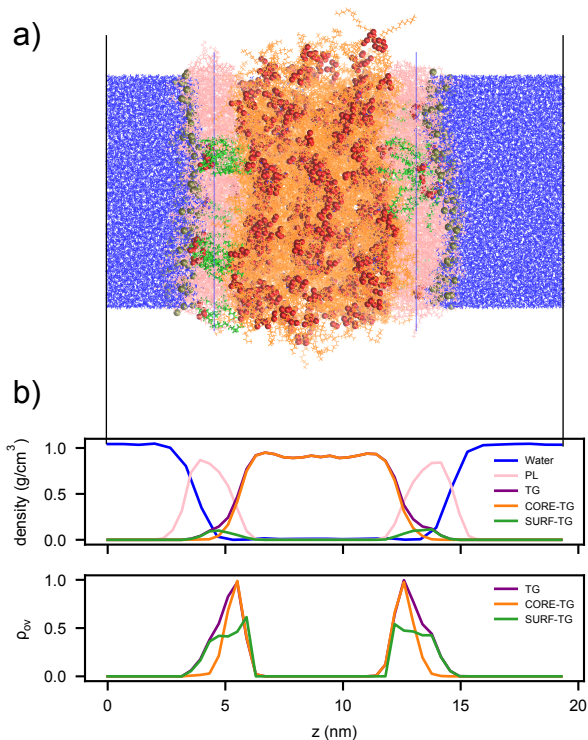
The presented results provide new perspectives on LD surfaces in the context of protein targeting mechanisms. A previous study proposed that proteins containing amphipathic helices can bind to the packing defects of LDs, which were therein reported to be slightly larger and more persistent than bilayer membranes (82). Our results, however, suggest there are markedly larger packing defects than previously reported, in addition to chemically unique defects created by SURF-TG. We anticipate the targeting mechanisms will be significantly influenced by these SURF-TG defects. For example, the first step and rate-limiting step of lipolysis is carried out by adipose triglyceride lipase (ATGL), which converts a triglyceride into a diglyceride and fatty acid (141-143). However, how ATGL resting on LD surfaces can reach the glycerol moiety of TG across the PL monolayer is unknown. With this new model, the SURF-TG molecules, and especially the carbonyl carbons, are exposed to the cytosol, easily accessible and in close proximity to the second hydrolysis substrate, water. This targeting model in which neutral lipids exposed at LD surfaces interact with a protein is further supported by the recent findings of LDs' differential protein recruitment. Chorlay and Thiam found different targeting preferences by amphipathic peptides to LDs with different neutral lipid compositions but the same phospholipid composition and density (34). This differential recruitment was maintained with different surface tension although the binding levels changed. Our results suggest that quite significant exposure of neutral lipids at the LD surface is a likely explanation of differential recruitment.

We also anticipate that SURF-TG may play an active role in modulating the surface tension in LDs. It is an outstanding curiosity how LDs, which shrink with lipolysis and grow with TG synthesis, maintain similar growth/shrinkage in their monolayer structures. For a given LD size, there is an optimal number of PLs in the monolayer (144). Our finding suggests that in the case of too few PLs, TGs can shift to SURF-TG to act as a secondary monolayer component and thereby reduce the monolayer surface tension. This is consistent with experimental data that report the surface composition of water/(PL + TG)/air as a function of surface tension (33). When the surface is under expansion, both TGs and PLs are mixed and surface-active. However, when it is compressed, TGs demix from the PL layer and form a separate phase in air because PLs are more surface-active than TGs. Thus, we predict that SURF-TG can reduce surface tension by being a secondary monolayer component when the supply of PLs is limited or during the initial stages of the LD formation (38, 47). In turn, we also anticipate that the percentage of exposed SURF-TG may influence the LD lifecycle, e.g., recruiting proteins to synthesize more PLs.

Finally, increased water density is found in the LD core. Even though it is present in relatively small quantity, water in the LD core can play a crucial role in stabilizing charged or polar residues embedded in the LD monolayer. For example, proteins that target LDs from the ER often contain a hydrophobic hairpin motif, which has been proven to be sufficient for LD localization (65, 78). These hairpin motifs include highly conserved charged residues at hinge and mutation of those residues to hydrophobic residues cause the defects to LD targeting (78). In the ER membrane, those charged residues are stabilized by interacting with the phosphate group of the lower leaflet; water can do the corresponding work in LD core. Future work will explore the influence of mixing sterol esters into the pure TG core studied here, the influence of surface tension, as well as the ways in which the reported properties influence LD protein targeting.

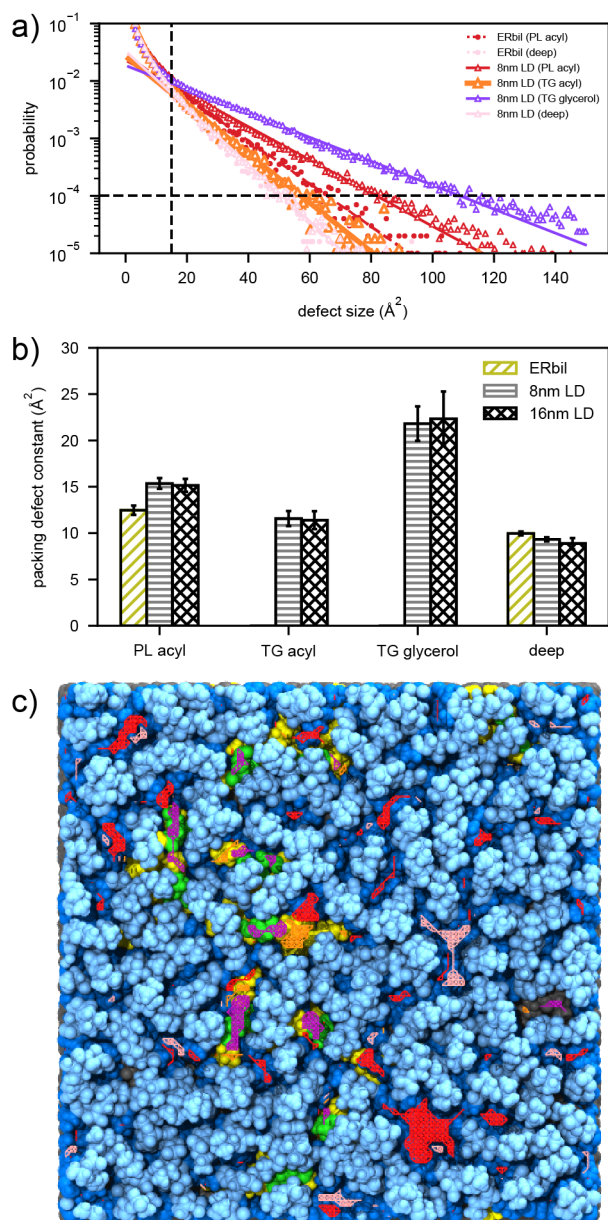
## **ACKNOWLEDGEMENTS**

This research was supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) through grant 5R01GM097194. We thank Chenghan Li, Dr. Jeeyun Chung, Dr. Zack Jarin, Dr. Rich Pastor, Dr. Bernard Brooks, Dr. Robert V. Farese, Jr., Dr. Tobias C. Walther, and Dr. Gregory A. Voth for useful discussions. S.K. acknowledges Won Hee (Harry) Ryu for reviewing the manuscript. The MD simulations in this research were performed on the Stampede2 supercomputer at the Texas Advanced Computing Center (TACC) through allocation MCB200018 with resources provided by the Extreme Science and Engineering Discovery Environment (XSEDE) supported by NSF grant ACI-1548562. The biased simulations were performed on the high-performance GPU cluster (GM4) at the University of Chicago Research Computing Center, supported by NSF grant DMR-1828629.



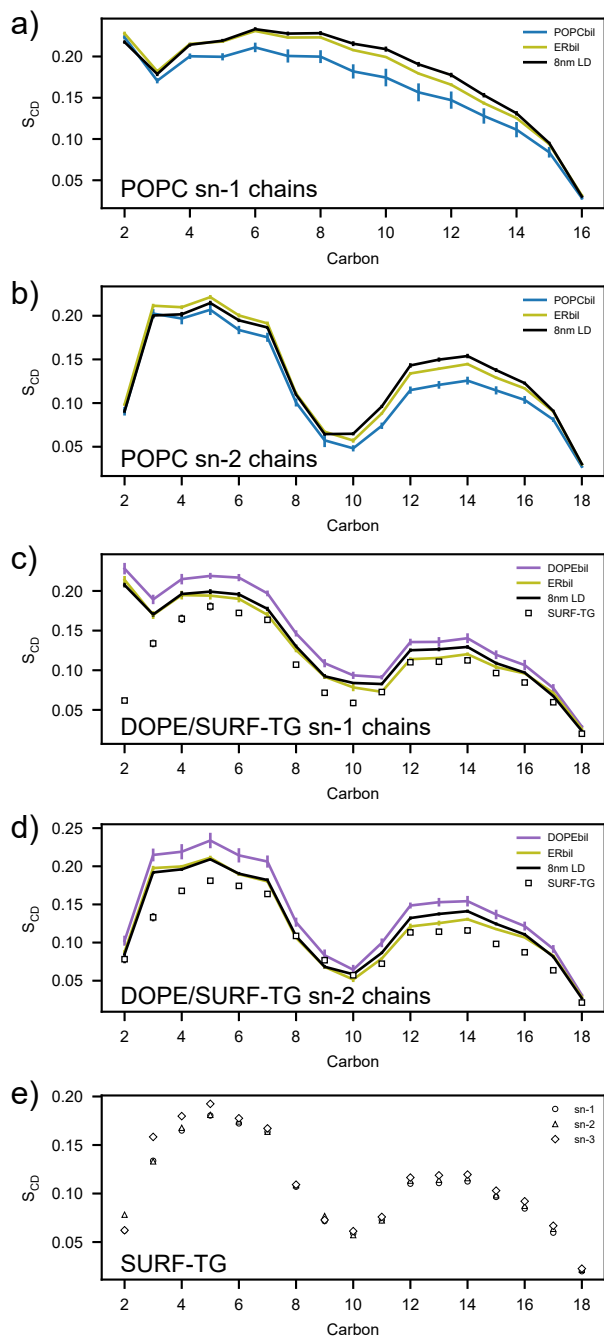
**Figure 3-1** LD structure and density profile.

(a) The last snapshot of 8nm LD simulation. Water molecules are shown in blue and PLs in pink. Gray and red space-filling representations indicate phosphorous atoms of PLs and oxygen atoms of TG molecules, respectively. SURF-TG is shown in green and CORE-TG in orange. Blue lines indicate the average Z position of PL tails. (b) The corresponding density profiles (top) and overlap parameter profiles (bottom) along the Z coordinate of the snapshot of a). The black lines are for the alignment between a) and b). The time-averaged density and overlap profiles of 8nm LD are shown in Figure 3-10a.



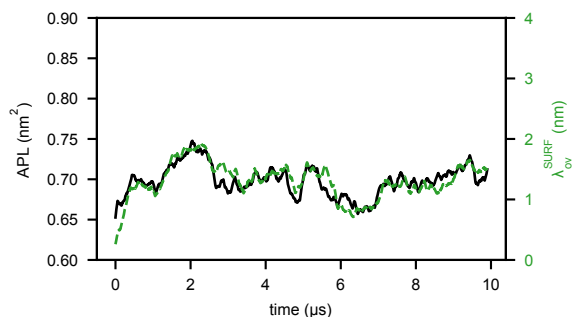
**Figure 3-2** Packing defects of ERbil and LDs.

(a) Packing defect distribution of heterogeneous ERbil (circle marker and dashed line) and 8nm LD (triangle marker and solid line) for PL acyl (red), TG acyl (orange; thicker than other lines), TG glycerol (purple), and deep (pink) defects. (b) The packing defect constants for PL acyl, TG acyl, TG glycerol, and deep defects. The error bars were obtained using the block averaging method with five blocks. (c) Co-localization of defects and the last snapshot of 8nm LD. Color coding for defects is as in a). The light and dark blue indicate polar groups (head groups and glycerol moieties) and acyl chains of PLs, respectively. The green and yellow indicate glycerol moieties and acyl chains of SURF-TG, respectively, and the black CORE-TG.

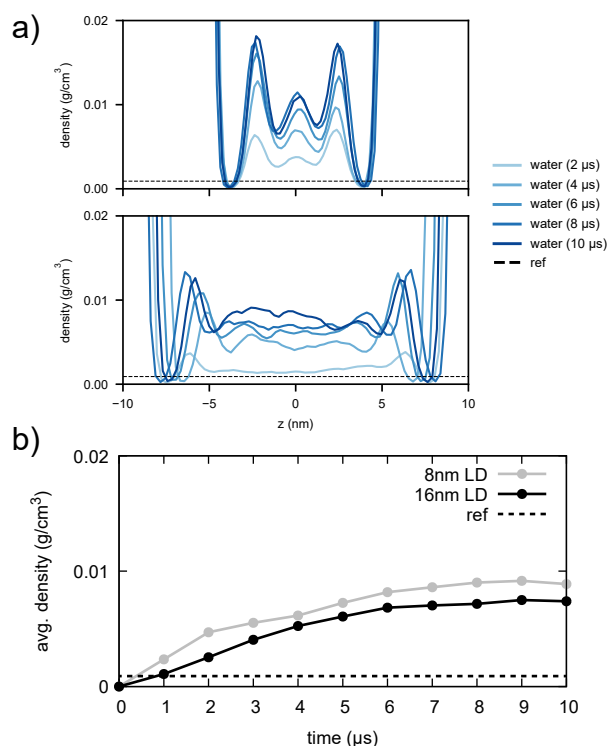


**Figure 3-3** The order parameters of POPC, DOPE, and TG molecules in POPCbil, DOPEbil, ERbil, and 8nm LD.

The error was estimated by taking the standard deviation from the five blocks of the equilibrated trajectories. (a) POPC sn-1 chains. (b) POPC sn-2 chains. (c) DOPE (lines) and SURF-TG (marker) sn-1 chains. (d) DOPE (lines) and SURF-TG (marker) sn-2 chains. (e) SURF-TG sn-1, sn-2 and sn-3 chains.



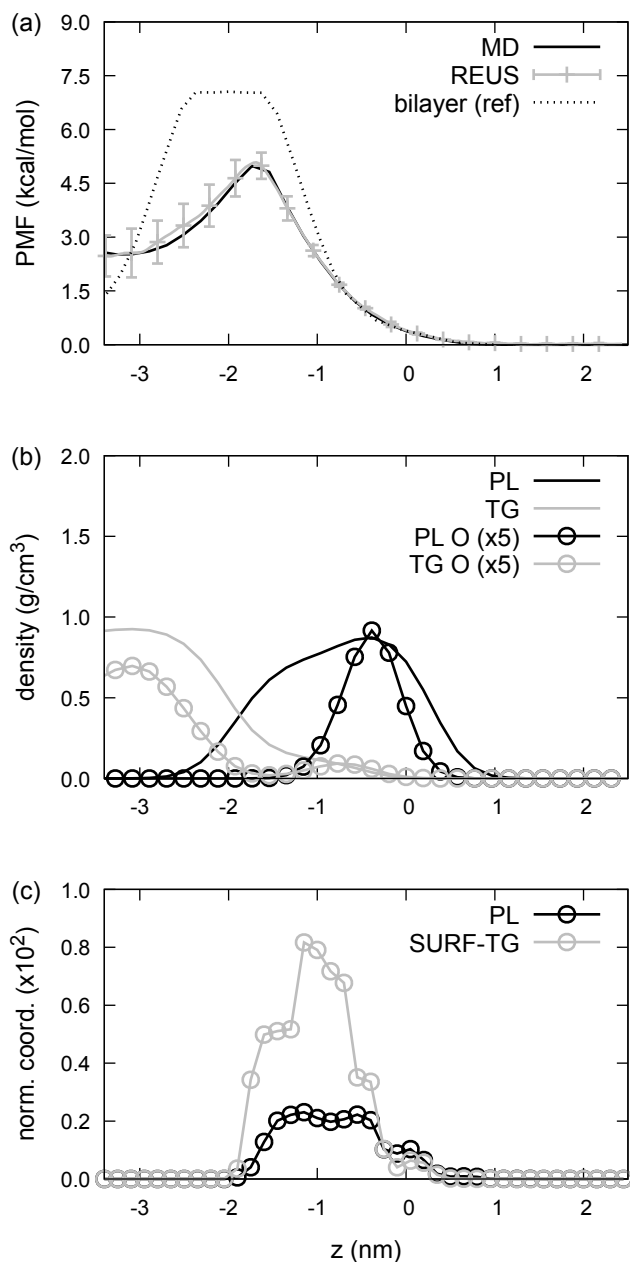
**Figure 3-4** The area per phospholipid (black line) and the amount of interdigitation of PLs by SURF-TG (green dashed line) for 8nm LD. For comparison, the APLs of POPCbil, DOPEbil, and ERbil were 0.645 nm<sup>2</sup>, 0.615 nm<sup>2</sup>, and 0.628 nm<sup>2</sup>, respectively.



**Figure 3-5** The water density profiles at simulation times.

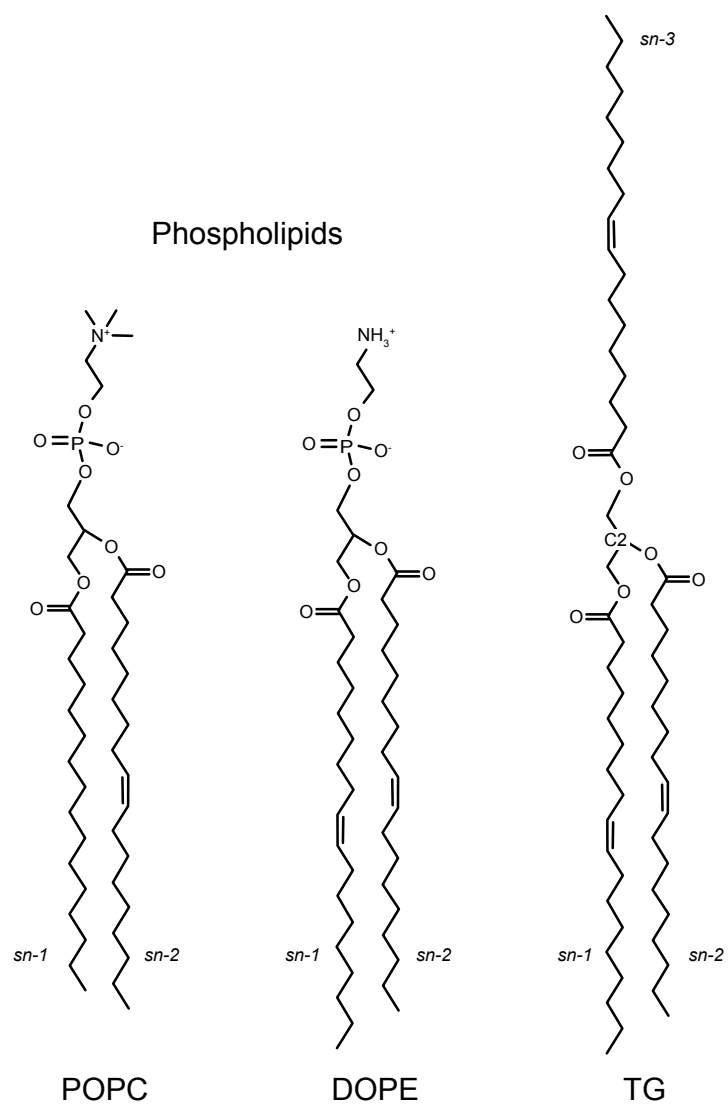
(a) The water density profiles, shown every 2 μs, of 8nm LD (top) and 16nm LD (bottom). The center of the phosphorus atoms (similar to the center of a LD) was zeroed. (b) The (spatially) average water density in the LD core, shown every 1 μs. The reference value of water (dashed) was from Bacle et al.



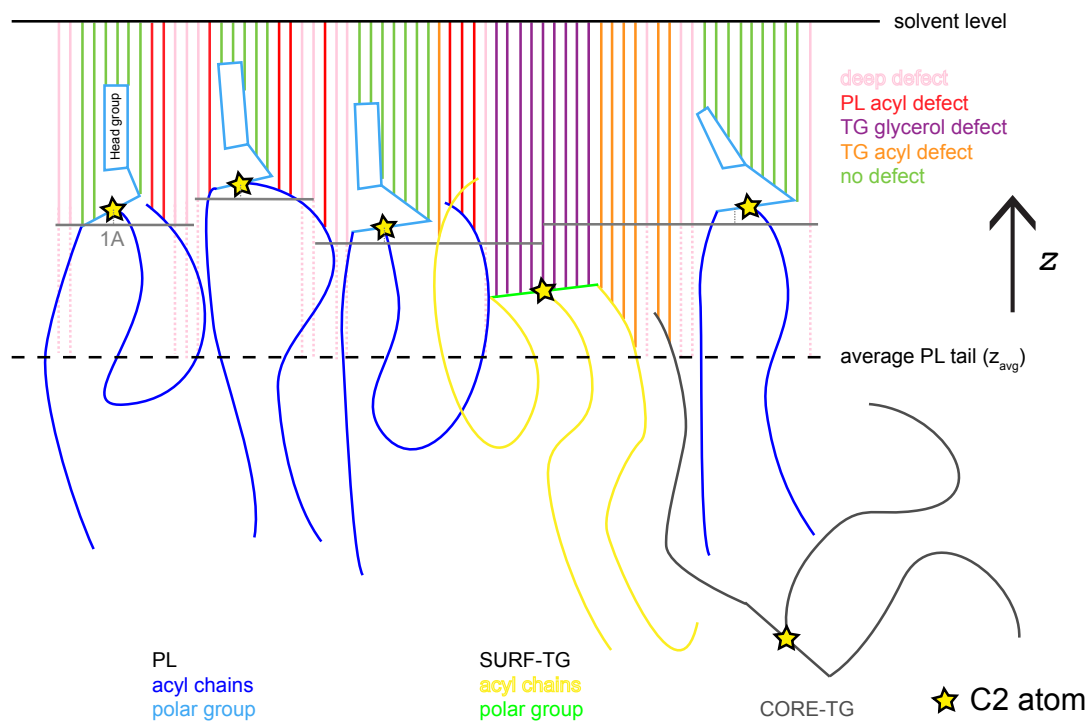


**Figure 3-6** Water permeation into LD.

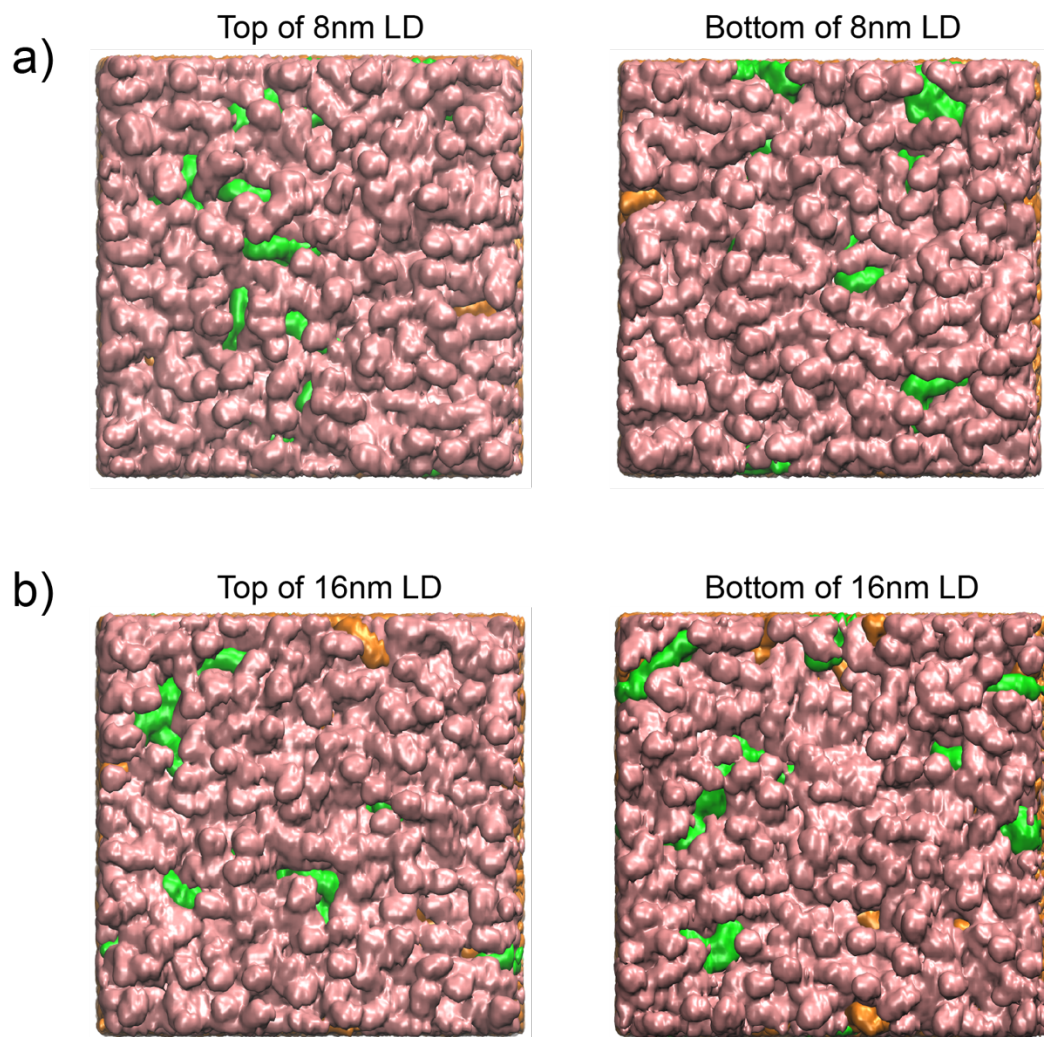
(a) PMFs obtained from the last 2  $\mu\text{s}$  of the 8nm LD simulation (black line) and from replica exchange umbrella sampling (gray). The error bars were obtained using the block averaging method. PMF for water permeation into a bilayer membrane (dashed black line) was obtained from Venable et al. (b) Density of PL (black line), TG (gray line), and respective oxygen glycerol atoms (circles) from the last 2  $\mu\text{s}$  of the 8nm LD simulation. For visual clarity, the oxygen density was multiplied by 5. (c) The normalized coordination number between the water molecule and a PL glycerol (black circles) or SURF-TG glycerol (gray circles). For visual clarity, the normalized coordination was multiplied by 100.



**Figure 3-7** The molecular structures of POPC, DOPE, and TG.

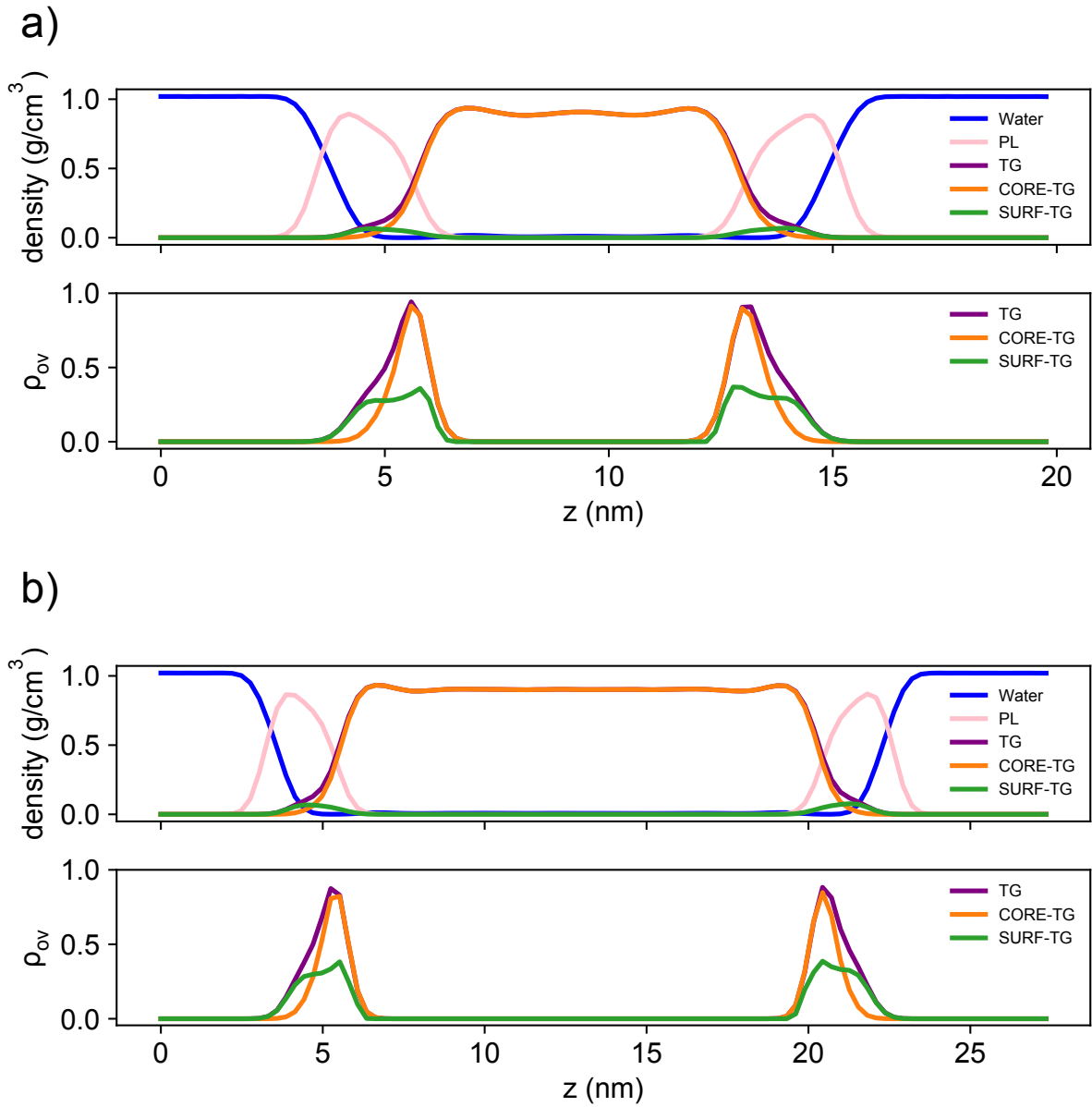


**Figure 3-8** The illustration of the packing defect analysis.

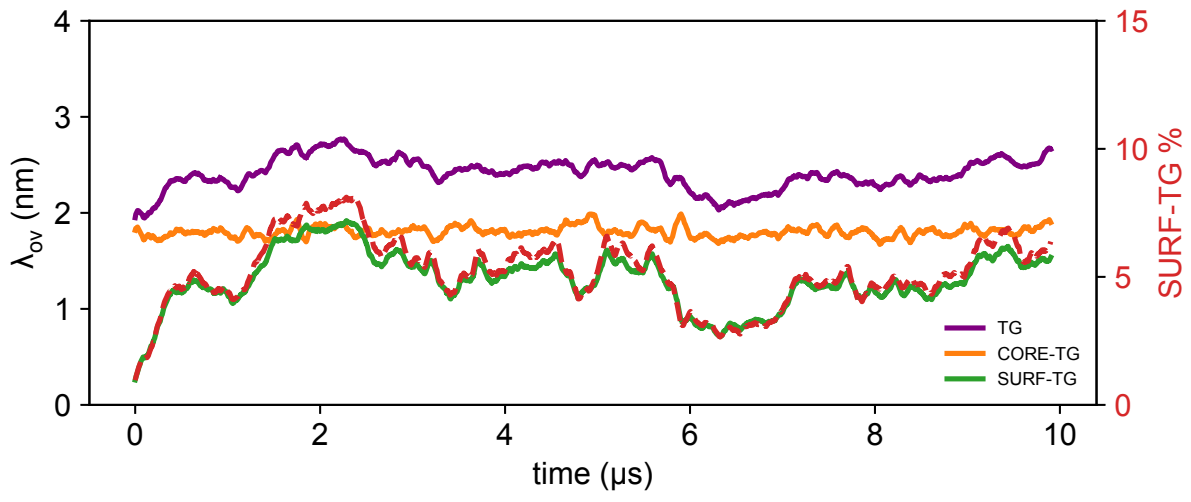


**Figure 3-9** The structures of the last snapshots of 8nm LD and 16nm LD.

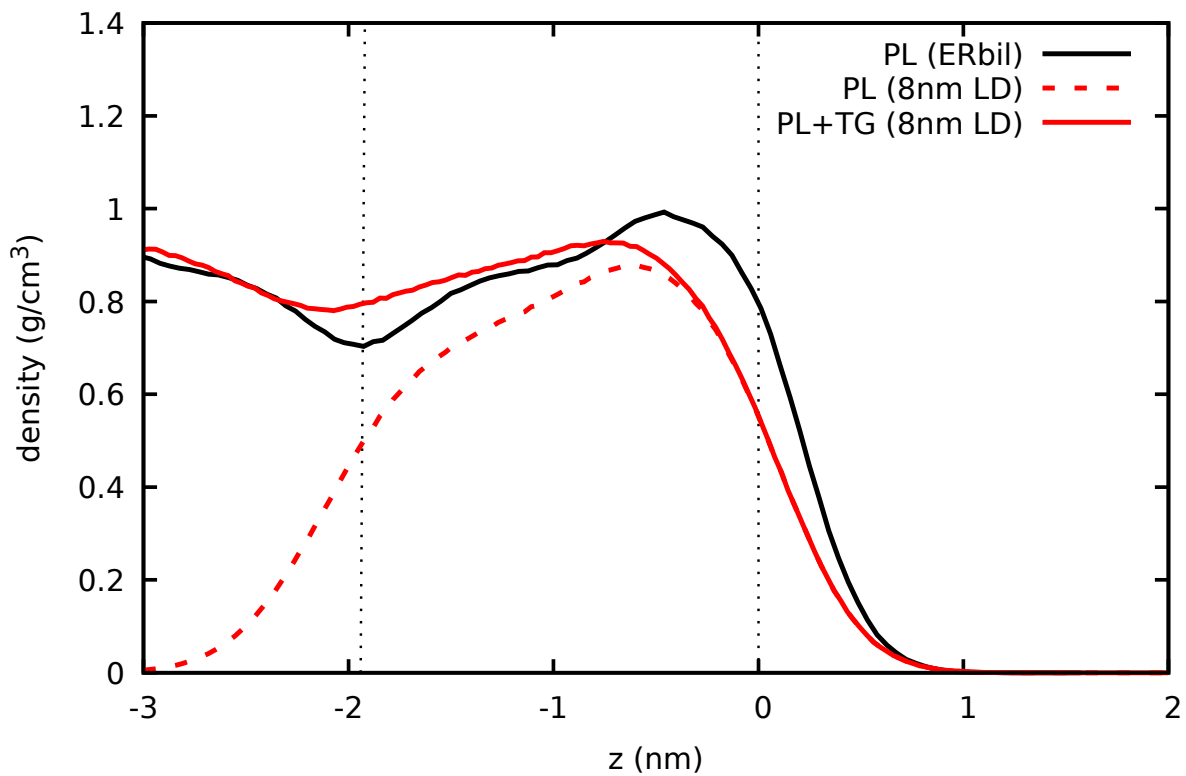
PLs, SURF-TG, and CORE-TG are shown in pink, green, and orange, respectively. Water is not shown for visual clarity. (a) The top view (left) and the bottom view (right) of 8nm LD. (b) The top view (left) and the bottom view (right) of 16nm LD.



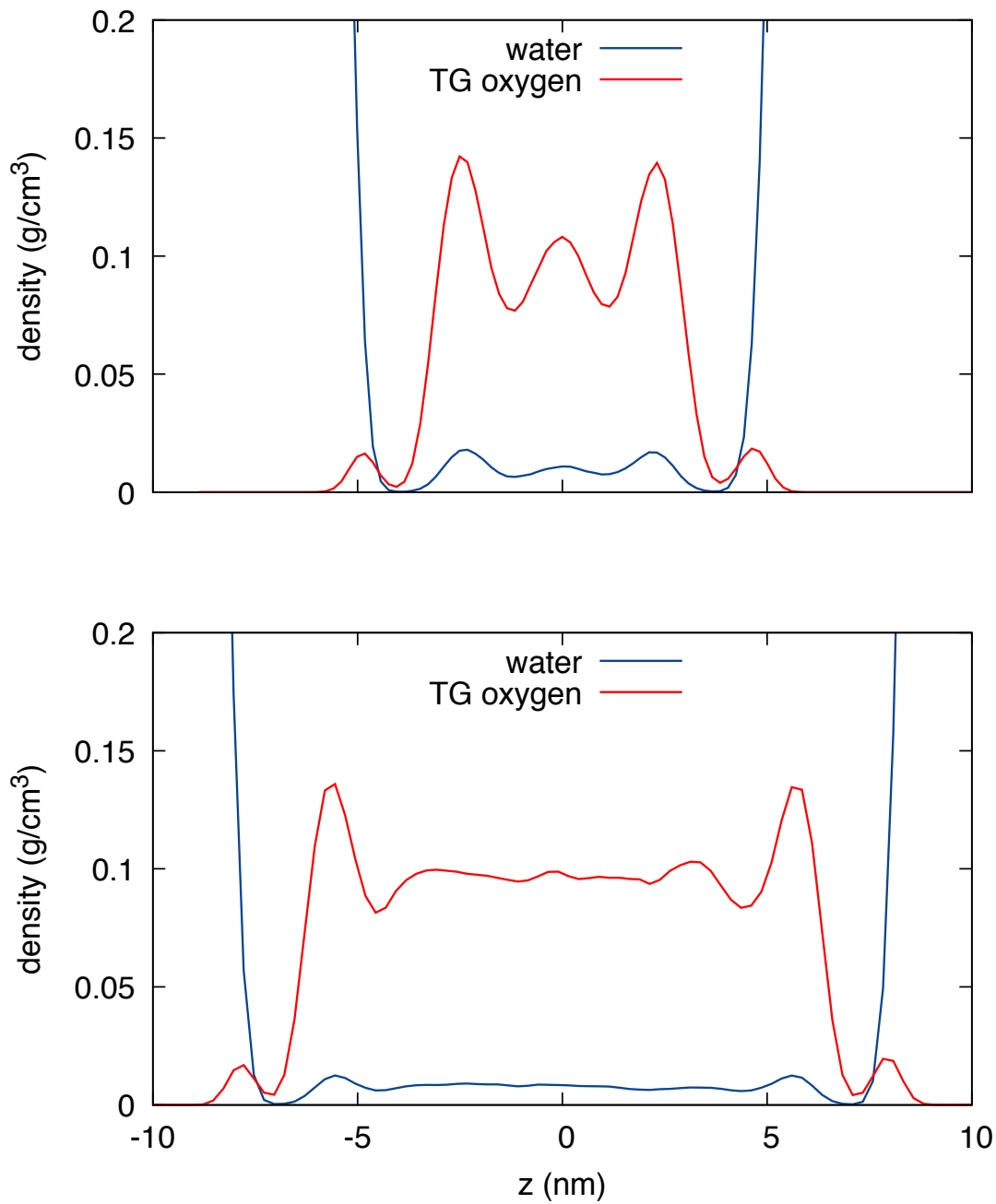
**Figure 3-10** The time-averaged density and overlap parameter profiles for (a) 8nm LD and (b) 16nm LD.



**Figure 3-11** The amount of interdigitation of PLs by TG (purple line), CORE-TG (orange line), and SURF-TG (green line) and the SURF-TG % (red dashed line) with simulation time for 8nm LD.



**Figure 3-12** The density of PL for ERbil (black solid line) and 8nm LD (red dashed line) and PL + TG for 8nm LD (red solid line), averaged from 8  $\mu$ s to 10  $\mu$ s. The average Z position of the phosphorus atoms of the upper leaflet was zeroed. Water is toward (+) direction and the membrane center toward (-) direction.



**Figure 3-13** The density of water (blue) and TG oxygen atoms (red) for 8nm LD (top) and 16nm LD (bottom), averaged from 8  $\mu$ s to 10  $\mu$ s. The center of the phosphorous atoms (similarly the center of a LD) was zeroed.



## Chapter 4 Determinants of endoplasmic reticulum-to-lipid droplet protein targeting

This chapter is adapted from [Olarde et al., 2020] (66).

### ABSTRACT

Lipid droplet (LD) formation from the endoplasmic reticulum (ER) is accompanied by the targeting and accumulation of specific hydrophobic, membrane-embedded proteins on LDs. The determinants of this process are unknown. Here, we study the hydrophobic membrane motifs of two *Drosophila melanogaster* proteins, GPAT4 and ALG14, that utilize this pathway, and we identify crucial sequence features that mediate LD accumulation. Molecular dynamics simulations and studies in cells reveal that LD targeting of these motifs requires deeply inserted tryptophans that have lower free energy in the LD oil phase and positively charged residues near predicted hairpin hinges that become less constrained in the LD environment. Analyzing hydrophobic motifs from similar LD-targeting proteins, it appears that the distribution of tryptophan and positively charged residues distinguishes them from non-LD targeting membrane motifs. Our studies identify specific sequence features and principles of hydrophobic membrane motifs that mediate their accumulation on LDs.

## INTRODUCTION

Lipid droplets (LDs) are unusual organelles, consisting of an organic phase of neutral lipids, such as triacylglycerols (TGs) and sterol esters, bounded by a monolayer of phospholipids (145, 146). This oil-water interface harbors specific proteins that vary in different cell types and species (24, 147-150). LD proteins number in the tens to hundreds, and many are involved in lipid metabolism. For instance, enzymes of both TG synthesis (65, 151) and lipolysis (152) localize to the surface of LDs. Additionally, enzymes for the biosynthesis of ergosterol (in yeast) and phosphatidylcholine localize to LDs (86, 148).

Proteins target LDs from the cytosol or from the endoplasmic reticulum (ER) bilayer membrane (28, 29, 153). LDs appear not to possess dedicated protein targeting machinery, such as the translocon in the ER, nor are they known to utilize biochemical landmarks, such as phosphoinositides, to recruit and bind proteins. Instead, proteins target LDs by recognizing the unique physical properties of the phospholipid monolayer. Molecular dynamics simulations showed that surface phospholipid packing defects are more frequent, more persistent, and larger on LDs than on bilayer membranes (36, 154). Cytosolic proteins with amphipathic helices recognize these phospholipid packing defects at the LD monolayer surface (28, 154).

Proteins may also access the LD monolayer surface from the ER bilayer membrane. Such targeting requires membrane continuity between the outer leaflet of the ER membrane and the monolayer surface of LDs, either when LDs bud from the ER or at later steps in LD biogenesis via ER-LD bridges (65, 77). Glycerol-3-phosphate acyl transferase 4 (GPAT4) catalyzes the first step in glycerolipid biosynthesis (155, 156) on the cytosolic leaflet of the ER membrane (157), and is an important substrate for this targeting pathway. In *Drosophila* S2 cells, during LD induction, GPAT4 relocates from the ER and accumulates on a subset of LDs, where it helps mediate their

expansion (65). Although this targeting reaction typically happens hours after LD formation is initiated, a 56-amino acid, hydrophobic, membrane-embedded motif of GPAT4 (known as *LiveDrop*) is sufficient to localize to LD surfaces during their formation (65, 158). Further, *LiveDrop* not only targets LDs but rapidly accumulates on their surface, making it an excellent model substrate to investigate the molecular mechanisms that drive the accumulation of hydrophobic, membrane-embedded motifs on the surface of LDs.

Here, we utilized *LiveDrop* and the membrane-embedded portion from another *D. melanogaster* LD protein, UDP-N-acetylglucosaminyltransferase subunit (ALG14), as model substrates to dissect the ER-to-LD targeting mechanism. We identify specific sequence requirements and define principles for the accumulation of these hydrophobic, membrane-embedded motifs on LDs. Analysis of motifs from other LD proteins utilizing this pathway suggests that these are general sequence determinants.

## RESULTS

### *The Hydrophobic, Membrane-Embedded Motif LiveDrop Accumulates on Nascent LDs*

We sought to identify a minimal sequence motif that is sufficient to mediate GPAT4 accumulation on LDs. Consistent with previous reports (65, 158), a central hydrophobic sequence of *D. melanogaster* GPAT4 (*LiveDrop*, amino acids 160–215), which is predicted to be membrane-embedded (Fig. 4-1A), targeted to and accumulated on LDs induced by oleate incubation (Fig. 4-1B, 4-1C). Similar to full-length GPAT4, *LiveDrop* exhibited little depletion of signal from LDs during repeated bleaching of the ER, indicating that the protein has a slow off-rate once it is bound to LDs (Fig. 4-1D). Also, *LiveDrop* targeted to LDs faster than full-length GPAT4; all nascent

LDs contained *LiveDrop* as early as 30 min after LD induction, whereas full-length GPAT4 accumulated on LDs after hours of incubation with oleate (Fig. 4-1B, 4-1C) (65). For this and similar experiments, we measured expression levels of the tested constructs by quantifying the fluorescence signal of each protein variant per cell. For instance, *LiveDrop* was expressed in cells at levels similar to or higher than full-length GPAT4. The different timing of LD targeting for full-length GPAT4 and *LiveDrop* suggests that there is a mechanism that prevents full-length GPAT4 from targeting to newly forming LDs.

To test whether the earlier targeting of *LiveDrop* to LDs may be due to its smaller size compared with full-length GPAT4, we increased its size by fusing fluorophores to both its N- and C-termini. This modification caused a modest reduction in LD targeting at 3 h, but the *LiveDrop* variant was still highly enriched on LDs, suggesting that features other than size prevent the targeting of full-length GPAT4 during initial LD formation. Indeed, a C-terminal segment of full-length GPAT4 (amino acids 216–458) was necessary and sufficient for delaying LD targeting. How this C-terminal region of the protein exerts that is currently unknown.

#### *Testing Several Models for LiveDrop Accumulation on LDs*

In one model for ER-to-LD protein targeting and accumulation, proteins equilibrate between the ER and LDs, but the ER pool is selectively degraded, leaving only the LD pool (159). To specifically test for a role of ER-associated protein degradation (ERAD) in LD targeting, we inhibited VCP/p97, an ATPase important for the extraction of proteins from the ER membrane (160). Inhibiting VCP/p97 led to the accumulation of ubiquitylated proteins (Fig. 4-1E) but did not increase the visible ER pool or notably change the LD enrichment of *LiveDrop* (Fig. 4-1F, 4-

1G, 4-S11). Thus, selective degradation of the ER protein pool is unlikely to explain the LD accumulation of *LiveDrop*.

In an alternative model, GPAT4 could bind to a protein on LDs, thereby causing its accumulation on LDs. However, this seems unlikely because overexpressing *LiveDrop* does not saturate LD targeting. It is also improbable because *D. melanogaster LiveDrop* targets LDs in evolutionarily diverse species, such as yeast (159) and human (38, 158), in which protein-protein interactions are generally not conserved.

We also considered whether global, structural features of *LiveDrop* account for its LD accumulation. Specifically, the *LiveDrop* sequence contains two hydrophobic, likely  $\alpha$ -helical stretches of amino acids that are separated near the midpoint by a methionine (M184) and a proline (P185) (Fig. 4-2A). With 21 amino acids on each side of these residues, the two hydrophobic stretches are relatively long for transmembrane domains. Thus, we considered whether this results in a mismatch between the transmembrane domain lengths and the ER membrane thickness, potentially putting *LiveDrop* under a strain that could be relieved by targeting to LDs, where the hydrophobic phase is essentially unlimited in thickness. This hypothesis has analogies to protein sorting along the secretory pathway, where the protein transmembrane domain length generally matches the bilayer thickness of the targeted organelle (161). However, shortening each of the hydrophobic  $\alpha$ -helical segments of *LiveDrop* by four amino acids (roughly one  $\alpha$ -helical turn) did not affect its LD targeting, suggesting that hydrophobic mismatch is not a major contributing mechanism. Moreover, M184 and P185 of the *LiveDrop* sequence were not required for LD targeting, further suggesting a rigid hairpin is not strictly needed for LD targeting.

### *LD Accumulation of LiveDrop Requires Specific Sequence Features*

Since the overall hydrophobic character and many specific residues of the *LiveDrop* membrane-embedded sequence are evolutionarily conserved (Fig. 4-2A), we generated a scrambled version of *LiveDrop* with the same length and hydrophobicity but different amino acid order (Fig. 4-2B). The scrambled variant inserted into the ER membrane with the same topology as wild-type *LiveDrop* (N- and C-termini in the cytosol) but failed to accumulate on LDs (Fig. 4-2C, 4-2D). Similarly, a synthetic hydrophobic motif, with a sequence of amino acids designed according to their propensity to form an  $\alpha$ -helix in bilayer membranes (Fig. 4-2B) (162), localized to the ER and had the same topology as wild-type *LiveDrop* but failed to accumulate on LDs (Fig. 4-2C, 4-2D). These results indicate that both the specific amino acid order and composition of the *LiveDrop* sequence are crucial for LD targeting.

### *Simulations of LiveDrop Suggest Conformational Changes and Energetic Contributions of Specific Sequence Features to LD Targeting*

To gain insights into how specific sequence features of *LiveDrop* determine its LD accumulation we performed molecular dynamics simulations. Since the structure of *LiveDrop* has not been resolved experimentally, we predicted it using the Rosetta *ab initio* structure generation tool (163, 164), and validated it with MEMSAT-SVM (165-167), TOPCONS (168), and RosettaMP (169) as described in Methods. The ten best-scoring structures were similar, showing two  $\alpha$ -helices closely connected by a hinge at P185 resulting in a hairpin. We next computationally inserted and relaxed *LiveDrop* in a bilayer membrane and LD monolayer, each with a composition resembling

the mammalian ER and LDs (170, 171). In the LD system, the neutral lipid layer was composed of 1:1 TG and cholesteryl oleate.

During equilibration there was a shift in the helix-helix separation of *LiveDrop* in the monolayer versus bilayer simulations. To verify this structural change was sufficiently sampled, we used umbrella sampling (132) to calculate the free energy profile (potential of mean force (PMF)) of helix-helix separation (Fig. 4-7A). This revealed that the most energetically stable conformation of *LiveDrop* is more open in the bilayer (~1.7-nm separation) than in the monolayer (~0.7-nm separation) (Fig. 4-3A, 4-7A). As a result, a number of residues changed orientation. W172 relocated from the bilayer midplane to just below the phospholipid glycerol plane in the monolayer (Fig. 4-3A). The positively charged R179 and R187, originally found at the luminal interface of the bilayer, relocated to the monolayer surface and the organic phase, surrounded by water molecules, respectively (Fig. 4-3A). A similar solvation effect was observed for P185, which transitioned from the luminal interface of the bilayer to the organic phase in the monolayer (Fig. 4-3A). These results were further validated in 1–2- $\mu$ s molecular dynamics simulations lacking the influence of sterol esters in the oil phase of the LD system (Fig. 4-7B) (see Methods). In both of these monolayer and bilayer simulations, the depth of the central P185 was stabilized (Fig. 4-3B).

Since it is not computationally feasible to calculate the free energy of membrane insertion for a protein of this size, we estimated the driving force for LD accumulation by calculating the free energy profiles (PMFs) of individual residues inserting into a bilayer membrane or LD monolayer (see Methods for justification). Permeation PMFs for seven amino acids of varying hydrophobicity were calculated using transition-tempered metadynamics (172). The PMFs for phenylalanine, tryptophan, and tyrosine demonstrated greater stability at all depths in the monolayer than in the bilayer (Fig. 4-7C). In the phospholipid tail region (~2-nm membrane depth)

particularly, the normally high energy barrier for bilayer permeation (173, 174) is significantly reduced in the monolayer (Fig. 4-7C) due to the formation of hydrogen bonds between the residues' side chains and the polar ester groups of phospholipids and core TG molecules. This stabilization effect continues into the neutral lipid layer where each of the reported aromatic residues has slightly lower free energy than in bulk water (Fig. 4-7C). Specifically, the imino group of tryptophan and the hydroxyl group of tyrosine are stabilized by forming hydrogen bonds with the glycerol moieties of TG and associated water molecules, suggesting these residues have a preference for the LD environment.

Besides this general trend, there were specific changes in free energy due to the conformational changes of *LiveDrop*. Based on the depth of each residue in the bilayer or monolayer minimum free energy structures (Fig. 4-3A), we calculated their respective insertion free energies from the single amino acid PMFs (Fig. 4-7C) (see Methods). The difference between those for the bilayer and monolayer estimates the change in free energy for *LiveDrop* to accumulate on LDs (Fig. 4-3C). Notably, the hydrophobic residues positioned in the middle of the  $\alpha$ -helices and, therefore, located near the energetically unfavorable bilayer midplane (Fig. 4-7C), have a large preference for the LD monolayer (Fig. 4-3C), where they redistribute to more favorable depths. Two large hydrophobic tryptophan residues (W172 and W197) in particular gained the largest free energy difference by relocating to the monolayer (Fig. 4-3C). This is consistent with amino acid distribution studies, which show that tryptophan and tyrosine residues preferentially occur at the interface regions of membrane proteins (175). Thus, the position of tryptophan (and potentially tyrosine) residues within *LiveDrop* contributes significantly to its accumulation on LDs.



We also addressed whether residue identity is important for the energetic stabilization of *LiveDrop* on LDs by calculating the free energy difference associated with changing F174, F176, F177, and F192, as well as W166, W172, and W197, individually to valine (Fig. 4-3D). Changing phenylalanines to valines in the *LiveDrop* sequence did not considerably alter the free energy differences between monolayer and bilayer (Fig. 4-3D). However, free energy calculations for mutating tryptophans to valines predicted a decrease in LD targeting for W172V and W197V, and an increase in LD targeting for W166V (Fig. 4-3D).

#### *Positively Charged and Tryptophan Residues Are Required for Efficient LD Targeting of LiveDrop in Cells*

To experimentally address the predictions derived from our simulations, we tested the LD targeting role of the positively charged and large hydrophobic residues in *LiveDrop*. Positively charged residues are unusual for a membrane-embedded protein segment, but can localize at the membrane-water interface by interacting with the negatively charged phospholipid headgroups, in a phenomenon known as “snorkeling” (176). We individually mutated each of these residues (i.e., K167, R179, and R187) to alanine and assessed the LD targeting capacity of the mutant variants. R179A resulted in a targeting defect, reducing LD accumulation by ~44% at 3 h and ~47% at 24 h of oleate treatment. In contrast, K167A or R187A had no apparent effect in LD targeting (Fig. 4-4A, 4-4B). Combining R179A with K167A and R187A did not affect ER targeting nor membrane protein topology, but it reduced LD accumulation to a similar extent as R179A alone (by ~40% at 3 h and ~47% at 24 h) (Fig. 4-4C, 4-4D). Thus, R179 appears to be the most relevant positively charged residue in *LiveDrop* for mediating LD accumulation.

Our simulations also predicted that large hydrophobic residues are important for *LiveDrop* accumulation on LDs. Consistently, when we mutated all large hydrophobic residues to small hydrophobic valine residues, the resulting protein variant appeared to insert into the ER membrane but failed to accumulate on LDs. To determine which specific type of the large hydrophobic residues was most crucial, we individually mutated the phenylalanines, tryptophans, or the tyrosine to valines (Fig. 4-4E). Mutating the three tryptophans (W166, W172, and W197) in the *LiveDrop* sequence abolished LD accumulation (Fig. 4-4F, 4-4G) without changing the ER membrane topology. In contrast, mutating either the phenylalanines or the tyrosine did not considerably alter LD targeting (Fig. 4-4F, 4-4G). Mutating each of the *LiveDrop* tryptophan residues individually revealed that W172, which is the conserved tryptophan with a calculated high free energy contribution (see Fig. 4-3C, 4-3D), was not essential for LD accumulation but caused the greatest reduction at early time points of LD induction (~54% at the 3 h time point) (Fig. 4-4H, 4-4I).

Our simulation and experimental results are consistent with the hypothesis that the bilayer anchoring and subsequent reorientation in the monolayer of positively charged residues, R179 and R187, and the differences in the free energies of tryptophans mediate LD accumulation of *LiveDrop*. To test whether these features are sufficient to mediate LD accumulation, we added either type of feature alone or in combination at their original positions to the scrambled-*LiveDrop* sequence (see Fig. 4-2B). The scrambled *LiveDrop* variant harboring both the positively charged and tryptophan residues showed modestly enriched signal on LDs over the ER protein pool when compared with the original scrambled *LiveDrop* variant (~39% increase at the 24 h time point). Thus, repositioning the positively charged and tryptophan residues enables some LD accumulation, yet not to the extent found for wild-type *LiveDrop*, which likely harbors additional determinants mediating LD accumulation.

We were curious to examine the contribution of the hairpin sequence motif to full-length GPAT4 targeting to LDs. Full-length GPAT4 lacking the *LiveDrop* motif failed to insert into the ER membrane but was able to target LDs from the cytosol at both the 3 h and 24 h time points, indicating sequences outside the hairpin motif can contribute to targeting. Mutating the three *LiveDrop* tryptophan residues in the context of full-length GPAT4 reduced but did not abolish LD accumulation. Together, these experiments indicate that targeting of full-length GPAT4 to LDs involves more determinants than those we uncovered for the *LiveDrop* motif alone.

*The Predicted ALG14 Hairpin Motif Is Sufficient to Target LDs and It Relies on Similar LD Targeting Features and Principles as LiveDrop*

To identify other proteins with motifs that are sufficient to target LDs from the ER during early time points of LD biogenesis, we examined the *D. melanogaster* LD proteins (177) with a predicted hydrophobic hairpin motif that are known to localize to the ER. Among these, we selected ALG14, which is an ER protein of the lipid-linked oligosaccharide synthesis pathway (178) and has been reported to be a LD protein in yeast (148) and humans (179). ALG14 is predicted to contain a hydrophobic hairpin motif (amino acids 83–137), consisting of two  $\alpha$ -helical segments separated by a proline residue in position 108 (P108).

We predicted the structure of the ALG14 hairpin with the RosettaMP *ab initio* tool (169). The two best-scoring structures were inserted into a bilayer membrane and LD monolayer for molecular dynamics simulations (see Methods). As shown in representative snapshots (Fig. 4-5A), two positively charged residues (R105 and R107) and one negatively charged residue (D106) localized to the luminal interface of the bilayer membrane and became solvated by water molecules

in the organic phase of the LD system. This hydration stabilizes the charged side chains, since they do not reach the phosphate level of the monolayer. One of the charged residues in particular, R105, dramatically reoriented toward the phospholipids in the monolayer (Fig. 4-5A), resembling the changes observed for R179 in *LiveDrop* (see Fig. 4-3A).

Same as for *LiveDrop*, we estimated which residues contribute to the LD accumulation of the predicted ALG14 hairpin by calculating the free energy changes associated with the different membrane depths of specific hydrophobic residues in the bilayer and monolayer environments. We confirmed that the individual residues had a consistent depth throughout the simulations, including the central P108 residue (Fig. 4-5B). Consistent with *LiveDrop* (see Fig. 4-3C), most of the energy stabilization leading to LD accumulation comes from the hydrophobic residues located in the middle section of each  $\alpha$ -helix (Fig. 4-5C). Specifically, a phenylalanine (F122), tryptophans (W94, W98), and tyrosines (Y101, Y124) are predicted to prefer the monolayer environment (Fig. 4-5C). However, similar residues (i.e., W85, F90, W104, and W129) found at different regions of the predicted ALG14 hairpin, such as the top ends and hinge region, do not contribute to LD accumulation (Fig. 4-5C), indicating that both residue identity and position within a hairpin motif are important for LD accumulation. Also similar to *LiveDrop*, a number of the hydrophobic residues showing an energetic preference for LDs (Fig. 4-5C) are conserved across species.

To test experimentally whether the large hydrophobic and positively charged residues are responsible for the LD accumulation of the predicted ALG14 hairpin, we first generated a sequence variant where the large hydrophobic tryptophans were exchanged for valines (Fig. 4-5D). This reduced LD accumulation by ~26% at 3 h and ~18% at 24 h of oleate treatment (Fig. 4-5E, 4-5F). Likewise, mutating the positively charged residues also reduced LD targeting (~27% at 3 h and ~17% at 24 h) (Fig. 4-5D, 4-5E, 4-5F).

### *The Distribution of Tryptophan and Positively Charged Residues Differentiates LD Targeting from Non-LD Targeting Hairpin Motifs*

To examine whether the features critical for LD accumulation of *LiveDrop* and the predicted ALG14 hairpin operate more broadly, we generated a list of candidates for ER to LD localization based on the *D. melanogaster* LD proteome and examining the literature. We expressed the predicted hairpins of each protein candidate and classified them based on their subcellular localization. We excluded predicted hairpin motifs that failed to insert into the ER (e.g., MCTP, dementin) or appeared to target LDs from the cytosol (e.g., ATGL/*Brummer*). Regarding the motifs that localized to the ER and targeted LDs, most showed low to moderate LD accumulation (e.g., UBXD8, AUP1, AAM-B, NSDHL). However, the predicted hairpin motifs of spastin and LD-associated hydrolase (LDAH) accumulated on LDs efficiently during early time points of LD biogenesis, similar to *LiveDrop*, constituting LD targeting motifs (Fig. 4-6A, 4-6B). Conversely, the predicted hairpin motifs of AGPAT3, DGAT2, FATP, and Selt-like protein appeared to insert into the ER but were not sufficient to target LDs (Fig. 4-6C, 4-6D), defining a set of non-LD targeting motifs.

Analyzing the sequences of the LD targeting and non-LD targeting motifs, we found no simple correlation between LD accumulation and the number of positively charged, large hydrophobic, or tryptophan residues. We, thus, determined a range of properties for these sequences, including molecular weight, hydrophobicity, net charge, number of charged residues, number of large hydrophobic residues, and residue distribution variables. Principal component analysis of these parameters with respect to LD accumulation yielded two major principal components that explain 30% and 21% of the overall variance between the eight motifs (Fig. 4-

6E). Strikingly, the motifs cluster in groups corresponding to the experimentally determined LD targeting and non-LD targeting groups (Fig. 4-6E). Although no single variable can differentiate the two groups, the overall number of positively charged residues and specific residue distribution variables, such the number of positively charged residues at the hairpin hinge region and the number of tryptophan residues in the middle section of the first hairpin  $\alpha$ -helix, make the strongest contributions to the two major principal components. Accordingly, tryptophan residues are found in different hairpin regions of the LD targeting motifs, possibly enriched in the first  $\alpha$ -helix of these predicted hairpins (Fig. 4-6F), whereas in the non-LD targeting ones, they are only found at the top ends and hinge region (Fig. 4-6G). Similarly, positively charged residues were consistently found at the hinge region of the LD targeting motifs (Fig. 4-6F), but they were absent from that region in all the non-LD targeting motifs (Fig. 4-6G).

## DISCUSSION

A number of proteins are originally inserted, via membrane-embedded motifs, in the ER and can subsequently relocate to the surfaces of LDs. This can occur during LD formation for some proteins, or at later time points for others, possibly via ER-LD membrane bridges (65). What drives proteins from the ER bilayer membrane to target to and accumulate at the LD monolayer is unknown.

Based on our data for the hydrophobic, membrane-embedded motifs of GPAT4 and ALG14, a model emerges. The pathway commences with the initial insertion of these motifs into the ER membrane. How *LiveDrop* or the hydrophobic motif of ALG14 insert into the ER membrane is unknown, but candidate pathways for this include the canonical translocon-mediated

pathway (180) and the Pex19/Pex3 pathway posited for other membrane-embedded LD proteins (181). The latter is apparently used for the ER insertion of UBXD8, a protein that targets to LDs from the ER, similar to GPAT4 (181). In the ER, these hydrophobic sequences likely adopt a hairpin conformation in which positively charged residues in the hinge region (such as *LiveDrop*'s R179) interact with the phospholipid headgroups in the luminal side of the ER membrane. This anchors the protein there and, thus, forces large hydrophobic residues into a less favorable environment within the phospholipid tail region of the membrane.

As LDs begin to form in the ER, these hydrophobic hairpins migrate onto the monolayer surface of LDs that are continuous with the ER bilayer membrane. Here, bilayer membrane anchoring is lost, and key residues are able to move to more energetically favorable positions in the environment of the monolayer-bound LD. In particular, large hydrophobic residues, such as tryptophans and sometimes tyrosines, achieve a lower free energy state in the oil phase of the LD than in the midplane of the bilayer membrane. Our simulations suggest that this is due to hydrogen bonding between specific hydrophilic groups in these amino acids (e.g., the imino group of tryptophans or the hydroxyl group of tyrosines) and the headgroups of phospholipids, the glycerol moieties of TG, or coordinating water molecules. Since mutating individual tryptophans of *LiveDrop* had little effect on LD targeting compared with the mutation of all three, the sum of free energies gained from multiple residues on LDs appears to be crucial for the preference of protein motifs for the LD surface. However, since the rescue of LD targeting by reintroducing the positively charged and hydrophobic tryptophan residues was modest, additional sequence features must contribute to *LiveDrop* accumulation on LDs. These could be, for instance, the sequence context around each key residue, necessary to position them in specific geometries, or the sum of minor free energy differences between the bilayer and monolayer for other residues.

Statistical analysis of biophysical and sequence-derived properties allowed us to classify the predicted hairpins of other LD proteins correctly. The main feature correlating with LD accumulation was the presence of positively charged residues around the hairpins' hinge region. In our *LiveDrop* simulations, the positively charged residues (R179 and R187) relocated from the luminal leaflet of the ER bilayer to the LD monolayer surface and to the hydrated glycerol moieties of the neutral lipids in the LD, respectively. Similarly, the positively charged residue R105 of the predicted ALG14 hairpin relocated from the luminal leaflet of the bilayer towards the monolayer surface, suggesting that this may be a general feature of the membrane-embedded motifs that move from the ER to the LD surface.

Among the hydrophobic ER-to-LD targeting motifs we analyzed, we found that spastin is an exception, since it does not contain tryptophans in its predicted membrane-embedded domain. However, this sequence is overall quite different from either *LiveDrop* or the predicted ALG14 hairpin, inasmuch as it is shorter and it does not contain a central proline. Thus, its conformation may differ from a hairpin and other LD targeting determinants may apply.

From our studies, a mechanism emerges for how ER membrane proteins target and accumulate on the monolayer surface of LDs. Specifically, membrane-embedded motifs of these proteins appear to recognize and conformationally adapt to the unique properties of the LD oil-water phase boundary. Future work identifying more sequences that are sufficient to mediate ER-to-LD targeting and accumulation will uncover additional targeting determinants and enable the refinement of this model.



## METHODS

### *Structure Generation*

The LiveDrop structure was generated using the Rosetta ab initio structure generation tool (163, 164, 182). Fragment files were created using the Robetta server (183, 184). Fifty thousand structures were generated, and the ten structures with the best scores were chosen for analysis. These ten structures contained the same main structural elements (i.e., a kink at P185 and two transmembrane  $\alpha$ -helices on either side of P185). The top-scoring structure was used for all simulations. We also predicted the structure with MEMSAT-SVM (165-167), TOPCONS (168), and the RosettaMP ab initio structure generation tool (169, 185-189), which gave structures with similar helicity and features as those from Rosetta. The best Rosetta generated structure was then solvated in a cubic water box and relaxed with 1 ns of NVT equilibration and 20 ns of NPT equilibration.

The ALG14 hairpin structure was generated using the RosettaMP ab initio structure generation tool. Fragment files were created using the Robetta server, and the membrane topology was predicted using OCTOPUS (190). More than ten thousand structures were generated and the two structures with the best scores were used. A low-resolution structure was refined to atomic level detail using high-resolution refinement as described (169). The three principal axes of the ALG14 hairpin structure were aligned with the x, y, and z directions before inserting into either a bilayer or monolayer system.

### *Molecular Dynamics Simulations*

Molecular dynamics simulations of LiveDrop (1  $\mu\text{s}$  in the bilayer and 2  $\mu\text{s}$  in the monolayer) and the predicted ALG14 hairpin (2  $\mu\text{s}$ ), in both a bilayer and monolayer system, were conducted. In the case of the LiveDrop simulations, the initial protein structure for the final production run was taken from the corresponding umbrella sampling minimum representation (see below). For the ALG14 hairpin simulations, the initial protein structures were predicted using RosettaMP as described above. The CHARMM-GUI membrane builder (108) was used to insert the protein motifs into a bilayer membrane. The bilayer system was composed of 133–137 POPC, 55 DOPE, and 16 SAPI molecules per leaflet in TIP3P water (115) and 0.15 M NaCl solution. A detailed description of the lipid composition used in each simulation is shown in Table 4-1. The monolayer system was generated by including an 8-nm-thick TG layer between the two leaflets of the original bilayer, so that both systems have the same lipid composition. To build a monolayer system, three segments were prepared separately and later combined using VMD (191). For the first segment, a bilayer system, including a given protein motif, was built using the CHARMM-GUI builder. The protein, the upper membrane leaflet, and the water and ions above the bilayer midplane were selected and used as the upper component in a final monolayer system. For the second segment, a separate bilayer membrane was built using the CHARMM-GUI builder. The corresponding lower leaflet and water and ions below the bilayer midplane were selected and used as the lower component in a final monolayer system. The third segment, corresponding to the neutral lipid layer, was built using Packmol (114). All three segments were further constructed such that they had the same X and Y dimensions. The three segments were then combined with extra 1-nm spacing along the Z axis between the TG segment and the other membrane segments. Any TG molecules that had bad contacts with the membrane components were removed from the system. Additional ions were added, when needed, to make the system neutral.

All-atom simulations were performed using the GROMACS (version 2016 and 2018) simulation engine (117) with the CHARMM36 lipid and protein force fields (112, 118, 192). Simulations were integrated with a 2-fs timestep. The Particle Mesh Ewald algorithm (119) was used to evaluate long-range electrostatic interactions with a real space cutoff of 1.0 nm. Lennard-Jones interactions were cut-off at 1.0 nm with the potential-shift-Verlet method, and the neighbor list was updated every 100 steps. Long-range dispersion was corrected for energy and pressure. The pressure was maintained semi-isotropically using the Parrinello-Rahman barostat (120) at a pressure of 1.0 bar, a compressibility of  $4.5 \times 10^{-5}$ /bar, and a coupling time constant of 2.0 ps. Bonds to hydrogen were constrained using the LINCS algorithm (121). The temperature was maintained at 310 K using the stochastic velocity rescaling thermostat (122) with a coupling time constant of 0.1 ps. Biased simulations were also conducted in GROMACS, with the addition of the PLUMED2 plugin (135). For biased simulations, non-bonded van der Waals interactions were cut-off at a distance of 1.2 nm between atoms and were switched to zero in the case of distances between 1.0 and 1.2 nm. The Particle Mesh Ewald algorithm was used with a real space cutoff of 1.2 nm. The pressure was maintained semi-isotropically with a coupling time constant of 5.0 ps.

### *LiveDrop Umbrella Sampling Simulations*

To confirm that the conformational changes in the LiveDrop  $\alpha$ -helices converged, umbrella sampling simulations (132) were run on LiveDrop. The protein was first inserted into the bilayer and monolayer systems using the procedure described by Javanainen (193). Briefly, the protein was placed next to the membrane patch, with the water and ions removed from the system. Restraints were placed on the protein backbone, as well as on the phosphorous atoms of the lipids. A lateral pressure of 1000 bar was applied to the system, pushing the protein into the membrane.

Three lipids were removed from the cytosolic membrane leaflet to maintain the area of both leaflets equal. The systems were then relaxed for 2 ns of NPT equilibration. TIP3P water and 0.15 M NaCl were added to both systems. Finally, the membrane-protein systems were equilibrated according to the CHARMM-GUI membrane equilibration procedure (56, 107-109). The bilayer membrane comprised 258 POPC, 107 DOPE, and 32 SAPI molecules (see Table 4-1). The monolayer system comprised 218 POPC, 91 DOPE, 26 SAPI, 91 TG, and 92 cholesteryl oleate molecules in total (see Table 4-1).

Umbrella sampling simulations were run biasing the distance between the center of geometry of the LiveDrop residues W166 and A203. Harmonic restraints with a force constant of  $10^4$  kJ/mol/nm<sup>2</sup> were placed on each of the ~80 umbrella sampling windows, which were spaced every 0.05 nm over a distance range of 0.5–4.5 nm. Each window was run for 50 ns in the bilayer and for 40 ns in the monolayer. The first 10 ns were considered equilibration and they were not used for calculating the potential of mean force (PMF). The starting structures for each window were generated by conducting metadynamics (194) on the same collective variable as the umbrella sampling, for both the bilayer and monolayer. Hills were deposited with a height of 0.003 kJ/mol/nm every 1 ps. The simulation length of each of the four metadynamics runs was at least 1  $\mu$ s. Each of the starting structures for the umbrella sampling was then chosen from the metadynamics runs, where the distance between residues W166 and A203 was within 0.02 nm of the equilibrium window distance. The PMF was then calculated using the Weighted Histogram Analysis Method (WHAM) package (195-197) with a bin spacing of 0.01 nm. The error in the simulations was estimated by dividing the equilibrated trajectories into four blocks, calculating the PMF for each block, and then determining the standard deviation of the block PMFs.

### *Single Amino Acid Permeation*

The PMFs for single amino acid permeating through a bilayer and monolayer were conducted using Transition-Tempered Metadynamics (172, 198), biasing the z-component of the position vector connecting the center of mass of the membrane and the center of mass of the amino acid. The bilayer system comprised 16 POPC, 7 DOPE, and 2 SAPI molecules per leaflet in TIP3P water and 0.15 M NaCl solution. The monolayer system included 35 TG molecules between the two bilayer leaflets. Acetylated N-terminus (ACE) and methylated C-terminus (CT1) patches were applied to the termini of all amino acids evaluated. For each amino acid, four independent simulation runs, each 2  $\mu$ s long, were conducted in each membrane system. The final PMF was obtained by averaging the PMFs obtained from the four simulations. The Gaussian function was deposited every 2 ps with a height of 0.001 kJ/mol and a scaled width of 0.025 for the bilayer and 0.018 for the monolayer. The bias factor was set to 10. The two transition basins were located at -0.2 and 0.2 for the bilayer and at -0.3 and 0.3 for the monolayer. The latter distance values are based on a scaled distance for biasing simulations, where the z-dimension of a system was considered equal to 1 and centered at 0. Simulations were conducted in the canonical ensemble (NVT). Single amino acid permeation calculations were run for most of the amino acids bearing hydrophobic side chains, including alanine, valine, isoleucine, leucine, phenylalanine, tyrosine, and tryptophan.

### *Free Energy Calculations*

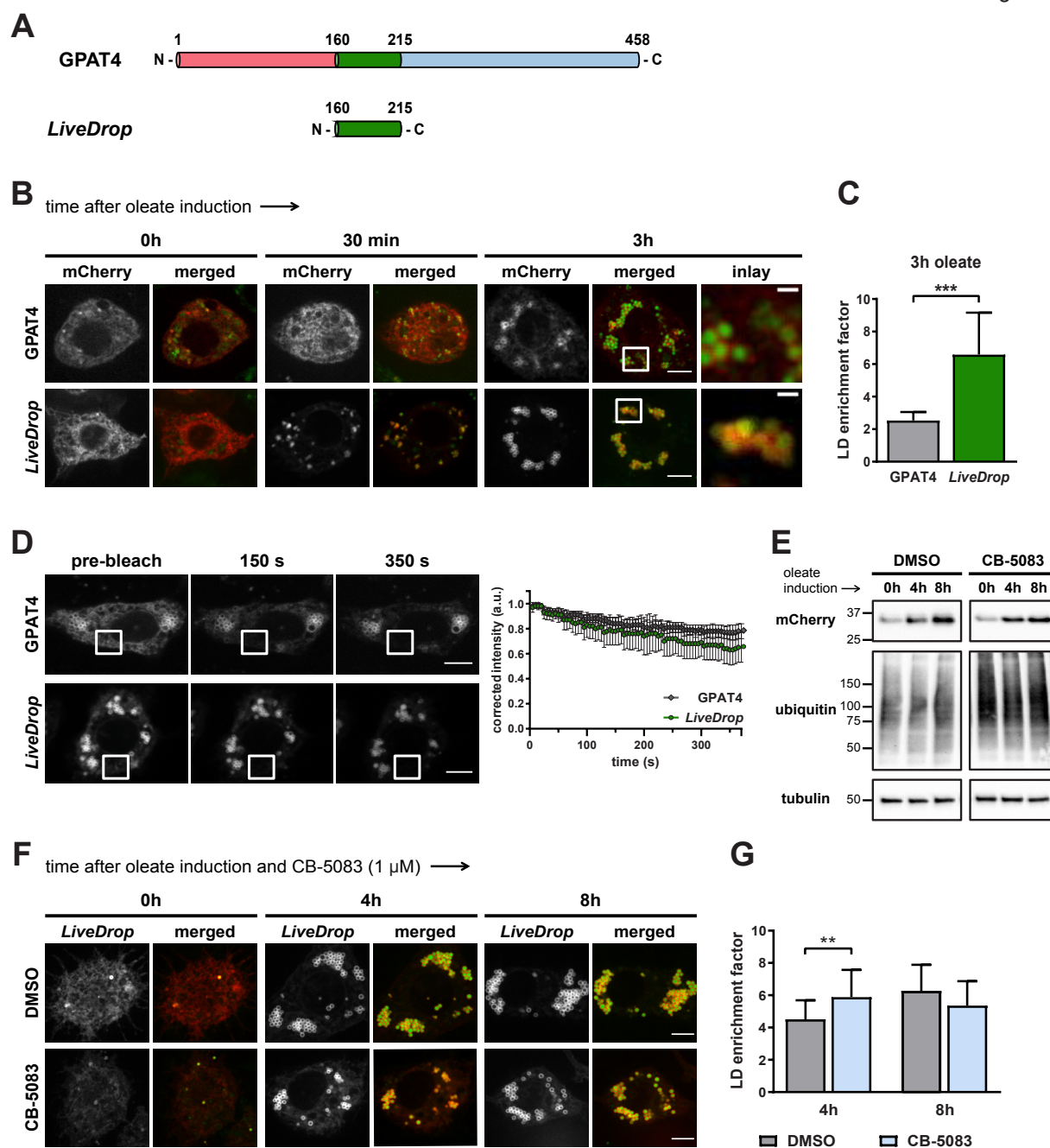
The thermodynamic force driving protein accumulation on LDs was approximated based on the changes in free energy for amino acids transitioning from a bilayer to a monolayer. These changes

in free energy are based on the bilayer and monolayer structures combined with the previously described permeation free energy profiles for single hydrophobic amino acids. To estimate these changes in free energy, the following process was used: (i) The depth of each residue was calculated as the z component of the position vector connecting the center of mass of the upper phospholipid phosphorus and the center of mass of a given residue. For LiveDrop, the residue depths were calculated using the minimum energy structures obtained from the umbrella sampling simulations, and for the predicted ALG14 hairpin, residue depths were calculated by time-averaging all the corresponding simulations from 500 ns to 2  $\mu$ s. (ii) The permeation PMF for a given residue was shifted so that the free energy is set to zero in bulk water, and the depth of the upper phosphorus atoms are set to zero to correspond with the structural depth. (iii) Based on the permeation PMF profiles, the corresponding free energy of each hydrophobic residue in the protein motifs was estimated at its calculated depth. Comparing the free energy of each residue in the bilayer and monolayer estimates its change in free energy when transitioning from the ER to LDs. Given the large changes in stability as a function of bilayer or monolayer depth, the changes in position are expected to dominate the change in free energy for a full protein to relocate from the ER bilayer to the LD monolayer. This is further justified by the thermodynamic cycle shown in Figure 4-8, which depicts the minimal change in association energies for amino acids in a bilayer versus a monolayer environment, compared to the large changes based on position and increased stability in the monolayer relative to the bilayer midplane (shown in Figure 4-7C).

## **ACKNOWLEDGEMENTS**

We thank Scot Stone (U. of Saskatchewan) for advice on the protease protection assays; Coline Prévost for help in image and movie quantification; Xihao Li (Harvard Chan School of

Public Health) and Wenting Lyu for help with statistical analyses; Chenghan Li, Jesper Madsen, and Zack Jarin (Voth group) for help with simulations; Julia Koehler Lemman (Flatiron Institute, Simons Foundation) for help with predicting membrane protein structures; members of the Farese & Walther laboratory, Stephen White (UC Irvine), and Gregory Voth (U. of Chicago) for useful discussions; and Gary Howard for editorial assistance. This research was supported by the National Institute of General Medical Sciences of the NIH through 5R01GM097194 (to TCW) and 5R01GM124348 (to RVF). TCW is an investigator of the Howard Hughes Medical Institute. The simulations used the Extreme Science and Engineering Discovery Environment (XSEDE), supported by NSF grant ACI-1548562. Specifically, it used the Stampede2 system at the Texas Advanced Computing Center (TACC) and the Bridges system at the Pittsburgh Supercomputing Center (PSC) through allocation number TG-MCA94P017. This work was also completed in part with resources provided by the U. of Chicago Research Computing Center.



**Figure 4-1** The Hydrophobic, Membrane-Embedded Motif *LiveDrop* Accumulates on Nascent LDs.

(A) Schematic representation of *D. melanogaster* GPAT4 (1–458 aa), including the *LiveDrop* motif (160–215 aa, green).



**Figure 4-1, continued.**

(B) *LiveDrop*, but not full-length GPAT4, targets LDs during their biogenesis. *Drosophila* S2 cells transfected with mCherry-tagged GPAT4 or mCherry-tagged *LiveDrop* (red) were imaged after incubation with oleate. LDs were stained with boron-dipyrromethene (BODIPY, green). Scale bar, 5  $\mu\text{m}$  (merged), 1  $\mu\text{m}$  (inlay).

(C) Mean values + SD ( $n > 15$ ) of the protein signal on LDs after 3 h of oleate treatment. \*\*\* $p < 0.001$ .

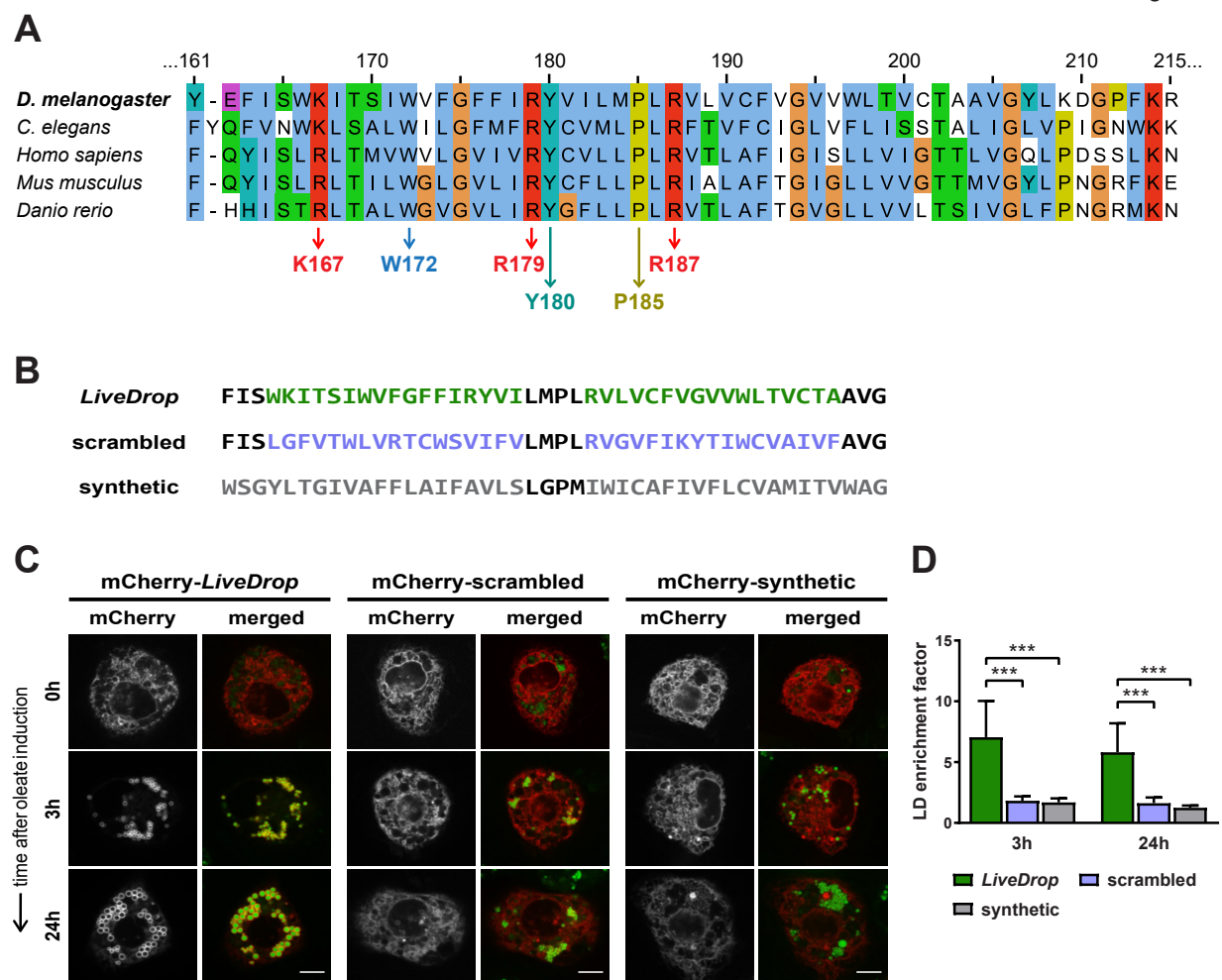
(D) (left) *LiveDrop* and full-length GPAT4 accumulate on LDs but do not move back to the ER. Repeatedly bleaching the mCherry signal in the indicated regions did not reduce the LD signal over time. Scale bar, 5  $\mu\text{m}$ . (right) Mean values  $\pm$  SD of the normalized fluorescence intensity on LDs for both mCherry-GPAT4 ( $n = 7$ ) and mCherry-*LiveDrop* ( $n = 9$ ) after 5 and 3 h of oleate treatment, respectively. a.u., arbitrary units.

(E) Inhibiting VCP/p97 with 1  $\mu\text{M}$  CB-5083 leads to accumulation of ubiquitylated proteins in S2 cells expressing mCherry-*LiveDrop*. The samples shown were ran on the same SDS-PAGE gel.

(F) Inhibiting VCP/p97 with CB-5083 does not significantly change the enrichment of *LiveDrop* (red) on LDs. LDs were stained with BODIPY (green). Scale bar, 5  $\mu\text{m}$ .

(G) Mean values + SD ( $n > 23$ ) of the protein signal on LDs after 4 and 8 h of oleate treatment. \*\* $p < 0.01$ .

Figure 2



**Figure 4-2** LD Accumulation of *LiveDrop* Requires Specific Sequence Features

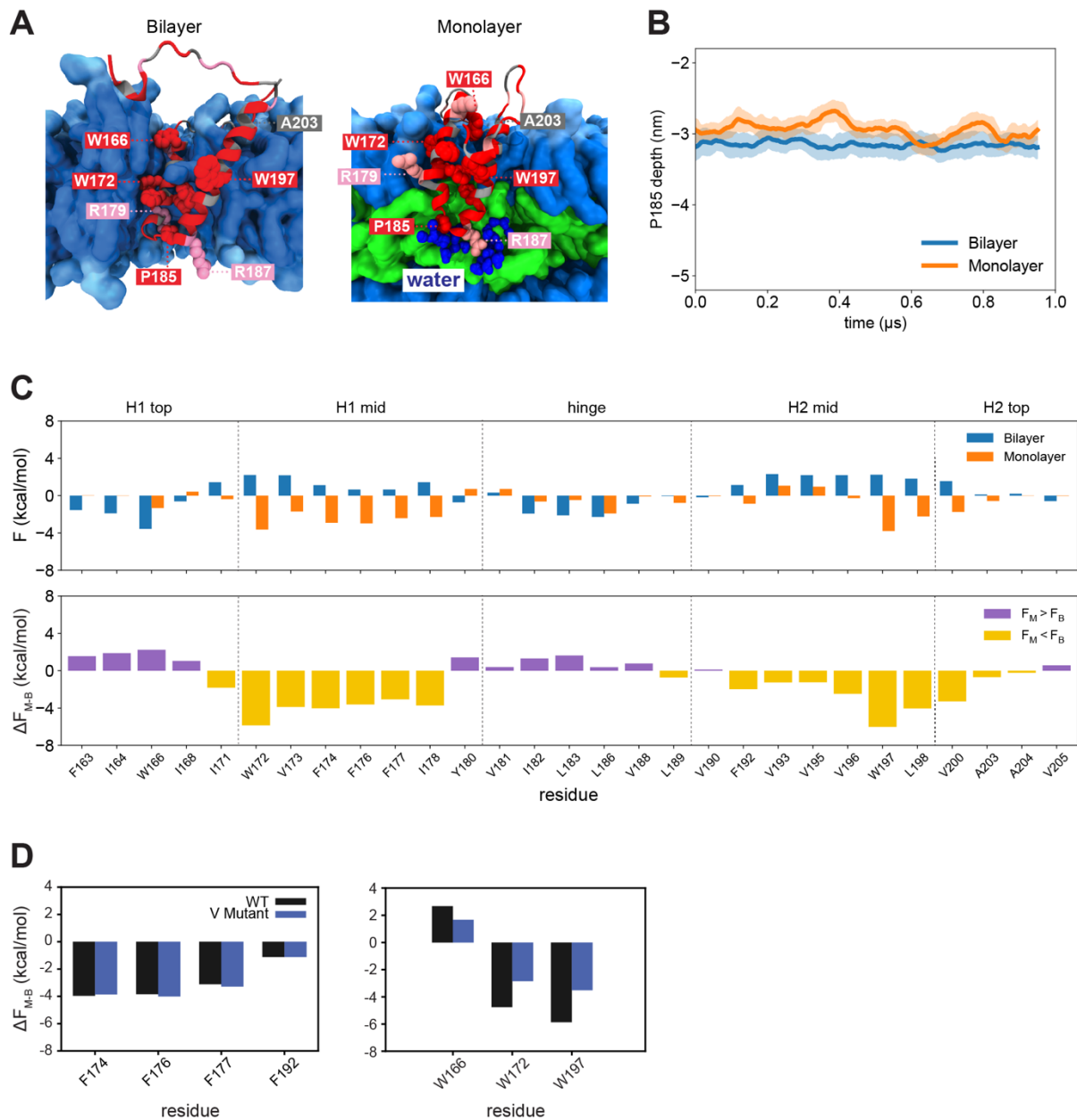
(A) Sequence alignment of the *D. melanogaster LiveDrop* motif (160–215 aa) and other GPAT4 ortholog sequences from representative species. Conserved residues are highlighted, including P185 (dark yellow), positively charged residues (K167, R179, R187; red), and large hydrophobic residues (W172, Y180; sky blue or teal).

(B) Amino acid sequences of *LiveDrop* (green), a scrambled *LiveDrop* variant (violet), and a synthetic hydrophobic  $\alpha$ -helical motif (gray).

(C) The scrambled *LiveDrop* variant and the synthetic hydrophobic motif do not target LDs. S2 cells transfected with mCherry-tagged versions of the protein motifs in (B) (red) were incubated with oleate throughout the indicated time points and imaged by confocal microscopy. LDs were stained with BODIPY (green). Scale bar, 5  $\mu$ m.

(D) Mean values + SD ( $n > 12$ ) of the protein signal on LDs after 3 and 24 h of oleate treatment. \*\*\* $p < 0.001$ .

Figure 3



**Figure 4-3** Simulations of *LiveDrop* Suggest Conformational Changes and Energetic Contributions of Specific Sequence Features to LD Targeting

(A) Equilibrated simulation structures of *LiveDrop* in a bilayer (left) and monolayer (right) highlighting key residues: R179 and R187 (pink); W166, W172, W197, and P185 (red); and A203 (gray). The phospholipid headgroups and tails are shown in light and medium blue, respectively, the neutral lipid phase in green, and water molecules in dark blue.

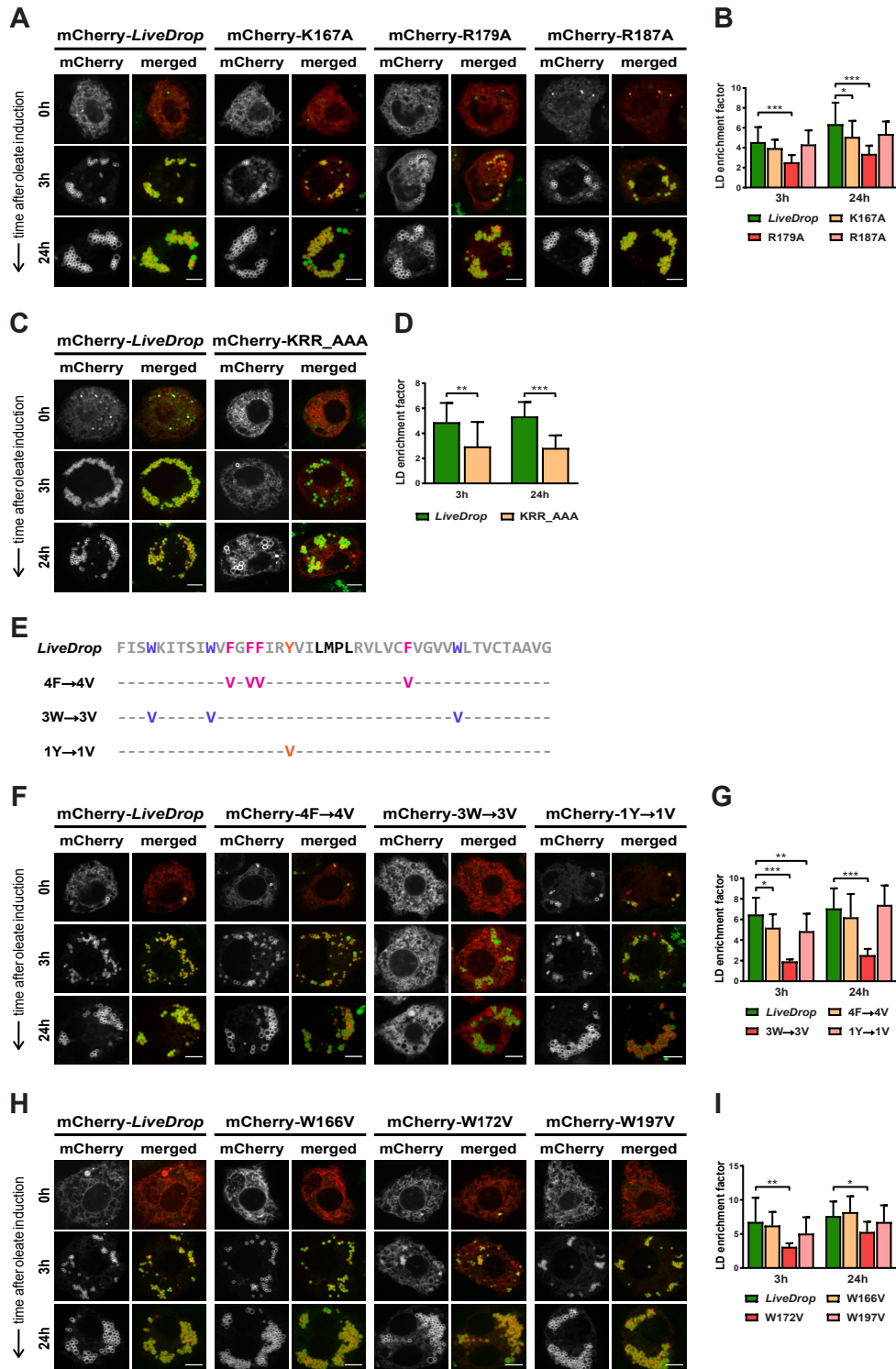
**Figure 4-3, continued.**

(B) P185 depth with respect to the upper phosphate plane of the bilayer (blue) and monolayer (orange). The solid lines show the running averages and the filled areas show the standard deviations from four and two simulations of the bilayer and monolayer, respectively. Extending the simulations in the monolayer to 2  $\mu$ s (not shown) did not reveal notable changes in the depth or conformation during the last 1  $\mu$ s.

(C) (upper panel) Free energy of insertion of *LiveDrop* hydrophobic residues into the bilayer (blue) and monolayer (orange). The defined hairpin regions were residues position ( $\alpha$ -helix top [H1 top, H2 top], middle [H1 mid, H2 mid], or hinge) are shown. (lower panel) Free energy differences between the monolayer and bilayer ( $\Delta F_{M-B}$ ). Bars are colored yellow or violet if the residue is more stable in the monolayer ( $F_M < F_B$ ) or in the bilayer ( $F_M > F_B$ ), respectively.

(D) Free energy differences between the monolayer and bilayer ( $\Delta F_{M-B}$ ) for the phenylalanine (left panel, black) and tryptophan (right panel, black) residues, and for their corresponding valine mutants (blue).

Figure 4



**Figure 4-4** Positively Charged Residues and Tryptophan Residues Are Required for Efficient LD Targeting of *LiveDrop* in Cells

**Figure 4-4, continued.**

(A) Of the three *LiveDrop* variants with single positive charge mutations (K167A, R179A, R187A), only R179A causes a reduction in LD accumulation.

(B) Mean values + SD (n > 18) of the protein signal on LDs after 3 and 24 h of oleate treatment. \*p < 0.05; \*\*\*p < 0.001.

(C) A *LiveDrop* variant with the three positively charged residues (K167, R179, R187) exchanged for alanines (KRR\_AAA) is compromised in LD accumulation to a similar extent as the single R179A mutation.

(D) Mean values + SD (n > 10) of the protein signal on LDs after 3 and 24 h of oleate treatment. \*\*p < 0.01; \*\*\*p < 0.001.

(E) Amino acid sequence of *LiveDrop* variants in which the phenylalanines (4F→4V, magenta), tryptophans (3W→3V, purple), and tyrosine (1Y→1V, orange) are individually mutated to valines. The predicted hinge of the *LiveDrop* sequence (gray) is shown in black. For each *LiveDrop* variant, the amino acid positions indicated with a hyphen (-) remain the same as in the original sequence.

(F) The *LiveDrop* variant with mutated tryptophans (3W→3V) does not accumulate on LDs, but the variants with mutated phenylalanines (4F→4V) and tyrosine (1Y→1V) target LDs in a manner comparable to the wild-type sequence.

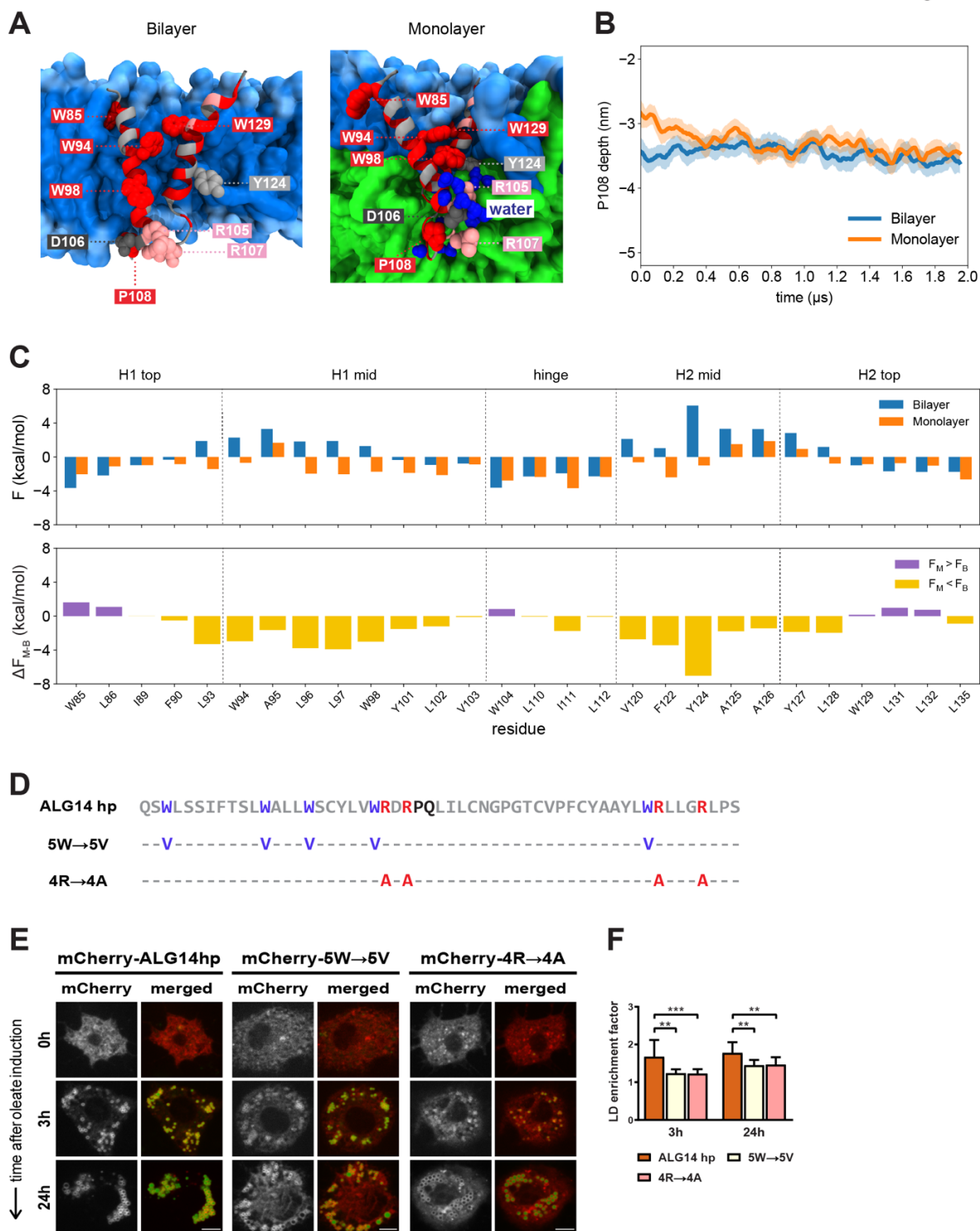
(G) Mean values + SD (n > 10) of the protein signal on LDs after 3 and 24 h of oleate treatment. \*p < 0.05; \*\*p < 0.01; \*\*\*p < 0.001.

(H) *LiveDrop* variants with single tryptophan mutations (W166V, W187V) target LDs similar to the wild-type sequence, except for W172V, which shows reduced LD accumulation.

(I) Mean values + SD (n > 10) of the protein signal on LDs after 3 and 24 h of oleate treatment. \*p < 0.05; \*\*p < 0.01.

For (A), (C), (F), and (H), S2 cells were transfected with mCherry-tagged versions of the *LiveDrop* variants (red), incubated with oleate throughout the indicated time points, and imaged by confocal microscopy. LDs were stained with BODIPY (green). Scale bar, 5 μm.

Figure 5



**Figure 4-5** The ALG14 Hairpin Is Sufficient to Target LDs and It Relies on Similar LD Targeting Features as *LiveDrop*

**Figure 4-5, continued.**

(A) Representative structures of the ALG14 hairpin in a bilayer (left) and monolayer (right) highlighting key residues: R105 and R107 (pink); D106 (dark gray); W85, W94, W98, W129, and P108 (red); and Y124 (light gray). The phospholipid headgroups and tails are shown in light and medium blue, respectively, the neutral lipid phase in green, and water molecules in dark blue.

(B) P108 depth with respect to the upper phosphate plane of the bilayer (blue) and monolayer (orange). The solid lines show the running averages and the filled areas show the standard deviations from four simulations of both the bilayer and monolayer, respectively.

(C) Same as Figure 4-3C for the ALG14 hairpin.

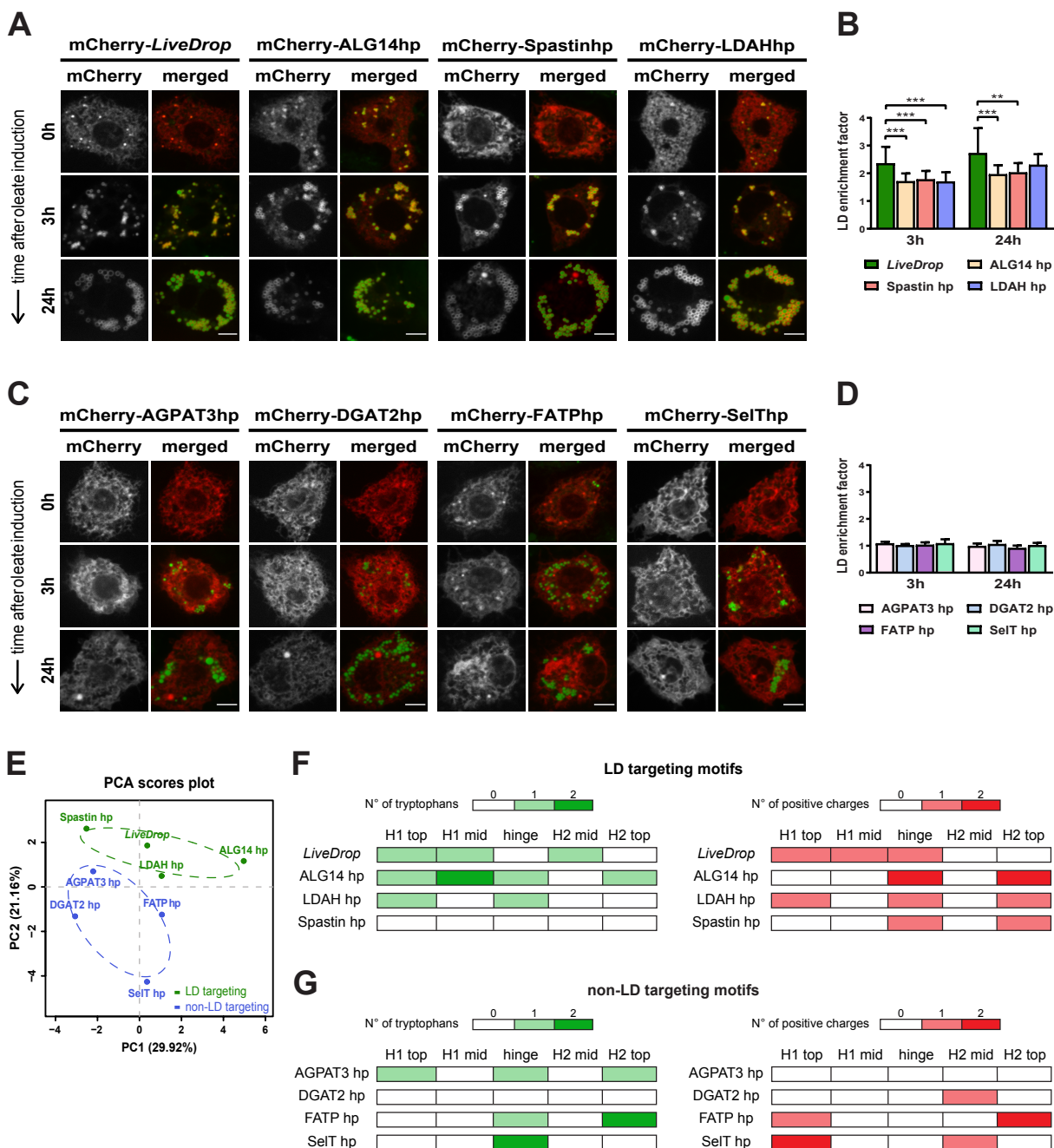
(D) Amino acid sequence of the ALG14 hairpin (ALG14 hp, gray) and its sequence variants in which the tryptophan (5W→5V, purple) and positively charged (4R→4A, red) residues are individually mutated to valines and alanines, respectively. The predicted hinge of the ALG14 hairpin sequence is shown in black. Amino acid positions indicated with a hyphen (–) remain the same as in the original sequence.

(E) Both of the ALG14 hairpin sequence variants with mutated tryptophans (5W→5V) or positively charged arginines (4R→4A) are compromised in LD accumulation. S2 cells transfected with mCherry-tagged versions of the ALG14 hairpin and its variants (red) were incubated with oleate throughout the indicated time points and imaged by confocal microscopy. LDs were stained with BODIPY (green). Scale bar, 5  $\mu$ m.

(F) Mean values + SD ( $n > 13$ ) of the protein signal on LDs after 3 and 24 h of oleate treatment. \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .



Figure 6



**Figure 4-6** The Distribution of Tryptophan and Positively Charged Residues Differentiates LD Targeting from Non-LD Targeting Hairpin Motifs

(A) The predicted hairpin motifs of the LD proteins spastin (spastin hp) and LDAH (LDAH hp) target the ER and efficiently accumulate on LDs.

**Figure 4-6, continued.**

(B) Mean values + SD ( $n > 19$ ) of the protein signal on LDs after 3 and 24 h of oleate treatment. \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

(C) The predicted hairpin motifs of the LD proteins AGPAT3 (AGPAT3 hp), DGAT2 (DGAT2 hp), FATP (FATP hp), and SelT-like protein (SelT hp) target the ER, but they are not sufficient to target LDs.

(D) Mean values + SD ( $n > 10$ ) of the protein signal on LDs after 3 and 24 h of oleate treatment.

For (A) and (C), S2 cells were transfected with mCherry-tagged versions of the predicted hairpin motifs (red), incubated with oleate throughout the indicated time points, and imaged by confocal microscopy. LDs were stained with BODIPY (green). Scale bar, 5  $\mu\text{m}$ .

(E) Scores plot from the principal component analysis (PCA) of the parameters derived from the predicted hairpin sequences previously classified as LD targeting and non-LD targeting motifs. The first two principal components (PC1 and PC2) are plotted against each other. PC1 accounts for 29.29% of the variance, and PC2 accounts for 21.16%.

(F) and (G) Distribution of the tryptophan (left, green scale) and positively charged residues (right, red scale) throughout defined regions of the indicated hairpin motifs.

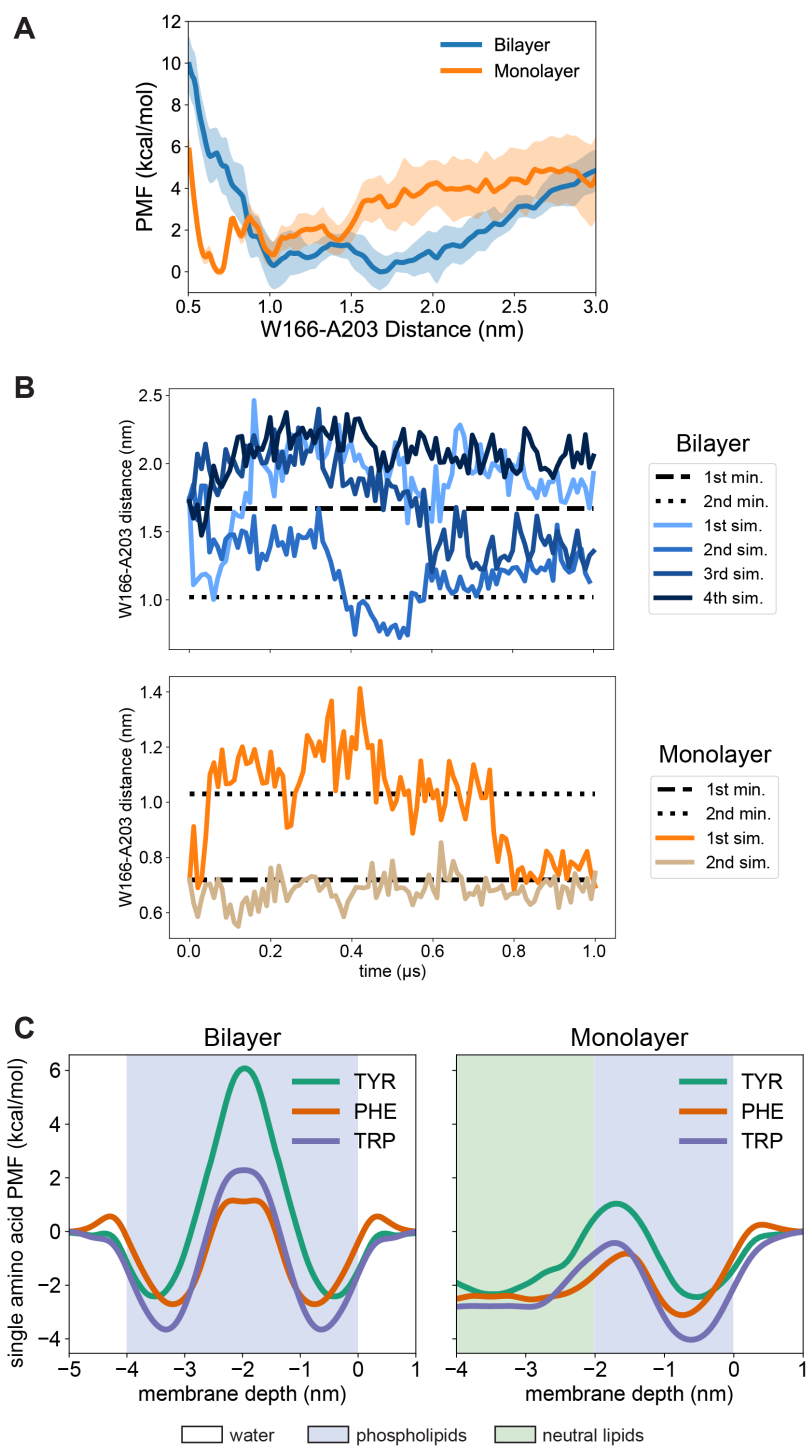


Figure S4. Related to Figure 3.

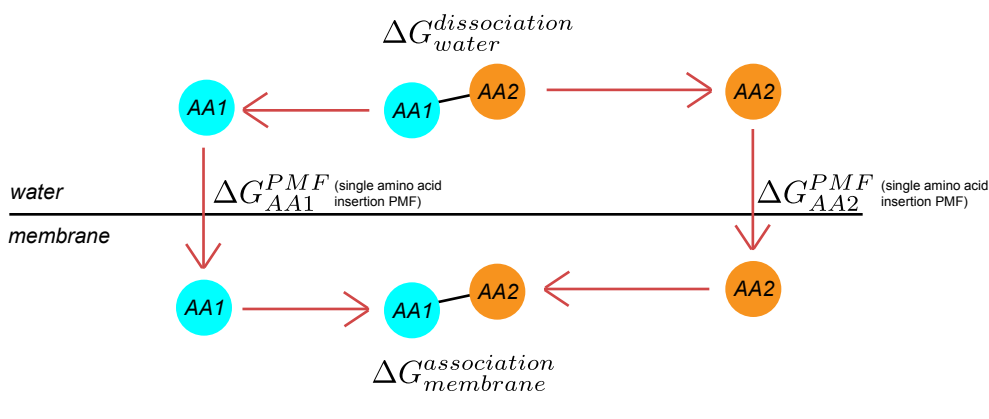
**Figure 4-7** Umbrella sampling simulations of *LiveDrop* and free energy profiles of single amino acids

**Figure 4-7, continued.**

(A) Potential of mean force (PMF) for the association of the *LiveDrop*  $\alpha$  helices. The distance between residues W166 and A203 (W166-A203 distance), both located close to the top ends of each helix, was used to approximate the degree of proximity between the  $\alpha$  helices of *LiveDrop*. The PMF was calculated as a function of this distance, and the error was estimated using the block averaging method, which entails dividing the equilibrated trajectories into four blocks. The  $\alpha$  helices of *LiveDrop* are closer together in the monolayer (orange), with a minimum in the PMF at 0.7 nm. Conversely, in the bilayer (blue), they are more separated.

(B) Distance between residues W166 and A203 throughout the molecular dynamics simulations (final production run) of *LiveDrop* in the bilayer (upper panel) and monolayer (lower panel). The initial protein structure used in each membrane system was taken from the corresponding umbrella sampling (US) minimum representation. The dashed line is the distance corresponding to the energy minimum in the US calculation, and the dotted line is the distance corresponding to the second energy minimum. As observed in the US result (see Fig. 4-7A), no significant barrier separates the two most stable states in the bilayer. Thus, the bilayer simulations explore the region broadly where the free energy is not too high. On the other hand, the two most stable states in the monolayer are separated by a 1–2 kcal/mol barrier, and the monolayer simulations mostly sample the first and second stable states.

(C) PMF for the permeation of single phenylalanine (orange), tryptophan (violet), and tyrosine (dark green) residues into the bilayer and monolayer. These free energy profiles, in combination with each residue's specific position (membrane depth) in the dominant bilayer and monolayer conformations, were used to estimate the change in free energy for each residue between the bilayer and monolayer environments. The error (not shown here) was estimated by taking the standard deviation from four independent simulations, resulting in values which were within 1 kcal/mol and 2 kcal/mol for the bilayer and monolayer, respectively.



$$\Delta G(BI) = \Delta G_{water}^{dissociation}(BI) + \Delta G_{AA1}^{PMF}(BI) + \Delta G_{AA2}^{PMF}(BI) + \Delta G_{membrane}^{association}(BI)$$

$$\Delta G(LD) = \Delta G_{water}^{dissociation}(LD) + \Delta G_{AA1}^{PMF}(LD) + \Delta G_{AA2}^{PMF}(LD) + \Delta G_{membrane}^{association}(LD)$$

$$\text{if } \Delta G_{membrane}^{association}(BI) = \Delta G_{membrane}^{association}(LD),$$

$$\Delta G(BI) - \Delta G(LD) = (\Delta G_{AA1}^{PMF}(BI) - \Delta G_{AA1}^{PMF}(LD)) + (\Delta G_{AA2}^{PMF}(BI) - \Delta G_{AA2}^{PMF}(LD))$$

**Figure 4-8** Thermodynamic cycle for the membrane insertion of a single amino acid

Diagram describing the insertion free energy of a peptide ( $\Delta G$ ) in terms of the insertion free energy of a single amino acid ( $\Delta G_{AA}^{PMF}$ ), similar to the permeation free energy profiles calculated in Figure 4-7C. In this thermodynamic cycle, a peptide, consisting of two amino acids (AA1–AA2) but extendable to a longer peptide, dissociates in water ( $\Delta G_{water}^{dissociation}$ ). Each single residue (AA1, AA2) inserts into the bilayer (BI) or LD monolayer ( $\Delta G_{AA}^{PMF}$ ), and subsequently associates with each other (AA1–AA2) once they are both in the bilayer or monolayer environment ( $\Delta G_{membrane}^{association}$ ). It is reasonable to assume that the free energy for two single amino acids to associate in the monolayer or in the bilayer is minimally different. Therefore, the free energy difference for the peptide to insert into the bilayer versus the monolayer is mainly dictated by the difference between the insertion free energies of each amino acid, as shown in the final equation.

**Table 4-1** Summary of simulations and lipid compositions used

<b>Table S5 (related to STAR Methods). Summary of simulations and lipid compositions used</b>							
<b>Protein</b>	<b>Membrane system</b>	<b>Simulation type</b>	$N_{POPC}^u: N_{DOPE}^u: N_{SAP1}^u$	$N_{POPC}^l: N_{DOPE}^l: N_{SAP1}^l$	$N_{TG}: N_{CLOL}$	<b>#</b>	<b>Time (ns)</b>
<i>LiveDrop</i>	bilayer	US	130:54:16	128:53:16	0:0	80	50–90
<i>LiveDrop</i>	monolayer	US	110:46:13	108:45:13	91:92	80	50–90
<i>LiveDrop</i>	bilayer	MD	133:55:16	138:56:16	0:0	4	1000
<i>LiveDrop</i>	monolayer	MD	130:54:16	136:57:17	651:0	2	1000
ALG14 hp	bilayer	MD	134:55:16	133:55:16	0:0	2	2000
ALG14 hp	bilayer	MD	134:55:16	134:55:16	0:0	2	2000
ALG14 hp	monolayer	MD	134:55:16	136:57:17	638:0	2	2000
ALG14 hp	monolayer	MD	134:55:16	136:57:17	640:0	2	2000

List of each type of simulation performed for *LiveDrop* and the predicted ALG14 hairpin (ALG14 hp) in both the bilayer and monolayer systems. ‘ $N^u$ ’ and ‘ $N^l$ ’ represent the number of molecules in the upper and lower phospholipid leaflets, respectively. The ‘#’ column represents the number of windows used in the umbrella sampling (US) simulations and the number of duplicates in the case of the molecular dynamics (MD) simulations. CLOL, cholesteryl oleate.

## **Chapter 5 Stressed Lipid Droplets: How Neutral Lipids Relieve Surface Tension and Membrane Expansion Drives Protein Association**

This chapter is adapted from [Kim et al., 2021] (68).

### **ABSTRACT**

Lipid droplets (LDs) are intracellular storage organelles composed of neutral lipids, such as triacylglycerol (TG), surrounded by a phospholipid (PL) monolayer decorated with specific proteins. Herein, we investigate the mechanism of protein association during LD and bilayer membrane expansion. We find that the neutral lipids play a dynamic role in LD expansion by further intercalating with the PL monolayer to create more surface-oriented TG molecules (SURF-TG). This interplay both reduces high surface tension incurred during LD budding or growth and also creates expansion-specific surface features for protein recognition. We then show that the auto-inhibitory (AI) helix of CTP:phosphocholine cytidyltransferase, a protein known to target expanding monolayers and bilayers, preferentially associates with large packing defects in a sequence-specific manner. Despite the presence of three phenylalanines, the initial binding with bilayers is predominantly mediated by the sole tryptophan due to its preference for membrane interfaces. Subsequent association is dependent on the availability of large, neighboring defects that can accommodate the phenylalanines, which are more probable in the stressed systems. Tryptophan, once fully associated, preferentially interacts with the glycerol moiety of SURF-TG in LDs. The calculation of AI binding free energy, hydrogen bonding and depth analysis, and in-silico mutation experiments support the findings. Hence, SURF-TG can both reduce surface tension and mediate protein association, facilitating class II protein recruitment during LD expansion.

## INTRODUCTION

Lipid droplets (LDs) are lipid and energy storage organelles, with cores composed of neutral lipids, such as triacylglycerol (TG) and sterol esters (SE), and bounded by a phospholipid (PL) monolayer (25, 47, 199, 200). LDs dynamically recruit proteins to their surfaces throughout their lifecycle in response to cellular needs. For example, when extra energy and/or lipids need to be stored, lipogenesis proteins associate with the LD surface. When energy is required, lipolysis proteins associate (20, 28). How LDs dynamically recruit different proteins depending on metabolic status is not fully understood and the focus of this research.

LD proteins are classified into two types depending on their origin. Class I proteins relocate from the ER membrane to LDs via ER-LD contact sites (66, 201). One example is glycerol-3-phosphate acyltransferase 4 (GPAT4), an enzyme that catalyzes one of the TG synthesis reactions (22, 65). They often contain a hydrophobic hairpin motif, which was proven to be sufficient for LD targeting by itself in the case of GPAT4 (66, 202). The main driving force of the ER-to-LD targeting has been suggested to be sequence and position-sensitive interactions between hydrophobic residues and the LD PL monolayer (66). This is supported by the finding that LD targeting can be abolished when the order of hydrophobic residues is shuffled. Tryptophan residues in particular were shown to gain significant energetic stabilization by relocating from unfavorable locations in the middle of the ER bilayer to a favorable region just below the glycerol groups in the LD monolayer. Class II proteins, in contrast, target LDs from the cytosol and generally contain one or more amphipathic helices (AHs). For example, the perilipin family exclusively target LDs via their conserved, 11-mer repeat regions (84, 203). CTP:phosphocholine cytidyltransferase (CCT) associates with expanding bilayers and LDs to cover the newly added surface area with



phosphatidylcholine (PC), but disassociates once expansion stops and surface tension decreases (86, 204, 205). Further information on LD targeting can be found in several reviews (28-30).

Over the last decade, AH–bilayer interactions have been extensively studied (206). The general mechanism of AH targeting to a bilayer surface is thought to consist of three steps: 1) sensing of packing defects by large hydrophobic residues, 2) peptide folding, and 3) stable association at the surface. Molecular dynamics (MD) simulations have provided insight into the molecular interactions involved in AH—bilayer targeting. All-atom resolution simulations have mainly focused on the first step because the timescale of peptide folding is much longer than most simulation limits (several microseconds). For instance, Voth and co-workers characterized packing defects using solvent-accessible surface area and colocalized those defects with hydrophobic residues upon the binding of an AH (207). Vanni and co-workers characterized packing defects using a Cartesian-based approach and proposed that deep defects are responsible for peptide targeting to bilayers (126). Monje-Galvan and Klauda approached peptide targeting from a thermodynamic context, computing interaction energies and contact frequencies for each amino acid with the membranes of the various PL compositions (208). For two bound conformations, they found that a Ser region or Tyr residue first approach the membrane and anchor the peptide, whereas charged residues, such as Lys, or bulky residues, such as Trp and Phe, contribute to stable association according to the calculated interaction energies (208).

The targeting of AHs to LDs is thought to be somewhat similar to that to bilayers. Previous work attributed the general AH targeting preference for LDs over bilayers to LDs' larger, more persistent packing defects (154). This general packing defect model can describe nonspecific adsorption of AHs to LDs; however, it does not explain differential targeting. For instance, the localization of perilipin family members can depend on the LD neutral lipid composition:

Perilipins 1a, 1b, and 5 associate with TG-abundant LDs while perilipins 1c and 4 associate with SE-abundant LDs (209). In contrast, Arf GTPase-Activating Protein 1 has a broad spectrum of association, targeting both bilayers and LDs (77, 210-212). This suggests that not all AHs are equal, and that their differences combined with bilayer/monolayer differences enable organelle-specific protein targeting.

Another example of differential targeting, this time based on the condition of membrane surfaces, is CCT (14, 213, 214). CCT catalyzes the rate-limiting step of the PC synthesis. It targets both expanding bilayers and LDs but disassociates when surfaces stop expansion (86, 204, 205). Biologically, CCT is up-regulated by PC deficiency in order to catalyze the PC synthesis pathway. Biophysically, the CCT M domain, which is responsible for membrane association, targets PC-deficient and negatively charged membrane surfaces. PC-deficiency is thought to play a role because PC is a cylindrical shape that creates few packing defects, while PC's precursors such as diacylglycerol (DAG) and phosphatidate (PA) or other lipids such as phosphatidylethanolamine (PE) are packing-discontinuity inducing lipids that facilitate CCT association (215). Negative charges are thought to attract the positively charged M domain (15 basic residues with a net charge of +1) (14, 216). Domain M of mammalian CCT $\alpha$  consists of three segments (14): The first segment closer to the N-terminus (residues 234-255) is basic and moderately amphipathic. The second segment (residues 256-271) is not conserved and nonessential for the binding of CCT $\alpha$ . The last segment (residues 272-293), the auto inhibitory (AI) helix, is highly amphipathic and contains a number of aromatic residues, including one Trp and three Phe residues. The AI helix is important in CCT $\alpha$  activation as well as membrane association (64, 217). In the cytosol, CCT $\alpha$  is inactive because the AI helix is docked onto the catalytic domain and thus suppresses the reaction.

On the other hand, when CCT $\alpha$  binds to a membrane surface, domain M becomes a long, unbroken helix and the AI helix is released from the catalytic domain (218, 219).

We recently characterized several physical properties of LD monolayers that are different from bilayers (68). Surface-oriented TG molecules (SURF-TG) were found to occupy 5-8% of the LD surface (under zero-surface tension) and to display order in an analogous manner to PLs. The exposed glycerol groups and acyl chains of these SURF-TG molecules in turn create packing defects that are chemically distinct from those found in bilayers. While SURF-TG molecules modulate the area per PL (APL), other TG molecules (considered CORE-TG) interdigitate dynamically with the PL tail region, increasing PL acyl defects and the degree of PL order. The hydration of the LD core was also found to be larger than previously thought, which plays an important role in class I targeting (66, 68). The display of TG at the LD surface suggests a plausible hypothesis for differential targeting in which neutral lipids either mediate LD protein interactions directly and/or alter monolayer properties to enable protein specificity. This idea is supported by recent experimental work that measured different binding affinities for a peptide to LDs with different neutral lipid compositions (34). The same order of binding affinities was demonstrated for the bare neutral lipids, implying direct interactions between a peptide and neutral lipids. Interestingly, changing the surface tension of these LDs shifted the binding affinities, however the order of the binding affinities was maintained.

Here we present all-atom MD simulations of LDs under an applied surface tension (i.e., stressed) to study the influence of surface expansion on the physical properties of stressed LDs and SURF-TG's involvement in CCT $\alpha$  targeting. We demonstrate that under an applied surface tension the LD system expands and more TG molecules are integrated into the monolayer surface. The integration of more SURF-TG thus decreases the applied surface tension, apparent in an

inverse correlation between the number of SURF-TG and surface tension once expansion is stopped until equilibrium is reached. Thus, a crucial interplay exists between surface composition and expansion that both enables stage-dependent protein recruitment and stress reduction under LD growth. We follow this with extensive MD simulations (13-14 replicas for each membrane) of the CCT $\alpha$  AI helix to study its association with both unstressed and stressed bilayers and LDs. Within the limited simulation time (1  $\mu$ s for each replica), we observe both the initial binding event and subsequent residue association are important for peptide targeting. Except for the unstressed bilayer membrane, which has the smallest packing defects, most of the other trajectories have at least one binding event. For the stressed surfaces, other residues can subsequently bind to the membrane surfaces, further leading to targeting; however, finding large defects near the initially inserted Trp/Phe residue was challenging for unstressed surfaces. These factors significantly contribute to the reported binding success to expanding LDs and bilayers. In the stressed bilayers, the initial binding event is mostly mediated by Trp278 (the only Trp in the AI helix) rather than any of the three Phe residues. This is likely due to its preference for membrane interfaces. In the stressed LDs, each Trp/Phe residue has an equal frequency of initial association, but the association site is preferentially TG defects. Once the AI helix is fully associated in the stressed LDs, Trp278 is highly coordinated with the glycerol moiety of SURF-TG via hydrogen bonding. The portion of time frames that have hydrogen bonding between SURF-TG and Trp278 was computed to be 0.25, which is relatively high provided that the LD surface under study contains  $\sim$ 18% SURF-TG. The depth analysis shows that the insertion depth of Trp278 is well aligned with the glycerol moiety of SURF-TG, but not with that of PLs. When Trp278 is mutated to Val278, the high coordination number disappears and the binding success decreases. Finally, the umbrella sampling simulation to compute the free energy of the AI binding to the stressed lipid droplet shows the good agreement

with the unbiased simulations. The binding free energy driving from water to the bound state is determined to be  $\sim 15.8$  kcal/mol. Collectively, our data demonstrate the role of SURF-TG as a stress reducer and peptide targeting mediator, additionally showing the importance of the specific interaction between neutral lipids exposed at the LD surface with the peptide in LD targeting.

## **METHODS**

### *System setup and simulation details*

Four different types of membranes were studied in this work: unstressed bilayers (BI0), stressed bilayers (BIe), unstressed lipid droplets (LD0), and stressed lipid droplets (LDe) where ‘e’ and ‘0’ represent expanding and zero-surface tension, respectively. Every leaflet of the membrane systems was composed of 88 3-palmitoyl-2-oleoyl-D-glycero-1-phosphatidylcholine (POPC), 37 2,3-dioleoyl-D-glycero-1-phosphatidylethanolamine (DOPE), and 10 phosphatidylinositol (SAPI) molecules, which is representative of the ER membrane and LD monolayer surfaces of human cells (4, 7). BI0 was prepared with the CHARMM-GUI membrane builder (55, 56, 220, 221). After 100 ns of the NPT (constant particle number, pressure, and temperature) simulation, a constant surface tension of 15 mN/m per leaflet (or 30 mN/m for the whole system) was applied in the XY-dimensions for 100 ns. The last structure was used for further simulations of the stressed bilayer (BIe). The LD system was modeled with a tri-layer structure made up of an 8nm-thick bulk TG region in between the two PL monolayers. The initial structure was taken from the end of a 10  $\mu$ s simulation described in previous work (68). More water was added in the Z-dimension, and the system was equilibrated for 1  $\mu$ s to create the LD0 system. A constant surface tension of 15 mN/m per leaflet was then applied for 2.3  $\mu$ s to model an expanding LD. The 300 ns structure from the constant surface tension simulation was taken for the further simulations of the LDe system. Of

note, the surface tensions of stable, non-expanding LDs (in vitro) and purified LDs (in vivo) were measured to be 1.63 mN/m and 3.5 mN/m, respectively (34, 35). The larger surface tension used herein was selected to capture expansion conditions. Importantly, no membrane instability such as pore formation was detected. TG was modeled with triolein in this work and the topology of triolein (TRIO.rtf) is available in <https://github.com/ksy141/TG>. All systems were solvated with TIP3P water (115, 222), and 0.15M NaCl was added. Details of the system setup are described in Fig. 5-15.

The AI motif of CCT $\alpha$  (275-295 amino acid) was taken from the crystal structure (PDB 4MVC) (64). The N- and C-termini were capped with acetylated N-terminus (ACE) and amidated C-terminus (CT2), respectively, to neutralize both ends. The helix was then located 1 nm above the highest atomic position of PL molecules, oriented such that the hydrophobic side of the AI helix faced the membrane surfaces. For each membrane system containing the AI helix, 13-14 replicas of MD simulations (NVT) were conducted with randomly generated initial velocities (each replica was run for 1  $\mu$ s). A point mutation (W278V) was also prepared to test the importance of this tryptophan via an in silico mutation experiment. The binding behavior of the mutant (W278V) with LDe was compared to that for the wildtype helix.

The MD simulations were performed using the GROMACS (version 2018) simulation engine (54) with the CHARMM36m lipid and protein force field (61, 63). Simulations were integrated with a 2-fs timestep. Lennard-Jones pair interactions were cut off at 12 Å with a force-switching function between 10 Å and 12 Å. The Particle Mesh Ewald algorithm was used to evaluate long-range electrostatic interactions (119). The LINCS algorithm was used to constrain every bond involving a hydrogen atom (223). Biased simulations (described below) were conducted with the external plugin, PLUMED2 (135, 224). The pressure was maintained semi-

isotropically using the Parrinello-Rahman barostat (120) at a pressure of 1.0 bar, a compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$ , and a coupling time constant of 5.0 ps. The temperature was maintained using the Nose-Hoover thermostat at 310 K (225, 226) with a coupling time constant of 1 ps. For NP $\gamma$ T (constant particle number, normal pressure, surface tension, and temperature) simulations, the Berendsen thermostat and barostat were used with the same parameters (227). The umbrella sampling (US) simulations were carried out with NAMD 2.14 to quantify the free energy of CCT $\alpha$  binding to LDe with NAMD 2.14 (228). The temperature was maintained using the Langevin thermostat at 310 K with a Langevin coupling coefficient of 1/ps (NVT ensemble) and no coupling for hydrogen atoms.

Molecular images included in this work were rendered using Visual Molecular Dynamics (VMD) (191) and PyMOL, and the trajectories were analyzed with MDAnalysis (62). Details of each simulation are described in Table 5-1.

**Table 5-1** Description of simulations.

<b>Simulations</b>		<b>Lipid composition</b> <sup>a)</sup>	<b>Time</b>	<b>Ensemble</b>	<b>Replicas</b>	<b>Initial struct.</b>
Bilayer (100 ns structure referred to as B10)	membrane	176 POPC + 74 DOPE + 20 SAPI	100 ns	NPT	1	CHARMM-GUI
Bilayer membrane (100 ns structure referred to as B1e)		176 POPC + 74 DOPE + 20 SAPI	100 ns	NP $\gamma$ T	1	From above sim.
Lipid (1 $\mu$ s structure referred to as LD0)	droplet structure	176 POPC + 74 DOPE + 20 SAPI + 429 TRIO	2 $\mu$ s	NPT	1	From previous simulation(68)
Lipid droplet (300 ns structure referred to as LDe)	structure	176 POPC + 74 DOPE + 20 SAPI + 429 TRIO	2.3 $\mu$ s	NP $\gamma$ T	1	From above sim.
B10 production run		Same as B10 or B1e	400 ns	NVT	1	B10
B1e production run		Same as B10 or B1e	400 ns	NVT	1	B1e
LD0 production run		Same as LD0 or LDe	2 $\mu$ s	NVT	1	LD0
LDe production run		Same as LD0 or LDe	2.3 $\mu$ s	NVT	1	LDe
CCT $\alpha$ folding PMF		N/A	100 ns <sup>b)</sup>	NVT	10 <sup>c)</sup>	CHARMM-GUI
CCT $\alpha$ binding PMF		Same as LD0 or LDe	162 ns <sup>b)</sup>	NVT	36 <sup>c)</sup>	LDe
B10 + CCT $\alpha$		Same as B10 or B1e	1 $\mu$ s <sup>b)</sup>	NVT	14	B10
B1e + CCT $\alpha$		Same as B10 or B1e	1 $\mu$ s <sup>b)</sup>	NVT	13	B1e
LD0 + CCT $\alpha$		Same as LD0 or LDe	1 $\mu$ s <sup>b)</sup>	NVT	14	LD0
LDe + CCT $\alpha$		Same as LD0 or LDe	1 $\mu$ s <sup>b)</sup>	NVT	13	LDe
LDe + W278V		Same as LD0 or LDe	1 $\mu$ s <sup>b)</sup>	NVT	13	LDe

a) The total number of lipid molecules of each type in a system.

b) Simulation time length for each replica or window.

c) The number of windows in the replica-exchange umbrella sampling simulation.

### *Packing defects*

We used the Cartesian-based algorithm to evaluate lipid-packing defects. For the upper leaflet, the lipid atoms (PL and TG) whose  $Z$  positions were greater than the threshold ( $z_{thr}$ ) were selected and projected onto a 1 Å-spacing two-dimensional (XY) grid. The threshold ( $z_{thr}$ ) was set to the average  $Z$  position of the phosphorus atoms of the upper leaflet minus 2 nm, which includes virtually all the lipid atoms of the upper leaflet. If a grid point overlaps with an atom (we define “overlapping” if the distance between the center of the atom and the grid point is less than the atom’s radius, which is taken from the CHARMM36 parameter set,(61) plus the half of the diagonal of the grid,  $\sqrt{2}/2$  Å), the  $Z$  position of the atom and the atom type are saved in the grid point. If a grid point overlaps with polar atoms (PL head groups or PL glycerol moieties), the grid



point is considered no defect. The other grid points were assigned to be PL acyl defects, TG acyl defects, or TG glycerol defects, based on the atom types and Z coordinates saved in the grid point. The atom which has the highest Z position determines the defect type. This algorithm is equivalent to finding a lipid atom that has the highest Z position at each grid point. The same procedure was repeated for the lower leaflet.

For each defect type, the neighboring elementary defects were clustered into one. If the clustered defect contains N elementary defects, it is considered to have a defect size of  $N \text{ \AA}^2$ . Consistent with other work (36, 68, 125-127, 207), the probability of finding a defect with the size of  $N \text{ \AA}^2$  was computed and fit to an exponential decay function:  $P(N) = ce^{-N/\pi}$ , where c is the normalization constant and  $\pi$  is the packing defect constant. If a defect is smaller than  $15 \text{ \AA}^2$  or the probability is lower than  $10^{-4}$ , the data was not included in the fit. The packing defect constant represents how slowly the decay function falls off. The higher the packing defect constant, the higher the probability of finding larger defects.

By using a deeper threshold ( $z_{thr}$ ) than previous applications of this algorithm (125, 127), every grid point overlaps with at least one atom, eliminating so-called ‘deep’ defects. (Previous papers set  $z_{thr}$  to 0.1 nm below the sn-2 carbon of the nearest glycerol, while we set it to 2 nm below the average phosphorus level of the either leaflet.) Deep defects have been useful in characterizing the very first step of peptide targeting (sensing of packing defects by large hydrophobic residues) (126). However, when a peptide is fully associated with a membrane, almost all grid points where the peptide is located become deep defects. Therefore, the concept of deep defects was eliminated in this work to avoid losing important information about the helix—packing defect interactions. The packing defect analysis script is available in

<https://github.com/ksy141/SMDAnalysis> with an example. The MDAnalysis (62) and parallel MDAnalysis (229) packages are required.

### *Helicity PMF*

To determine the stability of the AI helix in water, replica-exchange (131) umbrella sampling (132) (REUS) simulations were carried out to compute the potential of mean force (PMF) of the AI helix as a function of the distance between the carbon alpha atom of the first residue (Ile275) and that of the last residue (Pro295). Harmonic restraints with a force constant of 400 kJ/mol/nm<sup>2</sup> were placed in each of the 10 umbrella sampling windows with a 0.2 nm spacing over a range of 2.9 nm to 4.7 nm. An initial structure for each window was obtained from steered MD simulations, in which the same collective variable was biased with a force constant of 500 kJ/mol/nm<sup>2</sup> for 20 ns. The exchange was attempted every 1000 steps. The PMF was generated using the weighted histogram analysis method (WHAM) with a bin spacing of 0.02 nm (134, 230). The REUS simulations were run for 100 ns, and the trajectories were divided into five blocks, each containing a 20-ns trajectory. The first two blocks were considered equilibration and discarded. The error bars were the standard deviation of the block averages of the rest three blocks.

### *Binding PMF*

The free energy of binding of the AI helix to the LDe surface was estimated using umbrella sampling (US) simulations in the NVT ensemble. The PMF was calculated as a function of the perpendicular distance  $z_{CM}$  between the centers of mass (COMs) of the helix and the tri-layer system. 36 windows were prepared with a 0.1 nm spacing over a range of 3.5 nm to 7.0 nm. For

the constraints, a harmonic potential with a force constant of  $2.5 \text{ kcal/mol/\AA}^2$  was used in each window. An initial structure for each window was obtained from steered MD simulations, in which the alpha helix content was biased with a force constant of  $700 \text{ kcal/mol}/\alpha^2$  to maintain the helical structure. Each replica was then equilibrated for 90 ns. The PMF was generated using WHAM with a bin spacing of 0.01 nm. The US simulations were run for 72 ns (with no constraints on the alpha helical content), and the trajectories were divided into six blocks, each containing a 12-ns trajectory. The first 36 ns were discarded for the PMF calculation, and the error bars were obtained from the standard deviation of the last three blocks. For comparison, the normalized probability distribution of  $P(z_{\text{CM}})$  and the average location  $\langle z \rangle_{\text{P}}$  of the phosphorous atoms of the PL molecules in the binding leaflet were obtained from the unbiased MD simulations. The simulation described here was conducted with NAMD 2.14 (228).

### *Classification of TG*

TG molecules were categorized into two groups: SURF-TG and CORE-TG. The former is surface-oriented with their glycerol moieties exposed to the cytosol and acyl chains extended toward the LD core. The rest of TG molecules are considered CORE-TG. The designation was based on the distance between the TG glycerol moiety and the average position of PL tails (along the Z axis) of the closer leaflet. If the center of mass (COM) of the TG glycerol moiety is located above that of the PL acyl chains of the closer leaflet, it was classified as SURF-TG. Otherwise, it was denoted by CORE-TG.

### *Binding of AI helix*

Binding was defined based on the relative Z position of the Trp/Phe residues of the AI helix (Trp278, Phe285, Phe289, and Phe293) with respect to the phosphorus atoms. Specifically, successful binding was defined if the Z positions of all four residues are lower than the average Z position of the phosphorus atoms of the binding leaflet for more than 100 ns. Complete binding was defined if the average Z position of the four residues was less than the average Z position of the phosphorus atoms by more than 0.5 nm. Once the peptide binds to a membrane surface completely, the Z positions of all the residues remain the same. We used the 14 complete binding trajectories (9 and 5 for B1e and LDe, respectively) for the normalized coordination number analysis.

#### *Molecular groupings to analyze normalized coordination number*

To compute the interaction preference of the Trp/Phe residues (Trp278, Phe285, Phe289, and Phe293) with each membrane component, the coordination number was calculated between molecular groups. Each PL molecule was represented by 11 groups: polar head group, phosphate group, glycerol moiety, and four groups for each acyl chain. Similarly, each TG molecule was divided into 13 groups: glycerol moiety and four groups for each acyl chain. Each amino acid was represented by one backbone group and one side chain group. The coordination number between a side chain group of a Trp/Phe residue and each membrane group was computed by  $s = \sum [1 - (r_i/r_0)^6]/[1 - (r_i/r_0)^{12}]$ , where  $r_0 = 0.4$  nm and  $r_i$  is the distance between the side chain group and the membrane group. The normalized coordination numbers ( $\|s\|$ ) for POPC, DOPE, and SAPI were obtained by dividing the coordination numbers by the numbers of the corresponding PL molecules, which are, 88, 35, and 10 (in one leaflet), respectively. The

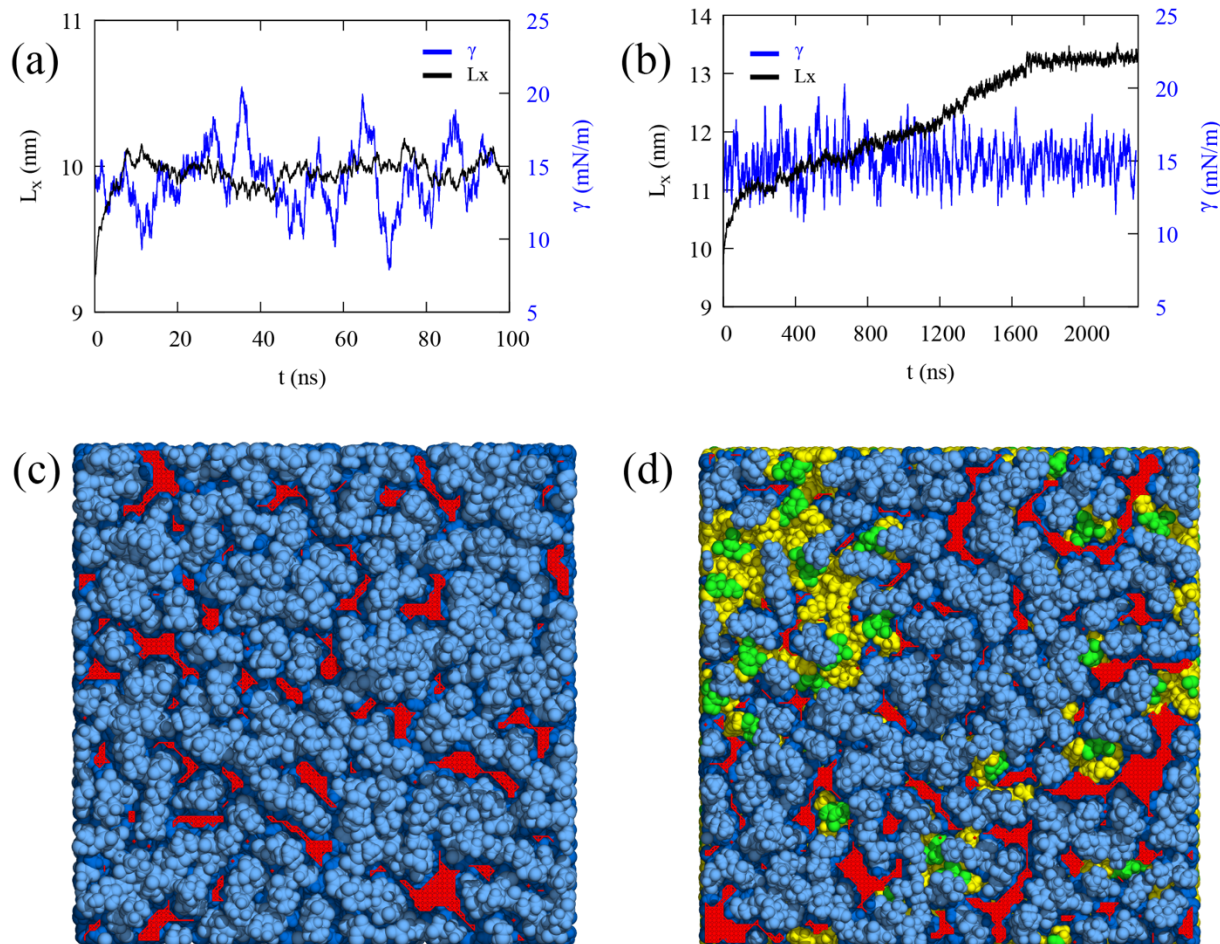
normalized coordination number ( $\| s \|$ ) for TG was obtained by dividing the number of SURF-TG molecules of the binding leaflet at each time frame.

## RESULTS AND DISCUSSION

### *Physical properties of unstressed and stressed bilayers and LDs*

MD simulations of the unstressed bilayer (BI0), stressed bilayer (BIe), unstressed LD (LD0), and stressed LD (LDe) were conducted to characterize the physical and structural properties of unstressed and stressed bilayers and LDs. The stressed bilayer and LD were prepared by applying the surface tension of 15 mN/m per leaflet to the equilibrated bilayer and LD, respectively (Fig. 5-15 and Methods). While the surface tension was maintained at 15 mN/m per leaflet in the NPT ensemble, different responses to the stress were apparent for the bilayer and LD systems (Figs. 5-1a and 5-1b). The X-dimension of the bilayer increases at the beginning and then levels off at ~10 ns, while that of the LD continues to increase with simulation time and takes longer to level off (~1.8  $\mu$ s). The bilayer response is expected. Under stress the bilayer expands, and more hydrophobic acyl chains are exposed to the cytosol (packing defects), but it can only expand so much. For the LD, on the other hand, stress induces much greater expansion than the bilayer membrane even though the same PL composition was used for both systems (7.9% increase for the bilayer and 38.1% increase for the LD in the X dimension, or equivalently 16.4% and 90.8% increase in the APL). The reason is that more TG molecules are pulled from underneath the PL monolayer (CORE-TG) to create more surface-oriented TG (SURF-TG) in response to the high surface tension. Since the simulation continues to keep the surface tension at the applied 15 mN/m per leaflet, the expansion continues, and more SURF-TG is created. Therefore, the stressed bilayer

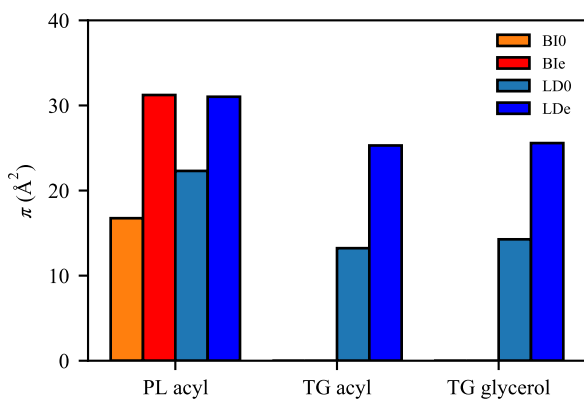
has increased PL acyl defects compared to the unstressed bilayer, while in the LD there is an increase in both PL acyl defects and TG-related defects (Figs. 5-1c, 5-1d, and 5-16).



**Figure 5-1** Bilayer and LD under the constant surface tension.

The X dimension  $L_x$  (black) and surface tension per leaflet  $\gamma$  (blue) for the bilayer (a) and LD (b) in the  $NP\gamma T$  ensemble ( $\gamma = 15 \text{ mN/m}$  per leaflet). The surface tension per leaflet (blue) was smoothed using a 2,000-point running average for the bilayer and a 6,000-point running average for the LD. The 100 ns and 300 ns structures of the  $NP\gamma T$  simulations for the bilayer (c) and LD (d), respectively, which are referred to as B1e and LDe in this study. The light and dark blue indicate polar groups (head groups and glycerol moieties) and acyl chains of PLs, respectively. The green and yellow indicate glycerol moieties and acyl chains of TG, respectively. PL acyl defects are shown in red and the other defects are not shown for visual clarity. The snapshots of unstressed bilayer (B10) and LD (LD0) are shown in Fig. 5-16.

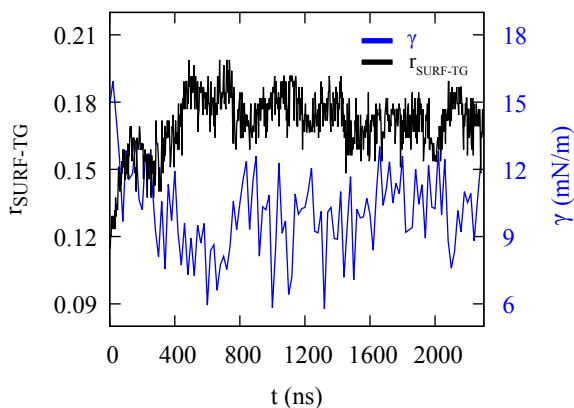
To compare the surface, molecular, and physical properties of the four membranes under a fixed level of expansion, NVT simulations were conducted (production run). The packing defect constants ( $\pi$ ) for four surfaces were evaluated (Fig. 5-2). From BI0 to BIe, the packing defect constant for PL acyl defects significantly increased (Fig. 5-2), consistent with the snapshots (Figs. 5-1c and 5-16). The packing defect constant for PL acyl defects also increased from LD0 to LDe, making them nearly equivalent to those in BIe. However, the increase from LD0 to LDe is not as large as the increase from BI0 to BIe, partially because LD0 has more PL acyl defects than BI0 due to the interdigitation with CORE-TG, as previously described (68). The increase is also less significant because the increased surface area in LDe includes increased TG acyl and glycerol defects (Fig. 5-2). Overall, the LD surfaces are more enriched with chemically distinct defects and therefore will more easily recruit certain proteins, as demonstrated in the previous work (154).



**Figure 5-2** Packing defect constants ( $\pi$ ) of PL acyl, TG acyl, and TG glycerol defects for BI0 (orange), BIe (red), LD0 (light blue), and LDe (dark blue).

In the NVT simulation of LDe, the surface tension relaxes to  $10.0 \pm 0.6$  mN/m from 15 mN/m in  $\sim 500$  ns (Fig. 5-3). In contrast, the surface tension of BIe under the NVT ensemble remained  $14.4 \pm 0.3$  mN/m. We suspected that the decrease in the surface tension could be due to increasing numbers of SURF-TG. We therefore define the SURF-TG ratio ( $r_{SURF-TG}$ ) with respect to PL as the number of SURF-TG ( $N_{SURF-TG}$ ) divided by the sum of the number of SURF-TG ( $N_{SURF-TG}$ ) and the number of PL ( $N_{PL}$ ):  $r_{SURF-TG} = N_{SURF-TG} / (N_{SURF-TG} + N_{PL})$ . Consistent

with our expectations, the amount of SURF-TG and surface tension are inversely correlated (Fig. 5-3), demonstrating SURF-TG does effectively reduce high surface tension. The system equilibrates when the induced surface tension from SURF-TG being a less effective surfactant than PLs balances the applied surface tension. Of note, we showed that SURF-TG can occupy 5-8% of the surface under zero-surface tension in our previous work (68). Under stress, the percentage of the SURF-TG in LDe increases to ~20% (Fig. 5-3). We expect the role of SURF-TG as a stress reducer is critical when LDs are budding or growing under the limited supply of PLs.



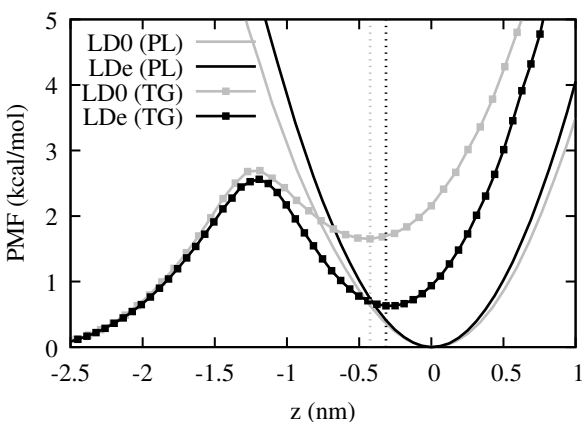
**Figure 5-3** NVT simulation of the stressed LD (LDe).

The SURF-TG ratio (black),  $r_{SURF-TG} = N_{SURF-TG} / (N_{SURF-TG} + N_{PL})$ , is inversely correlated with the surface tension per leaflet (blue). The surface tension per leaflet was block-averaged every 20 ns.

We then calculated the PMF of PL and TG with respect to the Z position by histogramming the Z position of C2 atoms for PLs and TGs (Fig. 5-4). In our previous work, we discussed how transitioning from CORE-TG to SURF-TG is a slow process such that SURF-TG only equilibrates after 2  $\mu$ s of simulation. Consistent with this, we argue the high transition barrier (~2.6 kcal/mol) from CORE-TG to SURF-TG in the unstressed LD has eluded the observation of SURF-TG in the previous shorter all-atom simulations (99, 154). SURF-TG is increasingly stabilized in the stressed system, making the transition barrier from SURF-TG to CORE-TG even higher in LDe. This implies SURF-TG stays longer at the LD surface in the stressed LD. The location of the local minimum of SURF-TG (indicated with dashed lines in Fig. 5-4) also shifts from -0.42 nm in LD0 to -0.31 nm in LDe. Thus, the depth of SURF-TG is more aligned with that of PLs in the stressed



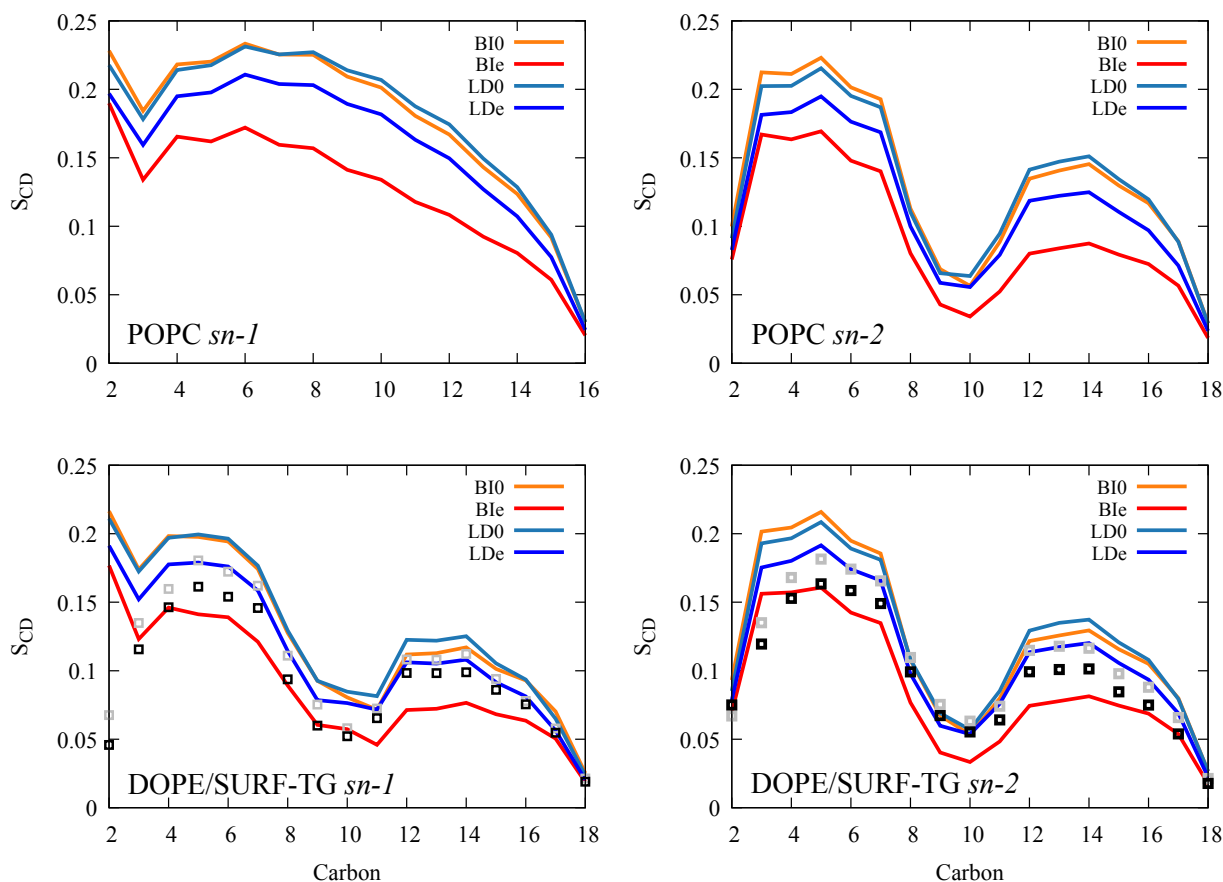
LD. We will later discuss how the insertion depth of Trp278 of CCT $\alpha$  aligns with SURF-TG's glycerol moiety depth, and they interact via hydrogen bonding. Finally, as one may expect, pulling PL from the averaged position resembles a harmonic potential (for the regions sampled) and is much more resistant than TG.



**Figure 5-4** Transition PMFs for PLs (continuous lines) and TG molecules (lines with squares) in the unstressed LD (gray) and stressed LD (black) from unbiased NVT simulations.

The average Z position of PL's C2 atom of either leaflet was set to zero and the positive Z is toward water. The vertical dashed lines indicate the local minima for the TG PMFs.

In order to probe the properties of PLs and SURF-TGs under the stress, the order parameters of PL and SURF-TG acyl chains were calculated (Fig. 5-5). There is a significant decrease in the PL's order parameters under surface tension for both bilayers and LDs. This is because the expansion in XY-dimensions causes the low density at each Z position (Fig. 5-17). Therefore, the acyl chains have more space for fluctuations, resulting in a reduction in their order. The order parameters of PLs in the unstressed LD (light blue) are slightly higher than those in the unstressed bilayer (orange), especially at the tail region, which is consistent with previous work (68, 99). This is because CORE-TG interdigitates with the PL acyl chains, resulting in a higher density in the lower tail region (-2 nm ~ -1.4 nm in Fig. 5-17). SURF-TG tracks DOPE's order parameters quite well for both stressed and unstressed LD, demonstrating its ability to act as a membrane component.



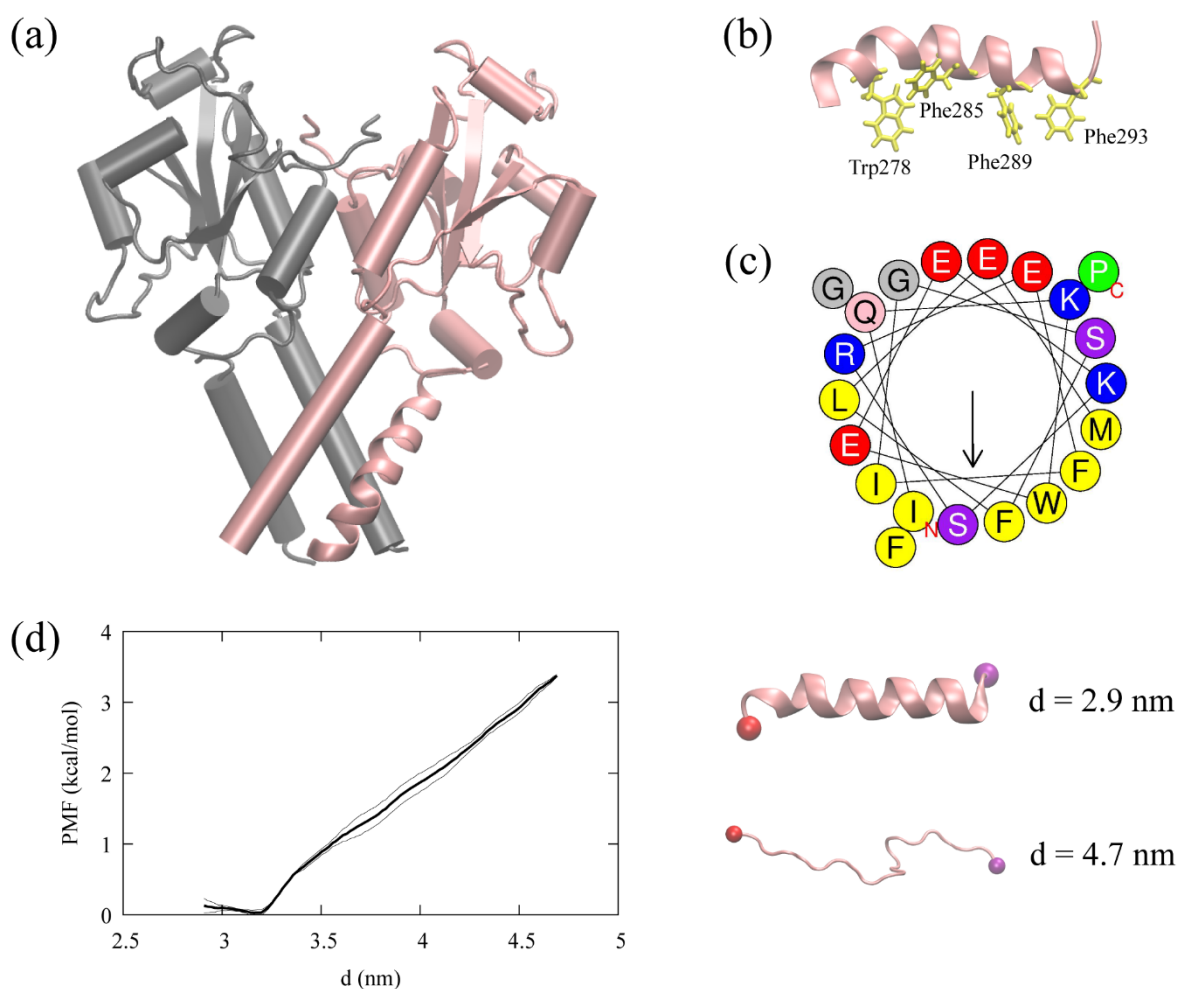
**Figure 5-5** Order parameters of POPC (top) and DOPE (bottom) from BI0 (orange), B1e (red), LD0 (light blue), and LDe (dark blue). Gray and black squares indicate the order parameters of SURF-TG from LD0 and LDe, respectively.

### *CCT $\alpha$ and the AI helicity*

We studied association of the CCT $\alpha$  AI helix (residues 275-295) to four the membrane systems (BI0, B1e, LD0, and LDe) using the extensive MD simulations. The AI helix was taken from the resolved structure of CCT $\alpha$  (PDB 4MVC) (Fig. 5-6a) (64). The AI motif is amphipathic and contains four Trp/Phe residues (Trp278, Phe285, Phe289, and Phe293), which play a key role in protein association with the membrane surfaces (Figs. 5-6b and 5-6c). Previous studies investigated the LD targeting behavior of CCT $\alpha$  P2, a longer peptide extending toward the N-terminus (residues 267-294) of the AI helix (154). The P2 peptide targets expanding LD surfaces,

but when the large hydrophobic residues (Ile, Leu, Phe, and Trp), including the four Trp/Phe residues (Trp278, Phe285, Phe289, Phe293), are mutated to Val, LD targeting is abolished. In this work, we also conducted the mutant experiment in silico with a single mutation, W278V, which will be discussed later.

Previous circular dichroism (CD) analysis of the peptides, CCT $\alpha$  PEPC22 (residues 267-288) and domain M (residues 237-293), demonstrated that these peptides are helical in the presence of anionic lipid vesicles and unstructured in water (231, 232). On the other hand, the AI motif remains  $\alpha$ -helical in the solution form of CCT $\alpha$  because it is docked onto the catalytic domain. When binding to a membrane, the M domain of CCT $\alpha$  becomes a long  $\alpha$ -helix (64, 217). Therefore, we hypothesized that the AI motif could remain helical over the course of CCT $\alpha$  membrane association. To test this hypothesis, we conducted a REUS simulation of the AI helix in an aqueous solution to obtain the PMF using the end-to-end distance as a collective variable. The biased simulations suggest that the helical structure of the peptide is stable in water within the accuracy of the force field (Fig. 5-6d). This motivates us to employ the helical structure as an initial structure of the AI motif in our membrane binding simulations.



**Figure 5-6** CCT $\alpha$  and the AI helix.

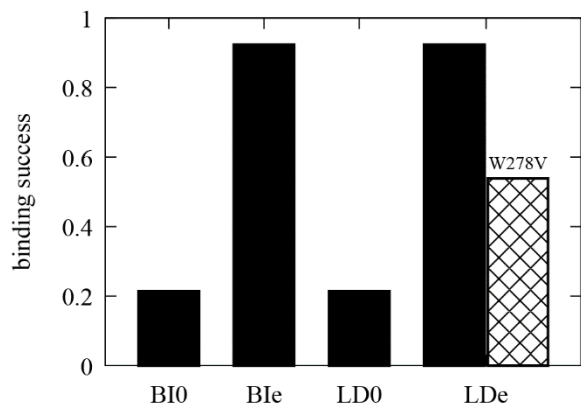
(a) The resolved CCT $\alpha$  structure (PDB code: 4MVC). CCT $\alpha$  is a dimeric protein, each monomer represented with a different color. The pink segment represented as a ribbon is the AI helix (residues 275-295) used in this study. (b) The AI helix in pink and four Trp/Phe residues in yellow. (c) The helical wheel of the AI helix computed using Heliquest. (d) The helical PMF (left) along the end-to-end distance,  $d$ , with the error bars (thin lines). The last structures of the first and the last window (right) with the carbon alpha atom of the first residue (red) and the last residue (purple).

*CCT $\alpha$  AI-membrane binding*

We first placed the AI helix 1 nm above the highest Z position of PLs in the unstressed and stressed bilayers and LDs. The AI helix was oriented such that the hydrophobic side faces the membrane surface. In order to obtain sufficient statistics, we conducted 13-14 replicas for each system, with each trajectory was run for 1  $\mu$ s. We define the binding of the AI helix to a membrane using the Z position of the average phosphorus atoms of the binding leaflet and the Z positions of the four Trp/Phe residues (Trp278, Phe285, Phe289, and Phe293). When the four residues remain below the average phosphorus level (referring to the upper leaflet) for more than 100 ns, the AI helix was considered to have successfully associated with the membrane. This definition is equivalent to the previous paper (154) except the previous paper used only one large hydrophobic residue while we used a stronger condition of all four Trp/Phe residues. The difference was chosen because the previous work used the random coil as an initial structure and studied the very first association of the binding, while we used the helical structure and therefore can better sample the phase space in which the peptide can fully associate. For the mutant (W278V) trajectories, two trajectories were excluded from the successful binding category that satisfy the above condition but have high fluctuations in the Val278 depth profile in the last 100 ns. Val278 effectively disassociates from the membrane surface, which did not occur in the natural (unmutated) simulations.

The binding success was simply calculated as the occurrence by dividing the number of trajectories that have successful association within our simulation length by the number of total trajectories (13-14 replicas). Consistent with the experimental observation (86, 204, 205), the AI helix shows the high binding success with the stressed bilayer and the stressed LD (~90%) in the timescale of microseconds (Fig. 5-7). The single mutation (W278V) reduces the binding success with the stressed LD to ~54%. Further analysis and simulations would be required to confirm the previous predictions that the AI helix should target LD0 over BI0 due to the bigger, more persistent

packing defects present at the LD surface (154). However, analyzing the depth profile of each Trp/Phe residue does support this argument. For 4 out of 14 BI0 trajectories, no association of any Trp/Phe residues was observed. On the other hand, 13 out of 14 LD0 trajectories have at least one initial association.

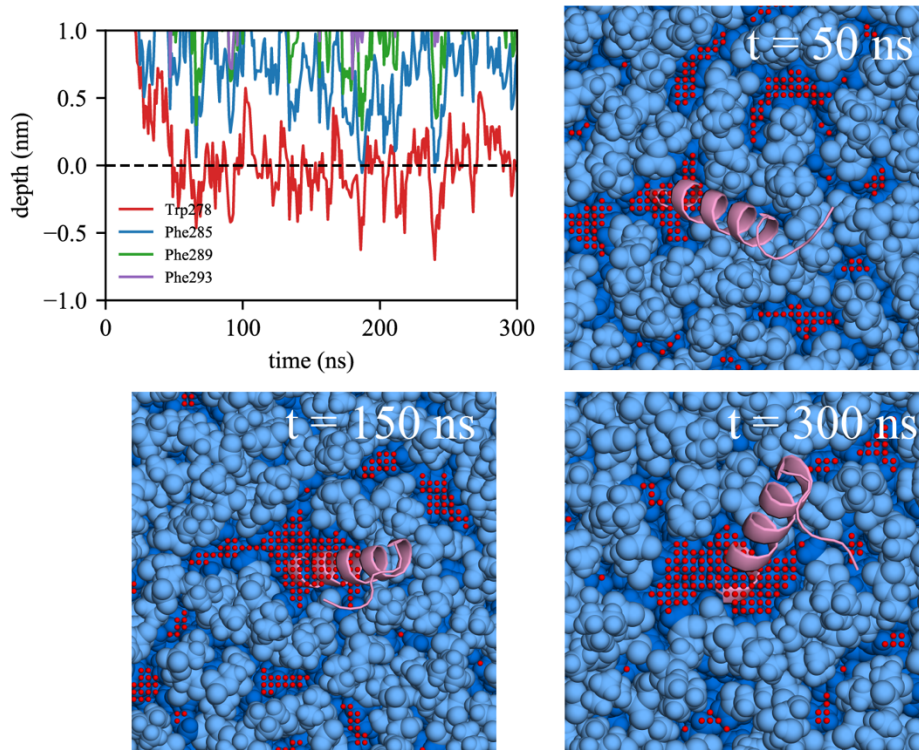


**Figure 5-7** Binding success (occurrence) of the AI helix (wildtype in black, mutant in crosshatch) to four different surfaces. For each membrane, the simulations were replicated 13-14 times.

Comparing association with BI0 versus BIe, the lack of initial association of any Trp/Phe residues in 4 out of 14 BI0 trajectories is likely due to the limited packing defects in the BI0 surface. In contrast, all the BIe trajectories include association with at least one of the Trp/Phe residues. We also observed the active role of Trp in peptide targeting in BIe using the depth profile analysis. Although the AI helix has three Phe and only one Trp, it was Trp278 that first associated in 8 out of 13 BIe trajectories. We expect this is because of Trp's preference for membrane interfaces (66, 233). In order to have successful binding, subsequent association of the other large hydrophobic residues should follow. We focused on two bilayer simulations, one with unsuccessful association to BI0 (Fig. 5-8) and another with successful association to BIe (Fig. 5-9). For both simulations, Trp278 was first inserted and the peptide stayed at the membrane surface. In the unsuccessful binding (Fig. 5-8), Trp278 is located at the average phosphorus level of the upper leaflet (a depth of 0 represents the average phosphorus level). However, it could not further

embed, nor could the other Phe residues embed due to a lack of large, neighboring packing defects. In contrast, in the successful simulation (Fig. 5-9), Trp278 was first inserted at 25 ns, followed by Phe285 at 70 ns. At 75 ns, a large, neighboring defect emerged near the defect that eventually accommodated the other Phe residues (Phe289 and Phe293) resulting in full association.

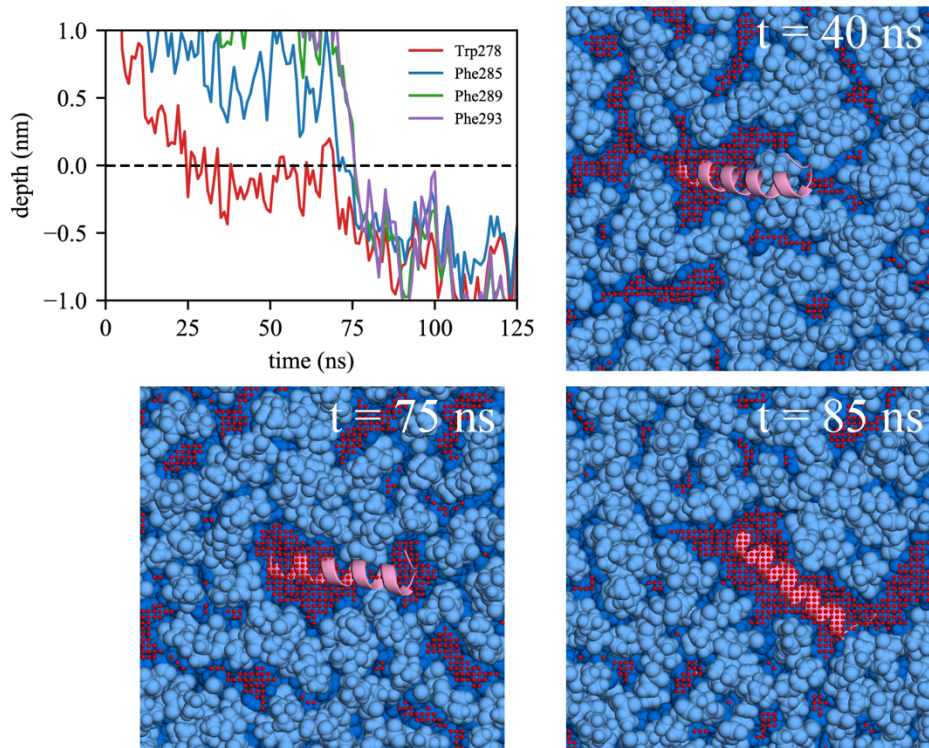
Focusing on the LDe simulations demonstrates that TG defects play a significant role in peptide targeting. In 9 out of 13 LDe trajectories, the initial binding happened at a TG defect, suggesting that SURF-TG facilitate the initial association of Trp/Phe residues. In contrast to the BLe simulations, the four Trp/Phe residues have an equal frequency of initiating binding to LDe. Specifically, the number of trajectories in which Trp278, Phe285, Phe289, and Phe293 initially associated was 4, 3, 3, and 3, respectively. We expect this is because of Phe's stabilization at the LD interface due to interactions with SURF-TG. Our previous work supports this argument showing greater stabilization of Phe during permeation into a relaxed LD compared to that into a relaxed bilayer (66). We expect the stabilization will be even greater in LDe because of the increased amount of SURF-TG compared to LD0. Finally, one successful binding trajectory to LDe is shown in Fig. 5-10. Trp278 initially associates with the TG acyl defects at 27 ns and other Phe residues target the PL acyl defects, emerging near Trp278 at 95 ns. Once the peptide is fully associated, the peptide preferentially interacts with the hydrophobic surface of the membrane via PL and TG acyl defects as shown in Fig. 5-10 (snapshot from 857 ns).



**Figure 5-8** Unsuccessful binding of the AI helix in BI0.

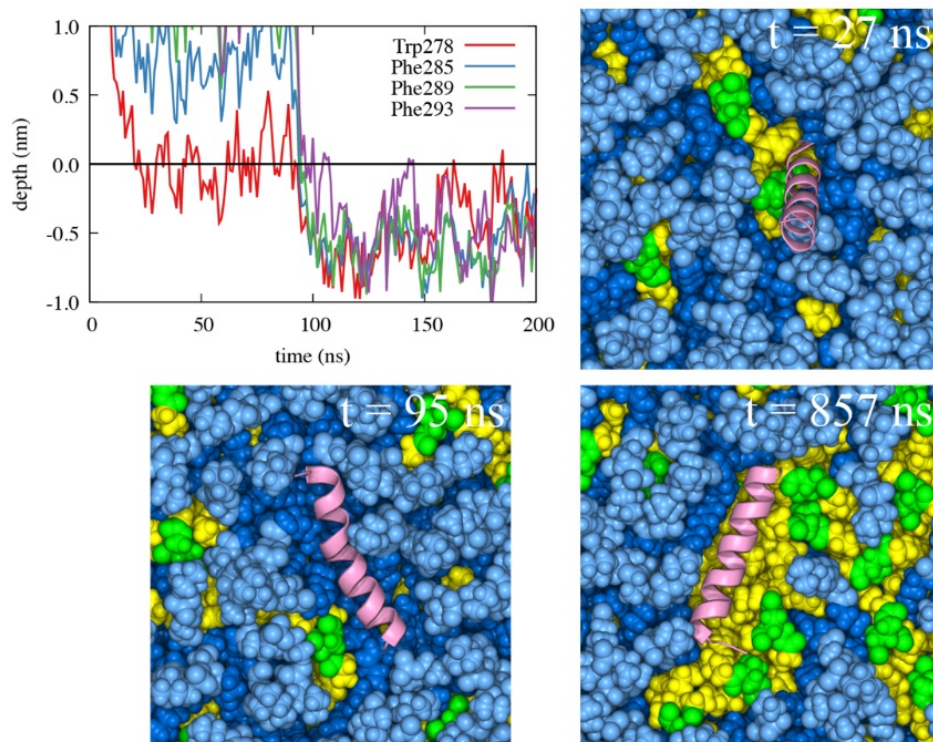
The depth profile is shown in top left. The average  $Z$  position of the phosphorus atoms of the upper leaflet was set to 0. The positive  $Z$  is toward water and the negative  $Z$  toward the membrane center. Snapshots with simulation times are shown. The light and dark blue indicate polar groups and acyl chains of PLs, respectively. The PL acyl defects (red) are colocalized. The AI helix is shown in pink. Only the part of the system is shown in this figure. The actual membrane system is larger.





**Figure 5-9** Successful binding of the AI helix in BLE.

The same color code was used as in Figure 5-8.



**Figure 5-10** Successful binding of the AI helix in LDe.

The same color code for PL and protein was used as in Figure 5-8. Additionally, the glycerol moieties and acyl chains of TG are shown in green and yellow, respectively. Defects are not shown for visual clarity.

To quantify the frequency of interactions between each Trp/Phe residue and each membrane group once the peptide is fully associated, we selected and analyzed the trajectories that have complete binding. We define the complete binding if the COM of the four Trp/Phe residues are 0.5 nm below that of the phosphorus atoms of the binding leaflet. This is an even stronger condition than that used to define the successful binding above. In these trajectories, the Z position of each residue remains almost the same following full association. Nine BIE trajectories and five LDe trajectories satisfied this condition. For these simulations, we first reduced the resolution of the all-atom trajectories to collective coordinates as described in the Methods and shown in Figs. 5-11a and 5-11b. This was done to reduce the amount of redundant information and increase the

efficiency of the coordination analysis. In each trajectory, the coordination number between each Trp/Phe residue and each membrane group was computed (averaged over the frames that have complete binding). The normalized coordination number ( $\langle |s| \rangle$ ) was obtained by dividing the coordination number by the number of molecules of each type at the binding surface (see the Methods). We then averaged the normalized coordination number over the trajectories that have complete binding for each system (Nine trajectories for B1e, five trajectories for LDe, and three trajectories for W278V-LDe). The error bars (Fig. 5-18) revealed that the peptide-SAPI interactions were not statistically reliable. This is not a surprise given that each trajectory samples different phase space of peptide targeting, and SAPI is the minor membrane component with only 10 molecules per leaflet. Thus, SAPI was excluded from this analysis. In general, the Trp/Phe hydrophobic residues have higher preference for membrane molecules with two or more unsaturated double bonds (DOPE, and TG) over POPE, which has only one double bond in the acyl chains (Figs. 5-11 and 5-18). Interestingly, the high contact between the glycerol moiety (GL) of TG and Trp278 was found in LDe simulations (Figs. 5-11c and 5-18). This is explained by hydrogen bonding between Trp278 and the GL of SURF-TG, which can be less than 2 Å in the trajectories, indicative of a stable hydrogen bond (Fig. 5-11d). In contrast, the phenylalanine residues which lack the hydrogen bonding ability do not show a clear preference for TG's GL group (Fig. 5-18).

This interaction is also apparent in the fraction of frames that have the hydrogen bond between Trp278 and the oxygen atoms of SURF-TG, which is  $0.244 \pm 0.125$  from the frames that have the complete binding in the LDe trajectories. The hydrogen bond was defined with the donor-acceptor distance cutoff of 0.4 nm and the angle cutoff of 150°. Thus, Trp278 forms a hydrogen bonding with SURF-TG in approximately one out of four frames, which is relatively high number

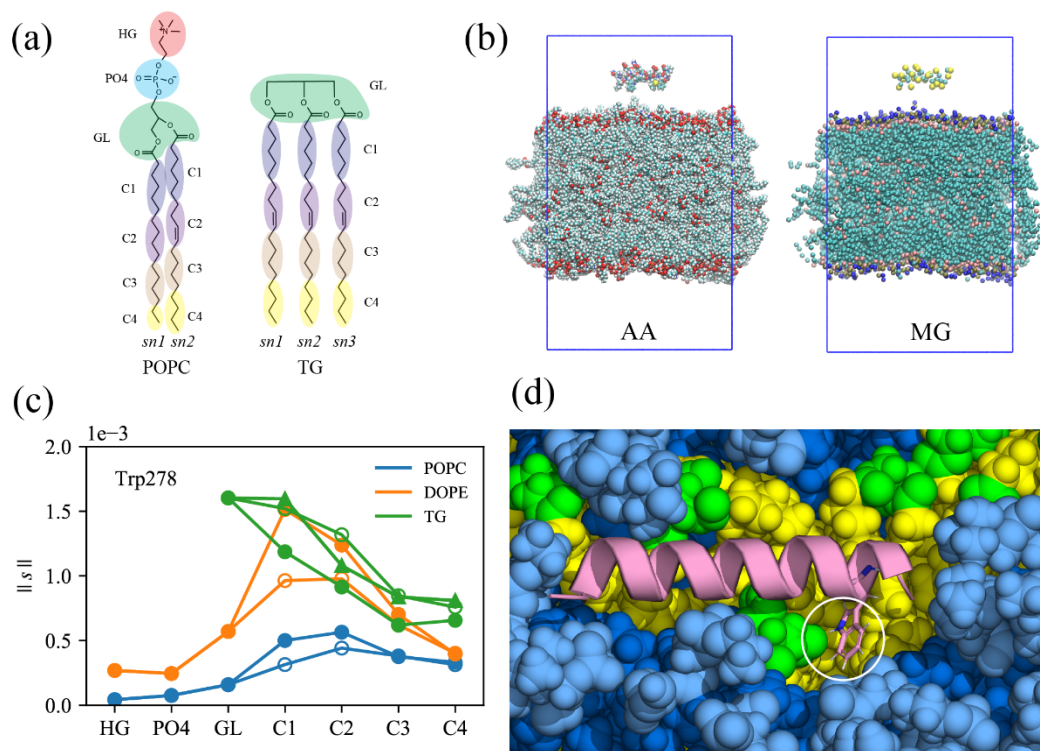
considering the lipid composition of the stressed LD surface is only ~15-18% of SURF-TG. Depth analysis further supports Trp278's preferential interaction with SURF-TG over the PL glycerol group. The preferred insertion depth of Trp278 (-3.7 Å averaged from bound simulations where the depth of 0 represents the average Z position of the PL C2 atom of the binding leaflet) is well aligned with the Z positions of the six SURF-TG oxygen atoms (centered at -2.9 Å), but far from that of the four PL glycerol oxygen atoms (centered at +1.3 Å).

The computed binding free energy of the AI helix (Fig. 5-12) also supports our findings. To ensure the findings from our unbiased simulations are statistically informative, we calculated the free energy profile (PMF) of the helix association with the LDe surface with US simulations. The distance  $z_{CM}$  between the COMs of the helix and the tri-layer system was used as the collective variable. The free energy minimum is located at  $z_{CM} \approx 4.35$  nm, which is below the average location of the phosphorous atoms of the PL molecules in the binding leaflet  $\langle z \rangle_P \approx 4.48$  nm (green dashed line) and consistent with hydrogen bonding between Trp278 and SURF-TG. The probability distribution  $P(z_{CM})$  of the helix from the unbiased simulations in which the AI helix is bound (shown by red bar graph) aligns well with the free energy profile. The slight shift toward the LD center in this distribution is expected given that the simulations in which association is weak or absent are not included. The alignment along with the barrierless PMF and large binding free energy (-15.8 kcal/mol) support the validity of the data gathered from the unbiased simulations and the observed rapid association with the LDe surface.

Furthermore, a point mutation (W278V) was prepared to confirm the importance of Trp278 via an in-silico mutation experiment. In the mutated trajectories (W278V), three LDe trajectories satisfied the complete binding condition in which the COM of the four Val/Phe residues were 0.5 nm below that of the phosphorus atoms of the binding leaflet. Consistent with our expectations,

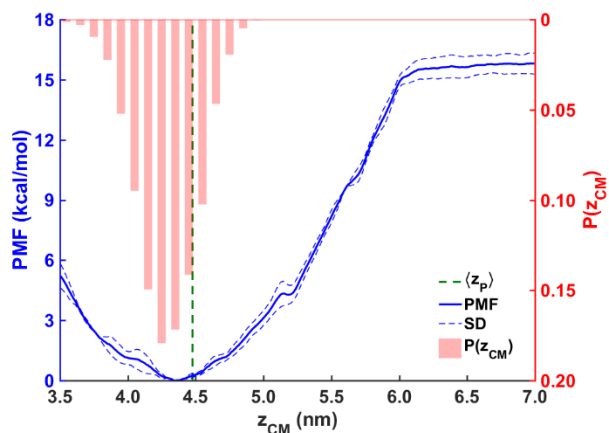
Val278 does not show a preference for the GL of SURF-TG due to the lack of the hydrogen bonding ability (Fig. 5-13). Taken together, this strongly supports the frequent interaction with the GL of SURF-TG being specific to tryptophan due to its combined features of being a large aromatic residue with the ability to hydrogen bond.

Finally, we characterized the nature of the defect-residue interactions by calculating the area of each defect type under Trp278 and Phe285 when they are fully associated (i.e. when the residue COM is 5 Å below the average phosphorus level of the binding leaflet). Figure 5-14 shows the normalized joint probability distribution  $P(a_{\text{PL}}, a_{\text{TG}})$  of each residue associating with PL and TG defects ( $a_{\text{PL}}$  and  $a_{\text{TG}}$ , respectively). The heat map adopts a lower triangular form, consistent with maximum areas of 18 Å<sup>2</sup> and 14 Å<sup>2</sup> for Trp and Phe, respectively. Lower cumulative values are due to overlap with PL head groups. Consistent with lipid packing defects, which are much larger than the dimensions of the residues and characterized by compositional heterogeneity, the residues often interact with a mixture of PLs and TGs. In fact, the probability that the residues reside on a mixture of PL and TG elementary defects, simply the integral of  $P(a_{\text{PL}} \neq 0, a_{\text{TG}} \neq 0)$ , turns out to be 74.5% for Trp278 and 60.1% for Phe285. Therefore, when Trp278 is completely bound to the stressed LD, its interaction with SURF-TG does not necessitate its localization to a homogeneous packing defect.



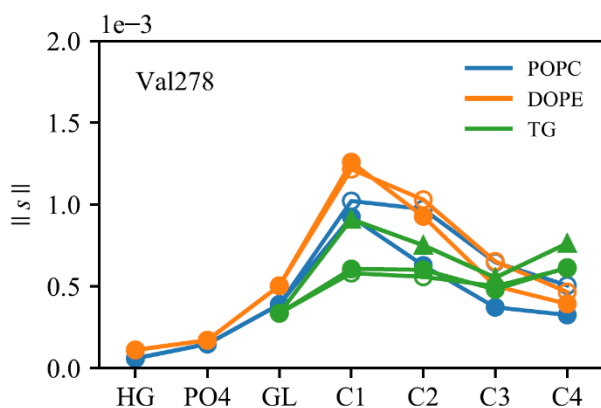
**Figure 5-11** Molecular groupings (MG) of all-atom trajectories to analyze the normalized coordination number.

(a) PL and TG, represented with 11 and 13 molecular groups, respectively. (b) The all-atom (left) snapshot of the first timeframe and the corresponding collective coordinates (right). Each amino acid consists of one backbone group and one side chain group. Water and ions are not shown for visual clarity. (c) The normalized coordination number ( $\|s\|$ ) between Trp278 and each membrane group, averaged over the trajectories that have the complete binding for the stressed LD. The filled circles, unfilled circles, and triangles represent *sn-1*, *sn-2*, and *sn-3* chains, respectively. (d) The snapshot showing the hydrogen bond between Trp278 (represented as sticks) and the glycerol moiety of TG. The same color code was used as in Fig. 5-10.



**Figure 5-12** PMF (blue) of the AI helix binding to the LDe surface, obtained with the umbrella sampling simulations.

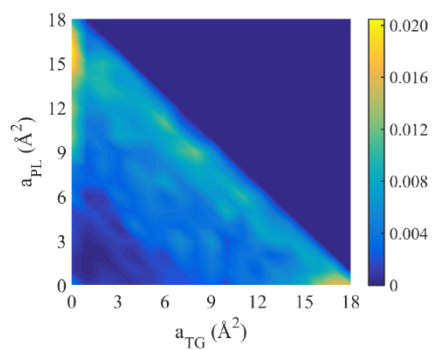
The error bars (blue dashed lines) were obtained from the standard deviation of the last three blocks. The normalized probability distribution of the AI helix  $P(z_{CM})$  (red bars) and the average location of the PL phosphorous atoms  $\langle z \rangle_P$  (green dashed line) were obtained from the unbiased MD simulations.



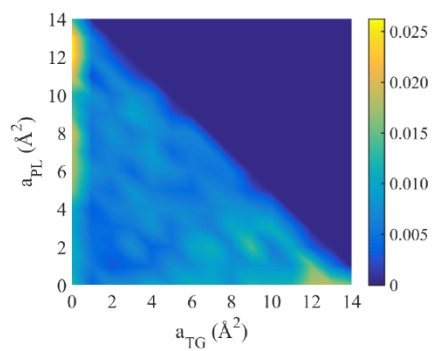
**Figure 5-13** Normalized coordination number between Val278 and each membrane group, averaged over the mutant trajectories (W278V) that have complete binding with the stressed LD.

The filled circles, unfilled circles, and triangles represent  $sn-1$ ,  $sn-2$ , and  $sn-3$  chains, respectively.

(a) Trp278



(b) Phe285



**Figure 5-14** Normalized joint probability distribution of the defect areas,  $a_{PL}$  and  $a_{TG}$  that Trp278 (a) or Phe285 (b) associates with, given that each residue is completely associated.



## CONCLUSIONS

LDs store neutral lipids (e.g. TG) surrounded by a PL monolayer. We recently demonstrated that TGs interdigitate with the PL monolayer as surface-oriented (SURF-TG) molecules creating chemically distinct packing defects from those found in PL bilayers (68). In this work, we further characterize the physical and surface properties of stressed LDs under surface tension, which is expected to increase during various LD lifecycle stages such as budding and growth. Although SURF-TG molecules are worse surfactants than PLs, and increase surface tension under non-expansion conditions, they can decrease applied high surface tension during expansion. As shown herein the SURF-TG percentage increases from 5-8% under zero-surface tension conditions to ~18% at ~10mN/m. Consistently, once expansion is stopped, the inverse correlation between the percentage of SURF-TG with the surface tension demonstrates that SURF-TG can effectively reduce high surface tension. Thus, the interplay between SURF-TG and surface tension will facilitate growth, such as during the initial stages of LD formation, by mediating stress alongside the association of proteins like lipid droplet assembly factor 1 (38). Similarly, growth and shrinkage will alter the LD surface, enabling stage-dependent protein recruitment.

We also investigate how SURF-TG is involved in AH targeting in both stressed and relaxed LDs. We chose the AI helix of CCT $\alpha$  as a test peptide, motivated by its conditional binding to the expanding (stressed) monolayer and bilayer surfaces. In water, the AI helix is helical due to its association with the catalytic domain of CCT $\alpha$  (64, 217). Therefore, we used the AI helix as our initial structure. We further support the validity of this choice by showing that the helix is also stable in water with enhanced free energy simulations. The advantage of being able to use the helix as the initial peptide structure is that it enables sampling of the membrane-associated phase space and bypasses the folding step common for other AH helices. This allows us to discuss residue—

membrane interactions in the fully associated state, which will be valuable when considering disassociation as well. The computational cost of folding an AH helix during membrane association is what limited most previous MD simulations to the very first step of peptide targeting (sensing packing defects by hydrophobic residues) (126, 154, 207).

Comparing trajectories for stressed and relaxed bilayers (B1e and B10, respectively) and LDs (LDe and LD0, respectively) reveals that the initial association is attempted more frequently for the stressed systems, and least frequently for the unstressed bilayer. Full association is dependent on the presence of large defects near the initially inserted residue, which is far less probable in the unstressed surfaces. Thus, the increase in size and prevalence of packing defects in expanding bilayer and monolayer surfaces is what drives conditional CCT $\alpha$  association during expansion. Our bilayer results further reveal that Trp is more “active” than Phe in the initial targeting. The number of trajectories in which Trp first associates with a surface is greater than those in which Phe first associates, despite the fact that there are three Phe residues to only one Trp. Additionally, the inserted Trp effectively holds the AI helix localized at the interface while the other large hydrophobic residues are looking for the neighboring defects. In contrast, when Phe first associates, it is sometimes kicked out of the membrane surface. The active role of Trp in peptide targeting is likely due to its preference for membrane interfaces (66, 233). It is interesting to note that Trp278 is also located at the bottom of the helix, closest to where CCT $\alpha$  would interface with the membrane. In contrast, the specificity of Trp in the initial targeting is reduced in the stressed LD, potentially because of Phe’s stabilization at the LD surface. The frequency that Trp or Phe residues initially associated was comparable, but both preferentially targeted TG defects over PL defects in our stressed LD simulations.

Once the AI helix is fully associated with the stressed LD, the interaction between the glycerol moiety of SURF-TG and Trp278 was notably high, and the interaction frequency was every out of four frames on average. This specific interaction is driven by hydrogen bonding, which is lacking for the other Phe residues (Phe285, Phe289, and Phe293). The preference of Trp278 for the glycerol moiety of SURF-TG over that of PL is further supported by the alignment of SURF-TG and the insertion depth of Trp278. Additionally, the single mutation (W278V) in the helix reduces the binding success from ~90% to ~54%. There is no preferential interaction between the glycerol moiety of SURF-TG and Val278. Finally, the US simulations confirmed the unbiased MD results, demonstrating the free energy difference between the unbound and bound states is 15.8 kcal/mol. Collectively, this suggests that tryptophan residues play a special role in membrane and LD association due to the combination of their bulky aromatic group and ability to hydrogen bond. We anticipate this hydrophilic interaction at the hydrophobic interface can increase binding affinities. For instance, the LD targeting domain of comparative gene identification-58 (CGI-58 or ABHD5), an activator of adipose triglyceride lipase (ATGL), is enriched with Trp and mutating those residues to Ala abolishes LD targeting (234, 235). Also, during lipolysis, CCT1 has the weakest binding affinity with LD surfaces and CGI-58 has one of the strongest among the proteins tested (75). The possible explanation could be the number of Trp residues and the interaction between Trp and SURF-TG. The whole membrane binding domain of rat CCT $\alpha$ , which ranges from residue 234 to 293, has only one Trp residue, whereas CGI-58 has three.

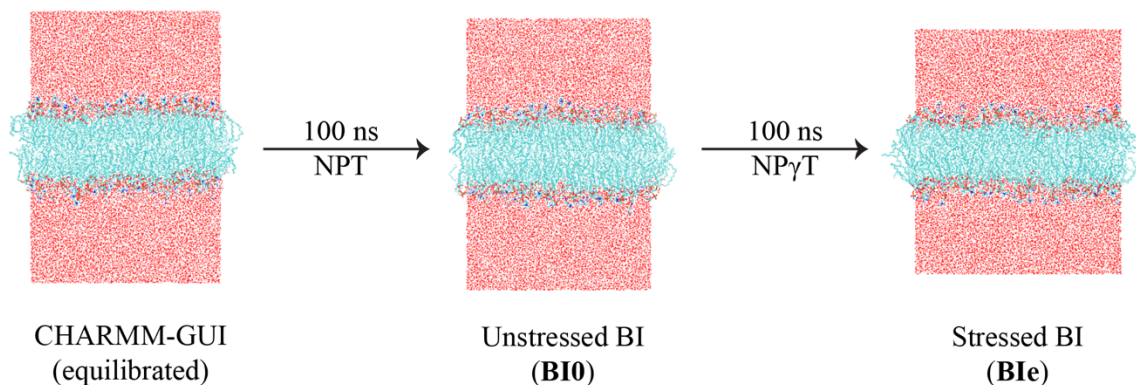
Taken together, our results provide new perspectives on the mechanisms behind preferential AH targeting to expanding bilayers and LD monolayers, the interplay between SURF-TG molecules and surface tension, and the specific role of SURF-TG molecules in LD protein association.

## **ACKNOWLEDGEMENTS**

S.K. thanks Dr. Rosemary Cornell for useful discussions on CCT $\alpha$ . We thank Dr. Jeeyun Chung, Dr. Rich Pastor, Dr. Robert V. Farese, Jr., Dr. Tobias C. Walther, and Dr. Gregory A. Voth for general discussions. This research was supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) through grant 5R01GM097194. S.K. was supported by the National Institute of General Medical Sciences of the National Institutes of Health under NIH award number R01-GM063796 (G. A. Voth, Principal Investigator). The MD simulations were performed on the high-performance GPU cluster (GM4) at the University of Chicago Research Computing Center, supported by NSF grant DMR-1828629 and the Stampede2 supercomputer at the Texas Advanced Computing Center (TACC) through allocation MCB200018 with resources provided by the Extreme Science and Engineering Discovery Environment (XSEDE) supported by NSF grant ACI-1548562.

## Bilayer membrane (BI) system setup

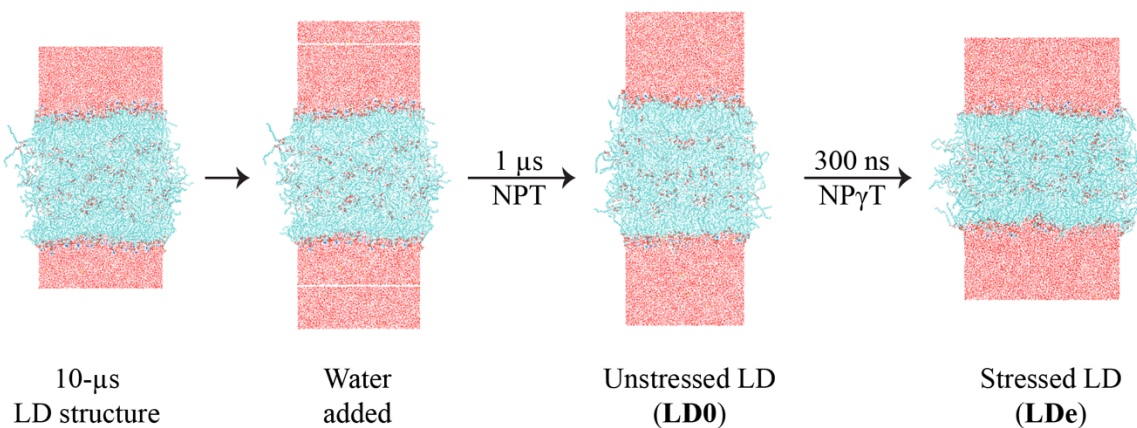
88 POPC, 37 DOPE, 10 SAPI per leaflet



---

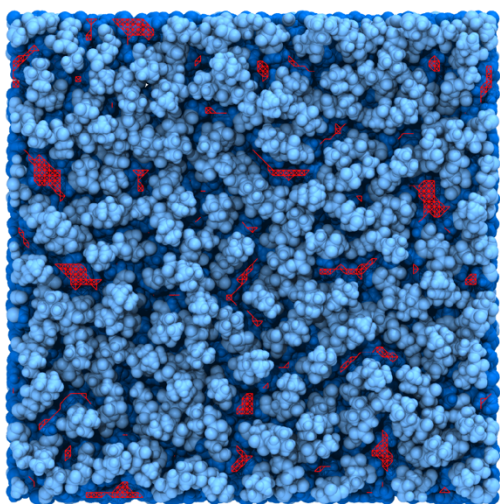
## Lipid droplet (LD) system setup

88 POPC, 37 DOPE, 10 SAPI per leaflet

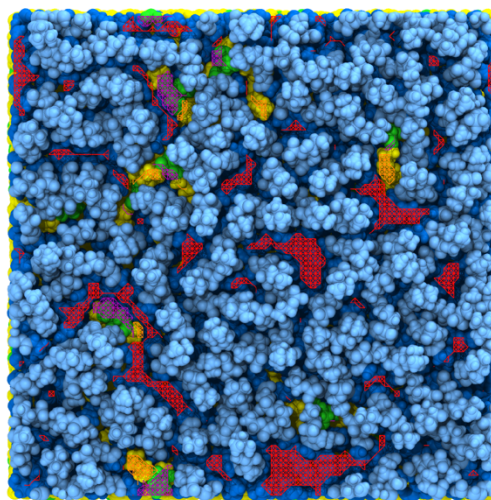


**Figure 5-15** System setup for the unstressed bilayer (BI0), stressed bilayer (BIe), unstressed LD (LD0), and stressed LD (LDe).

The initial structure of the bilayer was prepared with the CHARM-GUI membrane builder. The initial structure of the LD was taken from the previous work in which the LD system was equilibrated for 10 μs.



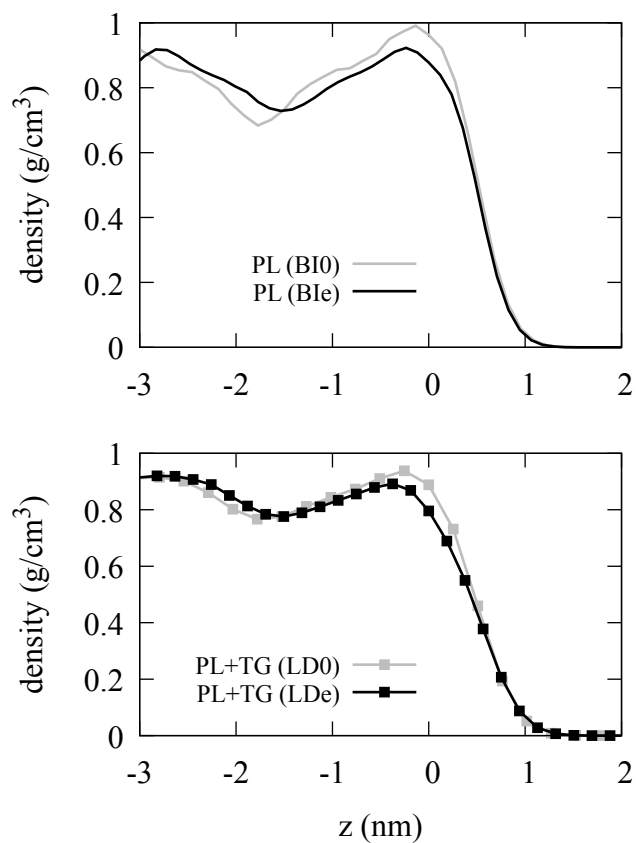
Unstressed bilayer (BI0)



Unstressed LD (LD0)

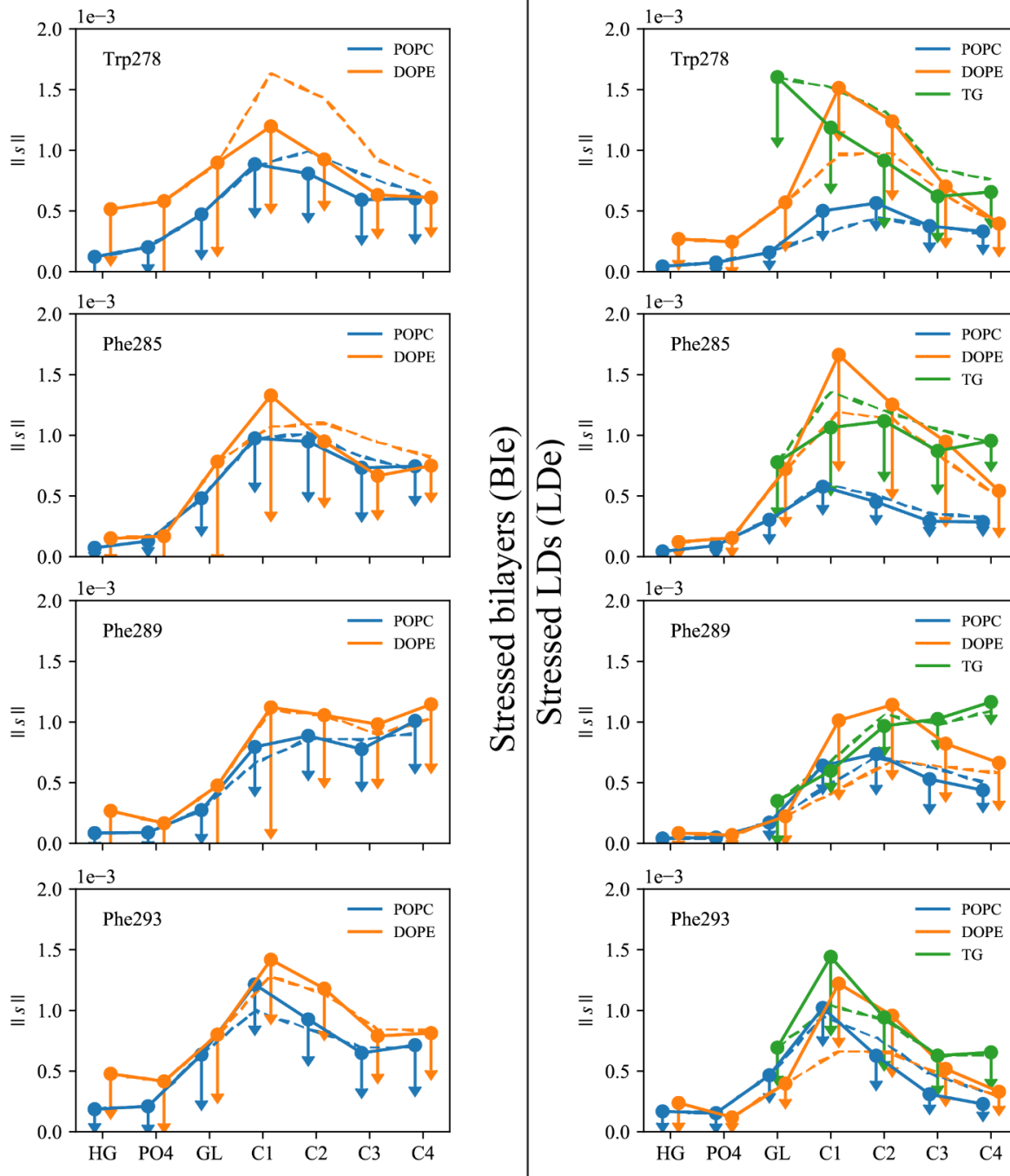
**Figure 5-16** Snapshots and defects of the unstressed bilayer and LD.

The color code is as follows: PL polar groups (light blue), PL acyl chains (dark blue), TG glycerol moieties (green), TG acyl chains (yellow), PL acyl defects (red), TG acyl defects (orange), TG glycerol defects (purple).



**Figure 5-17** Density profile of the bilayer membranes (top) and LDs (bottom).

The average Z position of the PL C2 atoms of the upper leaflet was set to 0. The positive Z is toward water and the negative Z toward membrane center.



**Figure 5-18** Normalized coordination number between tryptophan and phenylalanine residues and each membrane group in stressed bilayers (left) and stressed LDs (right), averaged over the trajectories that have the complete binding.

The error bars represent the standard deviation of the normalized coordination number results from the complete binding trajectories. The *sn-1* chains are shown in filled circles with solid lines while the *sn-2* chains are shown with dashed lines. For visual clarity, only the error bars of the *sn-1* chains are shown and the *sn-3* chains of TG are not shown.



## **Chapter 6 Physical Characterization of Triolein and Implications for Its Role in Lipid Droplet Biogenesis**

This chapter is adapted from [Kim et al., 2021] (69).

### **ABSTRACT**

Lipid droplets (LDs) are neutral lipid storing organelles surrounded by a phospholipid (PL) monolayer. At present, how LDs are formed in the endoplasmic reticulum (ER) bilayer is poorly understood. In this study, we present a revised all-atom (AA) triolein (TG) model, the main constituent of the LD core, and characterize its properties in a bilayer membrane to demonstrate the implications of its behavior in LD biogenesis. In bilayer simulations, TG resides at the surface, adopting PL-like conformations (denoted in this work as SURF-TG). Free energy sampling simulation results estimate the barrier for TG relocating from the bilayer surface to the bilayer center to be  $\sim 2$  kcal/mol in the absence of an oil lens. SURF-TG is able to modulate membrane properties by increasing PL ordering, decreasing bending modulus, and creating local negative curvature. The other neutral lipid, dioleoyl-glycerol (DAG), also reduces the membrane bending modulus and populates the negative curvature regions. A phenomenological coarse-grained (CG) model is also developed to observe larger scale SURF-TG-mediated membrane deformation. The CG simulations confirm that TG nucleates between the bilayer leaflets at a critical concentration when SURF-TG is evenly distributed. However, when one monolayer contains more SURF-TG, the membrane bends toward the other leaflet, followed by TG nucleation if a concentration is higher than the critical. The central conclusion of this study is that SURF-TG is a negative curvature inducer, as well as a membrane modulator. To this end, a model is proposed in which the accumulation of SURF-TG in the luminal leaflet bends the ER bilayer toward the cytosolic side, followed by TG nucleation.

## INTRODUCTION

Lipid droplets (LDs) are energy- and lipid-storing organelles surrounded by a phospholipid (PL) monolayer (20, 21, 25, 200). Enriched with neutral lipids such as triacylglycerol (TG) or sterol esters in their core, LDs are known to be formed from the endoplasmic reticulum (ER). However, little is known about LD biogenesis in spite of its implications for a number of metabolic diseases. The current model posits three distinct steps (44, 47, 199): 1) synthesis of TG by diacylglycerol *o*-acyltransferase (DGAT) at the ER, 2) phase separation of TG at a critical concentration to form nascent LDs, and 3) LD budding toward the cytosol. From nucleation to budding, the lipid droplet assembly complex is known to mediate LD biogenesis. Afterwards, the complex is separated into the ER resident protein seipin and the lipid droplet assembly factor (LDAF1) (38). The absence of this complex results in ectopic LDs; however, how this machinery catalyzes LD formation remains poorly understood (202).

Computer simulations have clearly proven to be highly valuable for the study of biomolecules such as lipid bilayers and proteins at the scale of up to  $\sim 100$  M atoms and  $\sim 1$  milliseconds, largely enabled by rapid advances in hardware, software, enhanced sampling, and supporting theories (52). As such, molecular dynamics (MD) simulations have provided valuable insights on both dynamical and structural properties. For instance, the physical properties of lipid bilayer membranes have been extensively studied involving varying sizes and lipid compositions, as well as lipid-protein interactions at various resolutions (see, e.g., refs (236-240)). In contrast, computational studies of LDs are only just beginning to receive attention; this is partly because the available force fields do not support certain neutral lipids. For this reason, prior computational studies have produced a new TG topology with or without a new parameter set. For instance, the Vanni group has been incorporating the Shinoda-DeVane-Klein (SDK) force field in their LD

studies, which involves the parametrization of the glycerol group to reproduce interfacial tension at the water interface (36, 42, 241, 242). The more common (and easier) approach is to create a new TG topology by replacing the head group of a PL with one of its tails without creating a new parameter set. This approach has been used in all-atom (AA) MD (66-68, 154) and coarse-grained (CG) MD simulations, the latter run with the MARTINI CG force field (43, 89, 243, 244).

Recent 10  $\mu$ s-long AA simulations of LDs report surface-oriented TG (called SURF-TG in this work) (68), which was also implicated in NMR experiments (37). The relative amount of SURF-TG to PLs, which ranges from 5-8% in the unstressed LD (68) to ~20% in the stressed LD (67), has been shown to modulate the area per lipid (APL) and interdigitation with PLs. The display of SURF-TG at the LD surface creates a unique composition and defect type that does not exist in other bilayer-bounded organelles, thereby providing a plausible explanation for LD-specific targeting or conditional targeting by certain proteins. For instance, the amount of SURF-TG increases in expanding LDs to reduce the surface tension (67). The highly amphipathic, autoinhibitory motif of CTP:phosphocholine cytidyltransferase (CCT) is predicted computationally to sense defects created by SURF-TG and preferentially binds to those defects when LDs are expanding (67). Once fully associated, the insertion depth of the Trp278 of CCT $\alpha$  is aligned with the glycerol moiety of SURF-TG to form hydrogen bonds, which may increase the binding affinity (67). Although additional research will be needed to confirm that surface-oriented neutral lipids at the LD surface promotes specificity in LD targeting, this latter CCT binding simulation study reinforces the possible importance of SURF-TG in peptide targeting.

In this paper, the focus is not on investigating SURF-TG as a stress reducer and peptide-targeting mediator but instead on SURF-TG as a membrane property modulator. In particular, the way in which its molecular geometry is implicated in LD biogenesis is discussed. Lipids that have

a conical shape such as phosphatidylethanolamine (PE), diacylglycerol (DAG), or TG do not usually form a lamellar structure by themselves (14, 213, 214). One notable example is 2,3-dioleoyl-D-glycero-1-phosphatidylcholine (DOPC) vs 2,3-dioleoyl-D-glycero-1-phosphatidylethanolamine (DOPE). The only difference between those two molecular structures is the head group, with DOPC having  $-\text{CH}_2-\text{N}(\text{CH}_3)_3^+$  in comparison to the  $-\text{CH}_2-\text{NH}_3^+$  head group of DOPE. Therefore, a single DOPC molecule features a cylindrical shape, while DOPE has a conical shape. This difference in molecular topology leads to two distinct phases when DOPC and DOPE exist in single-component bulk systems: DOPC forms a planar bilayer while DOPE mostly forms an inverse hexagonal phase (110, 111). The physical properties of these two phases have been characterized experimentally (110, 245-247) and computationally (248). Compared to DOPC or DOPE, DAG is more conical because of the absence of the head group, and SURF-TG is even more so due to the addition of the acyl chain. Using AA-MD simulations, we demonstrate here that these conical lipids are able to modulate membrane properties, such as bending modulus, and populate negative curvature regions of the membrane.

The present study also describes the development of a phenomenological CG model for PL and TG to access larger effective simulation time scales, wherein we show that TG nucleates at a critical TG concentration between the leaflets as opposed to the asymmetric distribution of SURF-TG, which induces membrane bending first. Collectively, we demonstrate that SURF-TG is a negative curvature inducer. To this end, we propose a model in which SURF-TG-driven bending helps budding and determines the directionality of LD budding.

## METHODS

### *AA-MD simulations and system setup*

In this study, dioleoyl-glycerol (DOGL) and triolein were used to represent DAG and TG, respectively. A system was prepared using either the CHARMM-GUI membrane builder (55, 56, 220, 221) or PACKMOL (60). The details for each simulation and system are summarized in Table 6-1. The TG molecular structure used for the PACKMOL input was based on a prior LD simulation featuring a trident structure (68). Three different TG models were used (Fig. 6-1). The first one was made based on the CHARMM36 (C36) lipid force field (61) by replacing the head group of DOPE with its sn-1 acyl chain. However, the structural and thermodynamic properties of nonpolar molecules are sensitive to the cutoff distance of Lennard-Jones (LJ) interactions, which was discussed in ref (249). Indeed, the bulk TG simulations, using the 1.2 nm cutoff distance of LJ interactions, which is consistent with C36, resulted in the low density and low surface tension at the vacuum interface (See Fig. 6-2 in the Results). Recently, a LJ cutoff-free version of C36 (C36/LJ-PME) was released (57, 58). The second TG model was made based on the C36/LJ-PME parameters in the same fashion used in the first model. However, both first and second models suffer from the low surface tension at the water interface (Fig. 6-2 in the Results). Therefore, we introduce the revised TG model (C36/LJ-PME-r), which was made based on C36/LJ-PME, however, with the reduced partial charge distribution at the ester group (Fig. 6-1). The new partial charge distribution was determined in an ad-hoc manner to match the surface tension of TG at the water interface. The parameters of the ester group were originally derived to best describe the PL behavior at the water interface, therefore, we hypothesized that the partial charges of the TG ester group should be reduced, which is a non-polar molecule. The suggested charge distribution (Fig. 6-1) was obtained by conducting a series of tests: Adding  $\Delta q$  to the oxygen atoms (O11, O12, O21,

O22, O31, and O32) and subtracting  $2\Delta q$  from the carbon atom (C11, C21, and C31). When  $\Delta q = 0$  (C36/LJ-PME),  $\Delta q = 0.05$ ,  $\Delta q = 0.10$ , and  $\Delta q = 0.20$  (C36/LJ-PME-r), the surface tension was 13.7 mN/m, 16.4 mN/m, 24.9 mN/m, and 32.2 mN/m, respectively. A systematic way to optimize parameters, not only partial charges but also LJ parameters, will be preferred in the future (58). The partial charges of PL were unchanged. Therefore, for the simulations that do not include TG, C36/LJ-PME and C36/LJ-PME-r are the same. All of those TG topologies are available in <https://github.com/ksy141/TG>. Of note, DAG is also not included in the original C36 force field release; nonetheless, the CHARMM-GUI developers (The Im group) created a new topology for DAG based on the existing parameter set.

**Table 6-1** Description of AA simulations.

AA Simulations	Lipid composition <sup>a)</sup>	Time	Ensemble	Replicas	Setup
Bulk TG	216 TG	100 ns	NPT	6 <sup>b)</sup>	PACKMOL
Bulk TG-vacuum	216 TG	200 ns	NVT	6 <sup>b)</sup>	From above sim.
POPC bilayer	128 POPC	200 ns	NPT	1	PACKMOL
POPC + TG bilayer	128 POPC + 1 TG	1 $\mu$ s	NPT	1	PACKMOL
POPC + TG bilayer	128 POPC + 1 TG	300 ns	NPT + biased	12 <sup>c)</sup>	From above sim.
POPC + TG bilayer	128 POPC + 1 TG	2 $\mu$ s	NPT + biased	5	From above sim.
POPC + TG bilayer	124 POPC + 8 TG	1 $\mu$ s	NPT	1	PACKMOL
POPC + TG bilayer	1984 POPC + 128 TG	600 ns	NPT	1	From above sim.
POPC bilayer	4050 POPC	300 ns	NPT	1	CHARMM-GUI
POPC + DAG bilayer (referred to as DAG 10%)	3200 POPC + 320 DAG	1 $\mu$ s	NPT	1	CHARMM-GUI
POPC + DAG bilayer (referred to as DAG 20%)	3200 POPC + 640 DAG	1 $\mu$ s	NPT	1	CHARMM-GUI
POPC + DAG bilayer (referred to as DAG 30%)	3200 POPC + 960 DAG	1 $\mu$ s	NPT	1	CHARMM-GUI
DAG 30% membrane deformation and then NPAT	3200 POPC + 960 DAG	20 ns + 1 $\mu$ s	NP $\gamma$ T and then NPAT	1	From above sim.
DOPC bilayer	128 DOPC	200 ns	NPT	1	CHARMM-GUI

a) The total number of lipid molecules of each type in a system.

b) With varying LJ cutoff distances

c) 12 windows

All simulations were run by GROMACS 2018 and 2020 (54) with the C36 (61), C36/LJ-PME (57, 58), or C36/LJ-PME-r force field. Simulations using C36 used the 1.2 nm LJ cutoff distance with a force-switching function between 1.0 nm to 1.2 nm, unless otherwise noted. All

simulations including DAG were performed with C36. Simulations using C36/LJ-PME and C36/LJ-PME-r used the real-space cutoff distance of 1.0 nm. Biased simulations (described below) were performed with the external plugin, PLUMED2 (v.2.6) (135). Simulations were evolved with a 2-fs timestep. The long-range electrostatic interactions were evaluated with the Particle Mesh Ewald algorithm (119), with the distance cutoff set to be the same as the LJ interaction. Every bond involving a hydrogen atom was constrained using the LINCS algorithm (223). The Nose-Hoover thermostat was used to maintain a temperature of 300 K (simulations involving DAG) or 310 K with a coupling time constant of 1 ps (225, 226). For the constant pressure simulations, the Parrinello-Rahman barostat was used to maintain a pressure of 1 bar with a compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$  and a coupling time constant of 5 ps (120). For membrane simulations, the pressure was coupled semi-isotropically, whereas during bulk simulations the pressure was coupled isotropically. For constant area simulations (NPAT), only the pressure of the Z dimension was controlled. In order to deform the DAG 30% membrane, a constant surface tension simulation (NP $\gamma$ T) was conducted for 20 ns with the Berendsen thermostat and barostat using the same parameters (227). The pressure was coupled semi-isotropically with the pressure in the X and Y dimension set to -100 bar during membrane deformation.

*Density, isothermal compressibility, and surface tension*

Density was obtained with the GROMACS analysis tool, which was outputted every 2 ps.

Isothermal compressibility was evaluated from

$$\beta = -\frac{1}{V} \left( \frac{\partial V}{\partial p} \right)_T = \frac{1}{k_B T} \frac{\langle V^2 \rangle - \langle V \rangle^2}{\langle V \rangle}$$

where the bracket represents the ensemble average and  $V$  denotes the volume of the system. Similarly, GROMACS was used to analyze the surface tension at the vacuum and water interfaces. Bilayer area compressibility was determined using the following equation

$$K_A = k_B T \frac{\langle A \rangle}{\langle A^2 \rangle - \langle A \rangle^2}$$

where  $A$  is the instantaneous total area of a membrane (250).

### *Order parameters*

Order parameters were evaluated from  $S_{CD} = 0.5 \times |\langle 3 \cos^2 \theta - 1 \rangle|$ , where the angle ( $\theta$ ) is between the Z axis and the position vector of an acyl carbon atom to a bonded hydrogen atom; the bracketed portion indicates the ensemble average (251-253).

### *Potential of mean force of TG flip-flop*

In order to estimate the preferential location of TG in a bilayer and its tendency to become SURF-TG, the potential of mean force (PMF) of the TG as a function of its Z position in the bilayer was calculated. A 3-palmitoyl-2-oleoyl-D-glycero-1-Phosphatidylcholine (POPC) bilayer membrane consisting of 128 POPC molecules and one TG molecule was prepared. Biased simulations were performed with the collective variable representing the Z distance between the center of the mass of phosphorus atoms and the center of the mass of the glycerol moiety of the TG molecule. Replica-exchange (131) umbrella sampling (132) (REUS) simulations were run for 300 ns with a harmonic restraint of a force constant of 47.8 kcal/mol/nm<sup>2</sup> for each window. The total of 12 windows was prepared with a 0.2 nm spacing over a range of 0.0 to 2.2 nm (0 is the average Z position of the



center of the mass of the phosphorus atoms, which is comparable to the center of the membrane). An initial structure for each window was prepared with a steered MD simulation. The exchange between windows was attempted every 2,000 steps. The weighted histogram analysis method (WHAM) was used for the calculation of the PMF with a bin spacing of 0.0275 nm (134). The Grossfield's WHAM implementation was used (230). The simulations were divided into five equal-length blocks to estimate the errors. The PMF was calculated for each block and then the standard deviation of block PMFs was reported as the errors. To support the PMF result, we also carried out efficient, fast-converging transition-tempered metadynamics (TT-MetaD) simulations with the same collective variable (254, 255). The Gaussian hill was deposited every 5K steps at a height of 0.29 kcal/mol and width of 0.08 nm or 0.1 nm. The bias factor of 6 and threshold of 2.4 kcal/mol for TT-MetaD were used. The position of the two basins were set to -1.1 nm and +1.1 nm, when the center of the mass of phosphorus was set to zero. Five simulation replicas, each 2  $\mu$ s long, were performed to estimate the averaged potential of mean force (PMF). The errors represent the standard deviation of the PMFs obtained from the five replicas.

### *CG simulations*

In order to study SURF-TG-driven membrane deformation and nucleation, a new phenomenological implicit solvent CG model was developed for the PL and TG. Each PL and TG molecule consisted of 11 and 13 CG beads, respectively. Each acyl chain was composed of 4 CG beads, with 2 different CG types. The approximate molecular groupings of the PL and TG structures are shown in Fig. 6-12. In this model, the attraction was modeled with the Gaussian function,  $-A \exp(-Br^2)$  where  $B = 2 \text{ nm}^{-2}$ , and the repulsion with the repulsive LJ interaction,  $4\epsilon (\sigma/r)^{12}$  where  $\epsilon = 0.0028 \text{ kcal/mol}$ . No CG bead carried any charge. The attraction only

occurs between hydrophobic CG beads (beads in the acyl chain or CT1 and CT2 in Fig. 6-12 of the Supporting Information), between the glycerol CG beads of TG, and between the glycerol CG beads of TG and PL (No attraction exists between the glycerol CG beads of PL); in contrast, repulsion occurs between all CG beads unless the pair is bonded. The attraction and repulsion parameters,  $A$  and  $\sigma$ , were primarily parameterized to reproduce the radial distribution function (RDF) from the mapped AA (C36) RDF (Figs. 6-13 and 6-14). Unless otherwise indicated, the repulsion parameter ( $\sigma$ ) of a pair composed of two different types of CG beads was determined using an arithmetic average, while the corresponding attraction parameter ( $A$ ) for such a pair was set explicitly. All CG bonds were modeled with the harmonic potential, with the equilibrium distance of 0.5 nm and the force constant of 500 kcal/mol/nm<sup>2</sup>. The angle potentials were modeled with the harmonic potential with a small force constant of 0.12 kcal/mol/rad<sup>2</sup>. However, we note that the properties of PL and TG in general were not sensitive to the bond and angle force constants. The details of this CG model can be found in Fig. 6-12. The potentials and inputs are available at [https://github.com/ksy141/SK\\_CGFF.git](https://github.com/ksy141/SK_CGFF.git).

All CG simulations were conducted with GROMACS 2018 using a 20-fs timestep (54). The cutoff distance of 2.4 nm was used for non-bonded interactions. Temperature was controlled via velocity rescaling with a stochastic term with the target temperature of 310 K and a time constant of 1 ps (256). A pressure of 1 bar was maintained with the Parrinello-Rahman barostat with a compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$  and a coupling time constant of 10 ps. For the PL or PL+TG bilayer simulation, the pressure was semi-isotropically coupled with the X and Y dimensions but was not coupled with the Z dimension (by setting compressibility in the Z dimension to 0). For the bulk TG simulation, the pressure was isotropically coupled. For bilayer simulations that have asymmetric distribution of TG between two leaflets, weak pressure and

temperature couplings were used to minimize perturbations of the evolution of the membrane remodeling. For those simulations that convey membrane deformation, time constants of 1 ns and 10 ns for temperature and pressure were used, respectively. Note that a previously reported implicit solvent CG model that detected membrane deformation used weak couplings as well (257). Finally, the TT-MetaD simulations were carried out to obtain the TG flip-flop PMF. The Z position of the TGL atom respect to the membrane center was biased with the same TT-MetaD parameters used in the AA simulations. The initial structures were prepared using PACKMOL (60) and MDAnalysis (62); simulation details are provided in Table 6-2.

**Table 6-2** Description of CG simulations.

CG Simulations	Lipid composition <sup>a)</sup>	Nsteps	Ensemble	Replicas	To compute
Bulk TG	1000 TG	5M	NPT	1	$\rho$ , $\beta$ , RDF
Bulk TG-water <sup>b)</sup>	1000 TG	1.5M	NVT	1	$\gamma_{\text{wat}}$
PL bilayer	200 PL	1.5M	NPT	1	$K_A$ , APL, RDF
PL bilayer	7200 PL	5M	NPT	1	$K_C$
PL + TG bilayer	128 PL + 1 TG	10M	NPT + biased	3	PMF
PL + TG bilayer (2%) <sup>c)</sup>	1764 PL + 36 TG	50M	NPT	3	Nucleation
PL + TG bilayer (5%) <sup>c)</sup>	1710 PL + 90 TG	50M	NPT	3	Nucleation
PL + TG bilayer (1%) <sup>d)</sup>	7128 PL + 72 TG	5M	NPT <sup>e)</sup>	2	Bending
PL + TG bilayer (5%) <sup>d)</sup>	6840 PL + 360 TG	5M	NPT <sup>e)</sup>	2	Bending

a) The total number of lipid molecules of each type in a system.

b) For implicit CG models, it is ambiguous to distinguish water and vacuum interfaces. Specific to TG, the surface tensions at those two interfaces happen to be comparable (See Fig. 6-2).

c) Symmetric SURF-TG distribution

d) Asymmetric SURF-TG distribution

e) Weak temperature and pressure couplings

### *SDK CG simulations*

To strength the findings and conclusions, CG simulations with the SDK force field (SPICA force field) (258-260) were carried out by LAMMPS (261). Two systems were prepared: One TG molecule in a bilayer containing 128 POPC molecules to obtain the TG flip-flop PMF and a POPC + DOPE bilayer immersed in water (or a nanodisc) to show bending. Developed by the Vanni

group, the TG force field that reproduces the surface tension at the water interface was used in the first system (36). In the second system, asymmetric distribution of PLs was used with the upper leaflet having 576 POPC molecules and the lower leaflet having 522 POPC and 54 DOPE molecules. Simulations were evolved with a 10-fs timestep. The Nose-Hoover thermostat (225) and barostat (120, 262, 263) were used to maintain a temperature of 310 K (first system) or 300 K (second system) and a pressure of 1 bar with damping parameters of 20 ps and 50 ps, respectively. The cutoff distance of 1.5 nm was used, and the long-range electrostatic interactions were evaluated with the particle-particle-particle-mesh (PPPM) solver. The force error of  $10^{-5}$  kcal/mol/Å, the third order, and the grid size of 20 Å in each dimension were used for the long-range electrostatic interactions.

### *Analysis and visualization*

Simulations were analyzed using MDAnalysis (62) and GROMACS (54). Unless otherwise noted, the standard errors (se) were reported in this work. Equilibrated trajectories were first divided into M blocks of equal length and the block average for each block was calculated. The standard errors (se) were estimated by  $se = 1/\sqrt{M} \text{ } sd$ , where sd is the standard deviation of the block averages. Five blocks (M = 5) were used in this work. Molecular images were captured using PyMOL.

## **RESULTS**

### *Physical properties of bulk TG*

In this section, we establish the quality of three TG models (C36, C36/LJ-PME, and C36/LJ-PME-r; see Fig. 6-1 and Methods) by characterizing four physical properties of bulk TG (Fig. 6-2):

density ( $\rho$ ), isothermal compressibility ( $\beta$ ), surface tension at the vacuum ( $\gamma_{\text{vac}}$ ), and surface tension at the water interface ( $\gamma_{\text{wat}}$ ). With a prior study demonstrating that the first three properties are sensitive to the LJ cutoff distances (249), we performed MD simulations (C36) of a bulk TG consisting of 216 TG molecules with varying LJ cutoff distances. The computed density and isothermal compressibility of the bulk TG were shown to converge when the LJ cutoff distance was increased from 1.0 nm to 2.0 nm (Fig. 6-2). The density of the bulk TG approached the experimental value (0.8991 g/cm<sup>3</sup> at 313 K) (116) when simulated with the LJ cutoff of 2.0 nm (0.898 g/cm<sup>3</sup>). The same calculations were carried out with C36/LJ-PME and C36/LJ-PME-r. The extended LJ interactions caused a higher density in the C36/LJ-PME results, however, the extra increase became compensated by the reduced charge distribution in C36/LJ-PME-r. The same behavior was observed in the RDF of the TG glycerol moiety (Fig. 6-15). While C36/LJ-PME increases the first peak due to the increased LJ range, the reduced charge distribution in C36/LJ-PME-r lowers the increased peak. Next, we performed MD simulations of the TG-vacuum interface by adding an empty space in the Z dimension to the bulk TG system. Consistent with density and isothermal compressibility, the surface tension of the TG-vacuum interface converges with the increased LJ cutoff distance and also agrees reasonably well with the results of C36/LJ-PME and C36/LJ-PME-r. In all cases, no TG molecules were released from the bulk TG. Although there are no experimental TG data for  $\beta$  and  $\gamma_{\text{vac}}$ , the results of C36 with the 2.0 nm cutoff distance, C36/LJ-PME, and C36/LJ-PME-r show the reasonable agreement for  $\rho$ ,  $\beta$ , and  $\gamma_{\text{vac}}$ .

*TG flip-flop*

We characterized the physical properties of TG in a bilayer membrane in order to investigate its implication in LD biogenesis. A bilayer membrane containing 128 POPC molecules and 1 TG molecule was prepared. We then calculated the PMF of the TG molecule (Fig. 6-3) as a function of its Z position within a bilayer membrane by performing the REUS simulations. Both C36 and C36/LJ-PME-r demonstrate the stability of SURF-TG in the absence of an oil lens although the C36 result overestimates the stability at the surface by  $\sim 1.5$  kcal/mol compared to the C36/LJPME-r result. The preferential location of the TG molecule is slightly different by 0.2 nm as well. We further verified our results with another biased methodology. Five replicas of TT-MetaD simulations were carried out with the same collective variable, each run for 2  $\mu$ s. The PMFs obtained from the REUS and TT-MetaD simulations show good agreement (Fig. 6-16). Finally, the 1- $\mu$ s unbiased MD simulation behaves as consistent with our findings. For the C36 simulation, the TG molecule was initially located at the center of the bilayer (CORE-TG), however, became SURF-TG rapidly ( $< 10$  ns). During 1  $\mu$ s, the TG molecule did not flip-flop but resided at the same leaflet. In contrast, the TG molecule visited the membrane center more frequently and eventually flip-flop at  $\sim 820$  ns with C36/LJ-PME-r. Taken together, both unbiased and biased results suggest that a single TG molecule is principally SURF-TG in a bilayer membrane, adopting the PL-like conformation, in the absence of an oil lens.

#### *Local membrane deformation by SURF-TG*

Conical molecules such as PE or DAG are known to be responsible for amphipathic peptide targeting by creating packing defects.(6, 126, 219, 264, 265) However, the precise mechanisms for how those molecules modulate membrane properties are poorly understood. The molecular shape of SURF-TG is inevitably conical, with the glycerol moiety forming the vertex of the cone

and the end group of three acyl chains forming a flat base while POPC features a quasi-cylindrical shape. In order to study how SURF-TG modulates the local properties of bilayers, we computed the height field of the phosphorus atoms of the upper leaflet in the bilayer containing 128 POPC molecules and 1 TG molecule. Histogram analysis was used for the 1- $\mu$ s C36/LJ-PME-r simulation to investigate the relative positions of the phosphorus atoms of the upper leaflet with respect to the glycerol moiety of the SURF-TG molecule (Fig. 6-4). Approximately one-fifth of the frames were not used for this analysis because in those frames SURF-TG transitioned to CORE-TG or resided at the other leaflet (flip-flop occurred at  $\sim$ 820 ns). The height of the phosphorus atoms near the SURF-TG molecule can be as low as  $\sim$ 0.15 nm compared to those of the other phosphorus atoms (Fig. 6-4). The range of the local deformation created by one SURF-TG molecule can be more than 2 nm in both the X and Y dimensions (Fig. 6-4). We anticipate that if SURF-TG molecules cluster at the surface there will be stronger and longer PL packing discontinuity. We confirmed that the same conclusion can be made with the C36 trajectory (data not shown). Our data indicate that SURF-TG locally pulls PLs toward the bilayer center, thereby creating local negative curvature. How this process serves to remodel bilayer membranes will be discussed later via our CG simulation results.

#### *Molecular and hydration properties of SURF-TG*

In order to characterize the molecular and hydration properties of SURF-TG, a new POPC bilayer membrane containing 6% TG (62 POPC and 4 TG molecules per leaflet) was prepared and compared with a pure POPC bilayer membrane (64 POPC molecules per leaflet). The system contained a higher concentration of TG than the critical nucleation concentration, which was reported to be 2.8% g TG or 2.4% mol TG.(37) However, within the AA simulation timescale, TG

nucleation did not occur, and all the TG molecules remained SURF-TG most of the times. The simulations were carried out with C36/LJ-PME-r. For a pure POPC bilayer membrane, C36/LJ-PME and C36/LJ-PME-r are equivalent.

The order parameters of SURF-TG and POPC were computed (Fig. 6-5). Interestingly, the order parameters of POPC in the POPC+TG bilayer (continuous lines) were found to be larger than those of POPC in the pure POPC bilayer (dashed lines), suggesting that SURF-TG increases the order of POPC molecules. As reported in a previous study, the increase in PL ordering in LDs compared to the bilayers can be attributed to the fact that CORE-TG interdigitates with the PL monolayer and increases the density of the low-density tail region.<sup>(68)</sup> Similarly, short LD simulations lacking any SURF-TG also demonstrated increased PL ordering.<sup>(99)</sup> Our results suggest that SURF-TG increased PL ordering as well. The SURF-TG order parameters have the same trend with PL; however, they are more reduced than PL. The reduced order parameters of TG compared to PL can be attributable to the higher degree of freedom in the Z dimension of TG molecules. While PL moves little in the Z dimension and therefore its order parameters are usually determined by the lipid-packing density in the XY dimensions, TG is relatively free in moving in the Z dimension.

In the same system, we characterized the orientation of each TG molecule by calculating the angle between the Z-axis and the positional vector of the center of the mass of the TG glycerol moiety from the center of the mass of the TG acyl chains. The angle as a function of the Z position is shown in Figure 6-17. Consistent with the TG flip-flop PMF (Figure 6-3), TG molecules principally reside at the membrane surface ( $z = \sim 1.2$  nm). Also, from the orientation analysis, we were able to confirm that TG adopts PL-like conformation in which the glycerol moiety is exposed to the membrane surface and its acyl chains are extended toward the membrane center.



The hydration properties of SURF-TG were evaluated in the same POPC+TG bilayer. It was suggested that the SURF-TG has the primary carbonyls in sn-1 and sn-3 chains that are more exposed to water than the secondary carbonyl in sn-2 chain, which would explain specificity for hydrolysis at the primary carbonyl position.(37) The computed RDF between each TG oxygen atom (O11, O21, and O31) and water shows reduced accessibility of water to the secondary carbonyl (Fig. 6-6). However, the RDFs between the other oxygen atoms (O12, O22, O32) and water were identical at the first peak (data not shown).

### *Impact of neutral lipids on membrane properties*

In order to systematically study the impact of neutral lipids on bending modulus ( $K_C$ ), we assembled large POPC bilayer membranes of varying DAG concentrations ranging from 0% to 30% and computed the bending modulus using the undulation spectrum.(266-268) On our simulation timescale (1  $\mu$ s), there was no evidence for phase separation in the DAG, despite its high concentration. With an increased concentration of DAG, the bending modulus decreased (Fig. 6-7), suggesting that DAG lowers bending rigidity. This agrees with the previous simulation showing the decreased bending modulus of a bilayer membrane containing DAG.(125)

We also confirmed a similar correlation for the bilayer containing 6% SURF-TG with both C36 and C36/LJ-PME-r (Fig. 6-7). The bending modulus of this particular bilayer was determined to be  $23.1 \pm 1.0 k_B T$  (C36) and  $22.4 \pm 2.5 k_B T$  (C36/LJ-PME-r), which is smaller than  $31.7 k_B T$  (ref(237)) or  $31.1 \pm 2.3 k_B T$  (Fig. 6-7), the bending modulus of the pure POPC bilayer. The results are supported by the recent experiments showing the reduced bending modulus of a bilayer membrane containing TG.(269)

Next, constant surface tension was applied semi-isotropically for 20 ns to induce deformation of the 30% DAG membrane. The resulting structure displayed a buckling that resembled a cubic function in the X axis (Fig. 6-8a). We then ran constant area simulation (NPAT) for 1  $\mu$ s, confirming that the overall buckling was maintained (Figs. 6-8a and 6-8b). As shown in Fig. 6-8b, the height field region in the upper leaflet where X is between 5 nm and 15 nm displayed a negative curvature; conversely, the region where X is between 25 nm and 30 nm displayed a positive curvature. In contrast, the curvature becomes inverted in the lower leaflet (Fig. 6-18). Consistent with our expectations, a conical DAG is less populated at the positive curvature (for instance, the region where X is between 25 nm and 30 nm in the upper leaflet), but more populated at the negative curvature (Fig. 6-8c and Fig. 6-18). The central conclusion derived from these simulations is that a neutral lipid reduces the bilayer bending modulus and behaves as a negative curvature inducer.

### *TG bulk and PL bilayer*

In order to further investigate the properties of SURF-TG in a bilayer membrane and establish their implications in LD biogenesis, we developed a new implicit solvent CG model for PL and TG as noted earlier. In this model, each PL and TG molecule consisted of 11 and 13 CG atoms, respectively, where each acyl chain was composed of 4 CG atoms and of 2 CG types. No charges were associated with CG beads. The interactions were modeled with a Gaussian function for attraction, and the LJ repulsive potential for repulsion. The details of the CG model can be found in Fig. 6-12.

The attraction and repulsion parameters were then primarily parameterized to reproduce the RDFs from the AA trajectories with the atoms mapped on the CG sites of the CG model (the so-called “mapped AA model”). A PL bilayer simulation consisting of 200 residues was performed for 1.5M steps, after which the computed 2-dimensional RDFs were compared with those from the mapped AA trajectory provided in Fig. 6-13. Similarly, a bulk TG simulation consisting of 1,000 residues was conducted for 5M steps; the computed RDFs from this trajectory were compared with those from the mapped AA trajectory shown in Fig. 6-14. The results showed reasonable agreement between the CG RDFs and the mapped AA RDFs, which were achieved both in the PL bilayer and bulk TG systems. We then determined the physical properties of the bulk TG and PL bilayer from the same simulations as shown in Table 6-3. Although the isothermal compressibility of the bulk TG obtained from the CG simulation was found to be a factor of three larger, and the bending modulus of the PL bilayer was recorded to be 1.4 times higher than the AA results, the remaining properties agreed well with the experimental or AA data.

**Table 6-3** Physical properties of DOPC bilayer and TG bulk from experiments, AA simulations, and CG simulations. Standard errors are given for the simulations that we performed.

	Exp.	AA	CG
DOPC bilayer			
APL [ $nm^2$ ]	0.674 <sup>a</sup> , 0.724 <sup>b</sup>	$0.70 \pm 0.00$	$0.72 \pm 0.00$
$K_A$ [ $mN/m$ ]	300 <sup>c</sup>	$251 \pm 20, 290^d$	$307 \pm 9$
$K_C$ [ $k_B T$ ]	21.2 <sup>e</sup>	29 <sup>d</sup>	$41 \pm 1$
TG bulk			
$\rho$ [ $g/cm^3$ ]	0.8991 <sup>f</sup>	$0.90 \pm 0.00^g$	$0.90 \pm 0.00$
$\beta \times 10^{-9}$ [ $m^2/N$ ]	N/A	$0.88 \pm 0.01^g$	$2.33 \pm 0.07$
$\gamma_{wat}$ [ $mN/m$ ] <sup>h</sup>	32 <sup>i</sup>	$32 \pm 0^g$	$38 \pm 0$

a) ref (270) b) ref (271) c) ref (272) d) ref<sup>(237)</sup> e) ref (273) f) ref (116) g) Taken from the C36/LJ-PME-r simulations. h) As it is an implicit CG model (no explicit solvent molecules), it is ambiguous to distinguish surface tension at the air interface and water interface. Specific to TG, those two values are comparable (see Fig. 6-2) i) ref (33)

### TG Nucleation

TG nucleation between ER leaflets represents the first step in LD biogenesis. A previous NMR study determined the maximum solubility of TG in a PL phase to be 2.8% g TG or 2.4% mol TG (37). Therefore, we expected that a bilayer containing TG at a higher than critical concentration would undergo TG nucleation. Accordingly, we prepared CG bilayer systems with varying TG concentrations to study TG nucleation. Consistent with our expectations, our CG simulations confirmed that TG nucleates and forms an oil lens between the leaflets within 20M time steps when the TG concentration is 5% but does not if it is 2% (Fig. 6-9). In order to estimate the degree of nucleation, the nucleation % was calculated by dividing the number of TG molecules in the

biggest cluster by the number of total TG molecules. If the glycerol moieties of two TG molecules are within 2 nm, they were assumed to be in the same cluster. Results for the 5% simulation indicate that the oil cluster is stable and recruits more TG with increasing simulation time until equilibrium is reached (Fig. 6-9). Importantly, TG molecules that are not in the oil cluster but dissolved in the membrane are primarily SURF-TG. This can be also seen in the 2% simulation where most of the TG molecules resided at the surface (Fig. 6-9).

In order to show the surface propensity of a TG molecule in our force field and to evaluate the quality of the TG model, the TG flip-flop PMF was calculated with the TT-MetaD simulations (Fig. 6-19). Although there is a disagreement between the AA and CG results at the membrane center, the PMF suggests that TG primarily resides at the membrane surface in the absence of an oil lens.

### *Conical lipid-mediated membrane bending*

As confirmed by our AA simulation results, SURF-TG locally induces a negative curvature. Using CG simulations, we investigated how local membrane deformation leads to global membrane deformation. First, we prepared an asymmetric bilayer wherein each leaflet contained a different number of SURF-TG molecules. For instance, in one of our CG simulations, the upper leaflet of the bilayer contained 3552 PL and 48 SURF-TG molecules, while the lower leaflet contained 3576 PL and 24 SURF-TG molecules. The biological rationale for the asymmetric distribution of SURF-TG between two leaflets will be discussed in the Discussion section later. Initially from a flat bilayer, the bilayer then rapidly (at 20K time steps) bent toward the lower leaflet containing less SURF-TG than the upper leaflet (Fig. 6-10). Interestingly, one of the regions in the upper leaflet

where the local concentration of SURF-TG was initially high ( $x = 0.6$  and  $y = 0.1$  in Fig. 6-10a) became the lowest point, as indicated in Fig. 6-10b and Fig. 6-10c. The other region that initially had the high local concentration of SURF-TG in the upper leaflet ( $x = 0.5$  and  $y = 0.9$  in Fig. 6-10a) became curved to the lower leaflet. In this particular simulation, TG nucleation did not occur because of the low TG concentration. Consistent with the AA results, TG remained SURF-TG, populated at the negative curvature during 5M steps (data not shown). The same simulation but containing the higher TG concentration demonstrates the same bending behavior, followed by nucleation.

To show the other example of conical lipid-mediated membrane bending, we prepared a bilayer immersed in water (nanodisc) and carried out CG simulation with the SDK force field (Figure 6-11) (258-260). Initially, the upper leaflet contains 576 POPC molecules and the lower leaflet contains 522 POPC and 54 DOPE molecules. After 20 ns (2M steps), the membrane becomes bent toward the upper leaflet and the bent structure was maintained toward the end of the simulation (28.3 ns). We expect the nanodisc will become flat eventually once the PL distribution becomes equal.

Based on the above results, we expect the timescale of membrane bending is faster than that of TG nucleation. In our CG simulations, membranes containing 5% TG undergo bending first, followed by TG nucleation. In the nanodisc simulation, membrane bending due to asymmetric distribution of conical lipids happen at the relatively fast timescale (20 ns). One may get the comparable insight via AA simulation results. In AA simulations, membrane undulation occurs within several hundred nanoseconds, thereby enabling us to evaluate the bending modulus from the undulation spectrum. In comparison, the bilayer containing 5% TG did not undergo TG nucleation within the AA timescale (microsecond). Therefore, both AA and CG simulations

suggest the timescale of TG nucleation is slower than that of membrane undulation. Taken together, SURF-TG will be important in the LD biogenesis because it can modulate the membrane properties before transitioning into an oil lens. Especially, membrane deformation driven by SURF-TG, which would precede TG nucleation, may determine the LD budding directionality.

## **DISCUSSION**

Using a revised AA-TG model that reproduces the experimental surface tension at the water interface, we have examined herein how a conical lipid modulates membrane properties and induces membrane remodeling. The AA simulations demonstrate that SURF-TG decreases bending rigidity and increases PL ordering. Due to its conical shape, SURF-TG induces a local negative curvature. We also confirmed using biased simulations that the energy barrier for relocating SURF-TG to the bilayer center is  $\sim 2$  kcal/mol in the absence of an oil lens. This finding is consistent with our experience that TG initially locates at the bilayer center, while becoming SURF-TG within 10 ns of unbiased MD simulations and it resides principally at the surface. Finally, we systematically demonstrated the impact of neutral lipids on the membrane properties with varying DAG concentrations by determining that neutral lipids populate and induce a negative curvature and reduce the bending modulus.

In order to access SURF-TG-mediated membrane remodeling, we developed a phenomenological implicit solvent CG model for PL and TG using a Gaussian function for attraction and the LJ potential for repulsion. Primarily parameterized to reproduce the RDFs from the atomistic CG mapped trajectories, the CG model was able to reproduce most of the physical properties, with the exception of the isothermal compressibility of TG, which was found to be a factor of three higher; additionally, the bilayer bending modulus was noted to be 1.4 times higher

compared with our AA results. Using our phenomenological CG model, we showed TG forms an oil lens between the leaflets when the TG concentration is 5% but does not if it is 2%. We also demonstrated that asymmetric SURF-TG composition induces membrane bending because SURF-TG works as a negative curvature inducer. For instance, when the upper leaflet of a bilayer contained more SURF-TG, the membrane would bend toward the lower direction. Consistent with the AA results, TG molecules that do not belong to the oil cluster are mostly SURF-TG.

Although a related hypothesis suggesting that asymmetric surface tension determines the directionality of LD budding has recently been proposed (274, 275), the cause of any asymmetric surface tension was not linked to lipid type or lipid geometry. A possible physiological explanation for the asymmetric distribution of SURF-TG between the luminal leaflet and cytosolic leaflet of the ER bilayer can be found in the recently resolved structure of DGAT1 (276, 277). Catalyzing the last step of TG synthesis, DGAT1 intakes DAG and outputs TG. The lateral gate, which is located closer to the luminal leaflet, is anticipated to be a pathway for the flow of the reactant and product. Therefore, we can propose a model in which a newly synthesized TG molecule bends the membrane toward the cytosol and determines the budding directionality based on the following steps: 1) The synthesized TG molecule is released through the lateral gate of DGAT1. 2) The released TG molecule becomes SURF-TG since it has an energy penalty of  $\sim 2$  kcal/mol when residing at the ER bilayer center. 3) Given that the lateral gate is closer to the ER luminal leaflet, it is more likely that the luminal leaflet will contain more SURF-TG than the cytosolic leaflet. 4) The accumulation of SURF-TG in the luminal leaflet bends the ER bilayer toward the cytosolic side. 5) The curved ER bilayer recruits the lipid droplet assembly complex (278).

This report makes reference to a number of papers that have described comparable results, although using different lipids and in different contexts (273, 279-283). In those studies, the



amount of polyunsaturated fatty acids or polyunsaturated PLs were correlated with bending modulus and the distribution of those lipids. The central conclusion from those studies is that polyunsaturated acyl chains add fluidity in the Z dimension, which implies that they can be more flexibly adapted for curvature than saturated acyl chains, thereby reducing bending modulus, facilitating lipid trafficking, and modulating membrane dynamics. In this paper, TG and DAG were modeled with triolein and dioleoyl-glycerol, wherein each chain was mono-unsaturated. Therefore, we suggest that our findings are more related to the molecular shape rather than to the degree of unsaturation. Nonetheless, a future study should be designed to investigate the interplay between those two factors.

Finally, we discuss here our future work on CG modelling. In this study, a phenomenological CG model was developed via hand-tuning. A follow-on study may incorporate a bottom-up approach that systematically constructs CG models based on underlying AA interactions. Such a methodology will include the force-matching method (284-288), relative entropy minimization (289), or a hybrid approach that utilizes both methods (290). Using a highly coarse-grained membrane, large-scale membrane deformations mediated by the Bin/amphiphysin/Rvs (BAR) domain-containing proteins have been successfully described with such methods (291-295). Also, one interesting potential direction would be to introduce internal “states” to lipid CG beads using the ultra-coarse-graining (UCG) theory in order to modulate the CG interactions in different chemical or physical environments (296-298). In particular, by assuming that the internal state dynamics remain in quasi-equilibrium (299-301), we expect that the UCG modeling can further capture different TG solubilities at PL bilayers and LD surfaces.

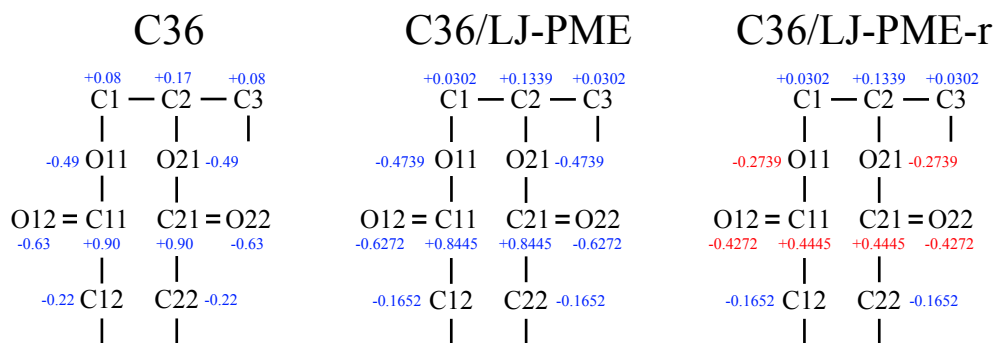
## CONCLUSIONS

In this study we show how neutral lipids are able to modulate the physical properties of bilayers. Results from our AA-MD bilayer simulations indicate that TG principally resides at the bilayer surface, adopting PL-like conformations, in the absence of an oil lens. SURF-TG is an innately conical lipid because of the addition of the acyl chain and the elimination of the head group compared to POPC; as such, a conical SURF-TG produces a local negative curvature and lowers bending modulus. We also find a conical molecule, DAG, populates the negative curvature and lowers bending modulus. In order to increase the accessible simulation time scale, a phenomenological CG model for PL and TG was developed by parameterizing the non-bonded interactions to reproduce the RDFs from the mapped atomistic trajectories and physical properties. In the CG simulations, TG molecules form an oil cluster when TG concentration is above the critical concentration. Our CG simulations of the bilayers, wherein each monolayer surface contained a different number of SURF-TG, confirmed SURF-TG-driven membrane deformation, which may determine the directionality of LD budding and catalyze TG nucleation. Consistent with the AA results, TG molecules that are not in the oil cluster but dissolved in the PL phase are mostly SURF-TG and populate the negative curvature. To conclude, this paper demonstrates how the conical shape of neutral lipids are implicated in the early step of LD biogenesis.

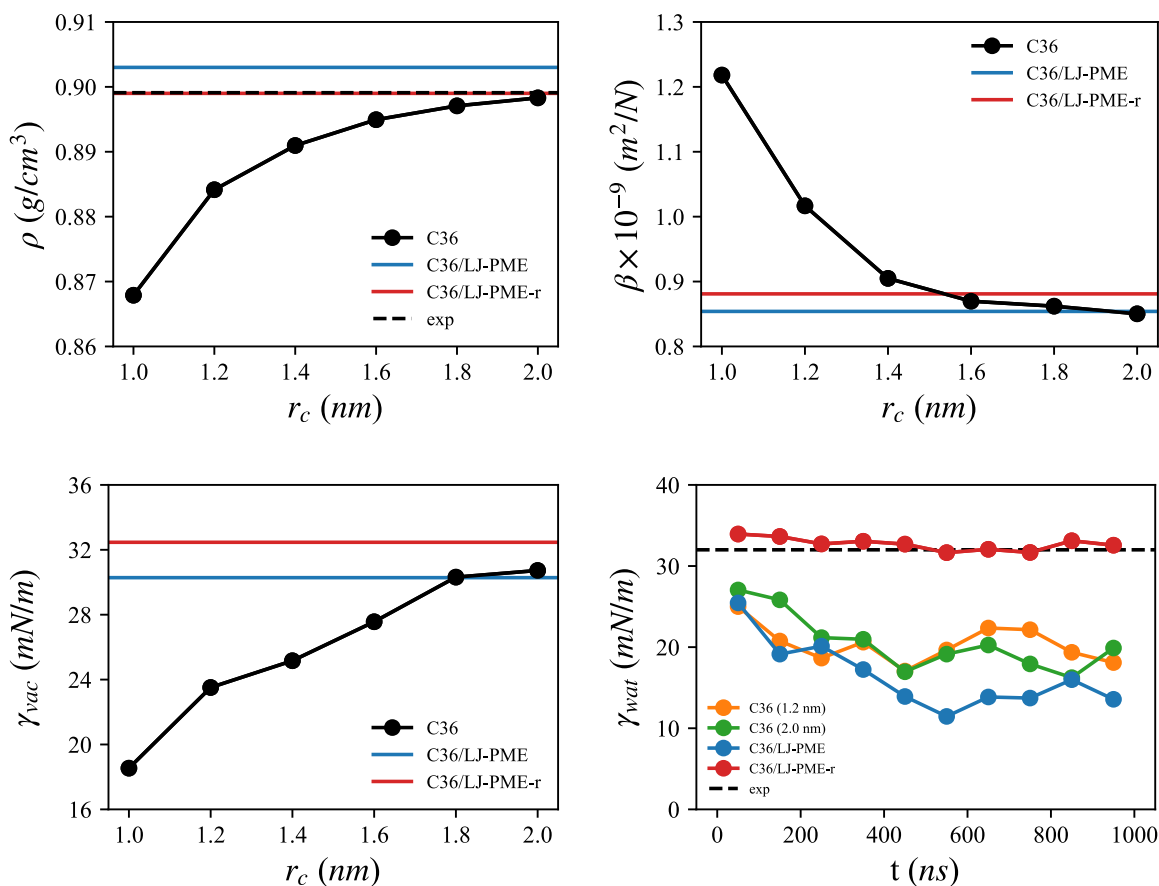
## ACKNOWLEDGEMENTS

This research was supported by a grant from the National Institute of General Medical Science (NIGMS) of the National Institutes of Health (NIH), grant R01-GM063796. The computer simulations were performed on the high-performance GPU cluster (GM4) at the University of Chicago Research Computing Center, supported by NSF grant DMR-1828629 and the Stampede2

supercomputer at the Texas Advanced Computing Center (TACC) through allocation TG-MCA94P017 with resources provided by the Extreme Science and Engineering Discovery Environment (XSEDE) supported by NSF grant ACI-1548562. We thank Chenghan Li, Jeeyun Chung, Zack Jarin, Myong In Oh, Jessica Swanson, Robert Farese, Jr., Tobias Walther, and Richard Pastor for their useful perspectives and discussions. We thank Stefano Vanni for providing his SDK force field of neutral lipids.

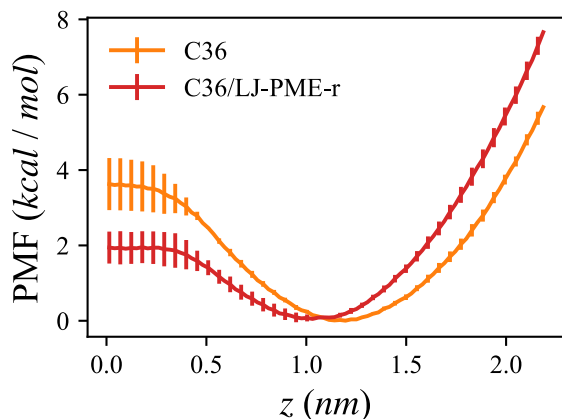


**Figure 6-1** The glycerol moiety of three TG models with partial charges. The revised TG model (C36/LJ-PME-r) has a significantly reduced charge distribution in the ester group (shown in red).



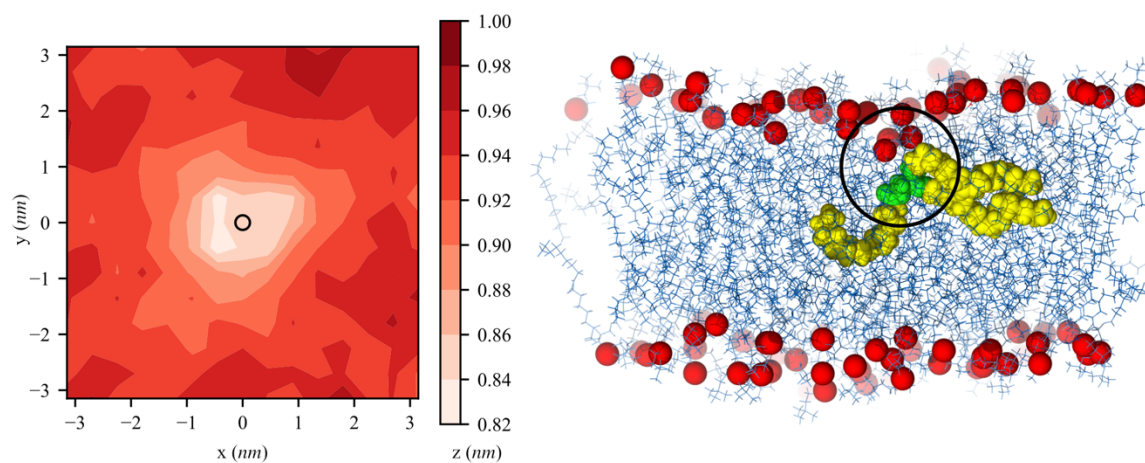
**Figure 6-2** Physical properties of the bulk TG evaluated with three different TG models.

For C36, the properties were calculated as a function of LJ cutoff distances and the force switching function was applied between  $r_c - 0.2$  nm and  $r_c$ . For the surface tension at the water interface, the values were averaged every 100 ns. C36 (1.2 nm) represent the simulation carried out with C36 using the cutoff distance of 1.2 nm.



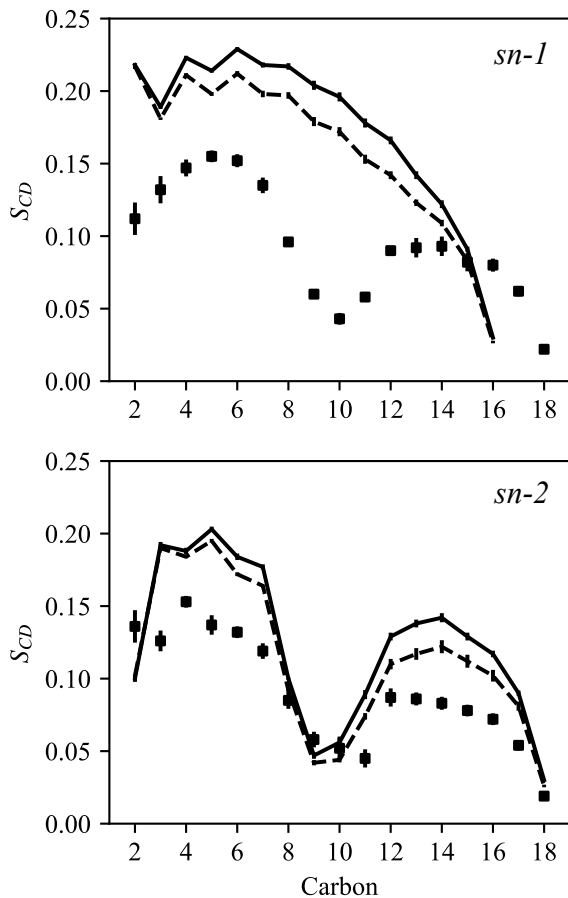
**Figure 6-3** PMF of a TG molecule in a POPC bilayer as a function of the Z distance from the bilayer center (TG flip-flop), calculated from the REUS simulations.

The Z position of the center of the phosphorus atoms was zeroed. The errors represent the standard deviation of the PMFs obtained from the five equal-length blocks.



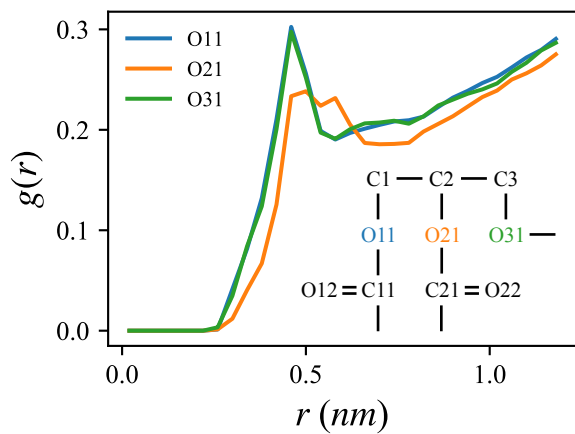
**Figure 6-4** Height field,  $z(x, y)$ , of the phosphorus atoms of the upper leaflet relative to the glycerol moiety of the SURF-TG molecule (circle at the origin).

Frames in which SURF-TG resided at the upper leaflet were used for the analysis. (*right*) Snapshot demonstrating the local deformation by SURF-TG. The red spheres are the phosphorus atoms and PLs are shown in blue lines. The glycerol moiety and acyl chains of TG are shown with green and yellow spheres, respectively. The simulation was carried out with C36/LJ-PME-r.



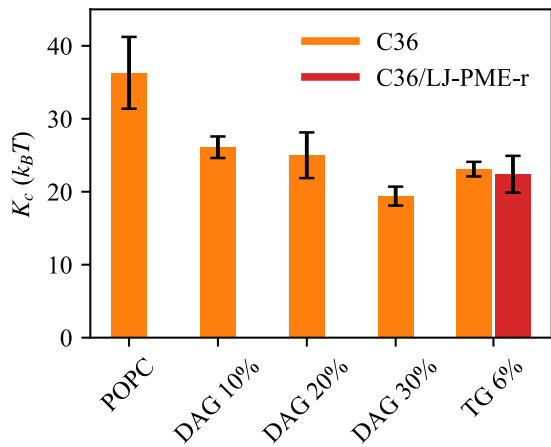
**Figure 6-5** Order parameters of *sn-1* (top) and *sn-2* (bottom) chains.

The POPC order parameters of the pure POPC bilayer (served as a reference) are shown in the dashed lines. The POPC order parameters of the POPC + TG bilayer are indicated in the continuous lines. The TG order parameters of the same system are shown in squares. The *sn-3* chain of TG is not included for visual clarity. The simulations were carried out with C36/LJ-PME-r.



**Figure 6-6** The RDFs of TG oxygen atoms (O11, O21, O31) with the oxygen atom of water.

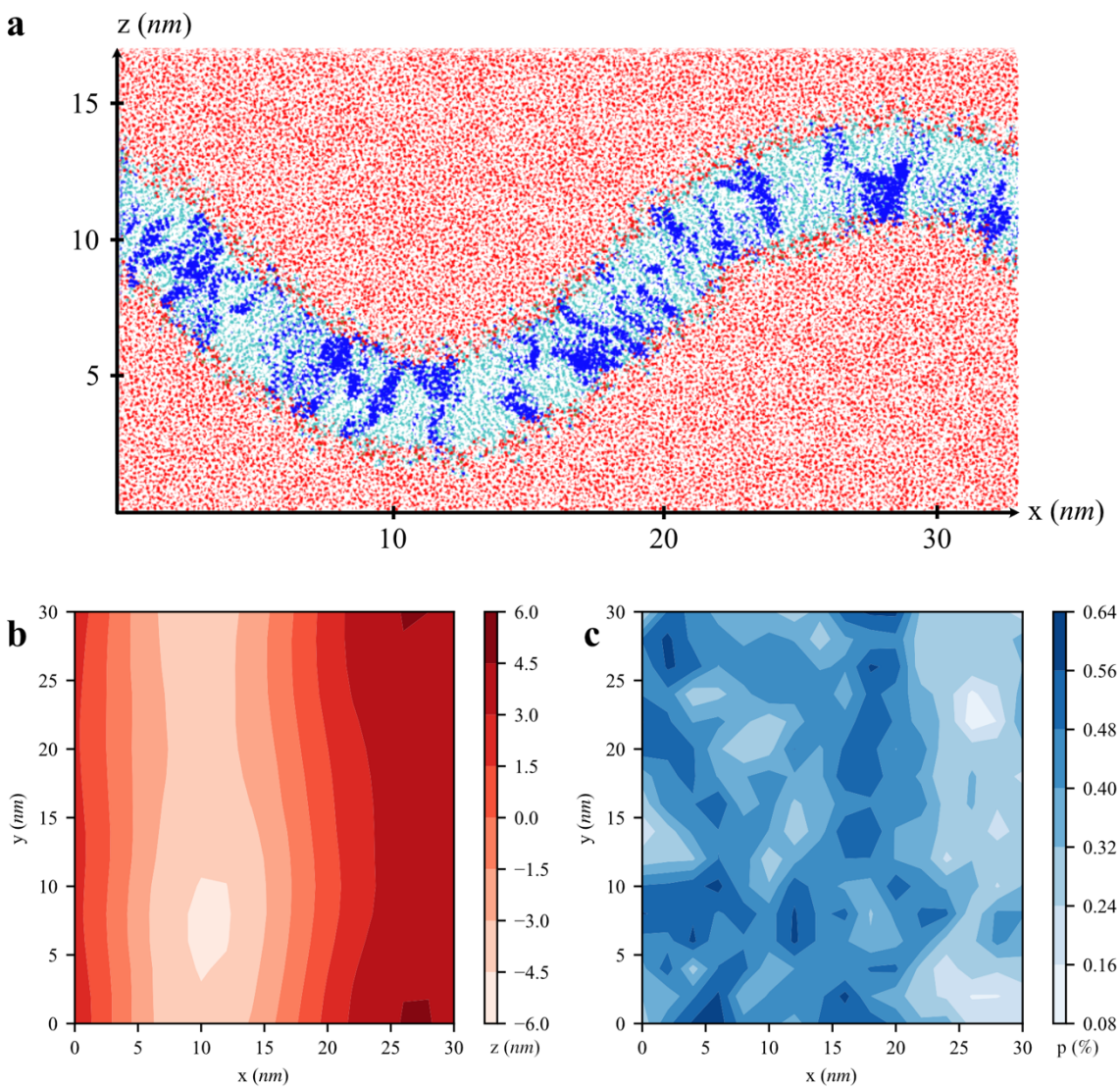
The structure of the TG glycerol moiety is illustrated in the inset. The simulation was carried out with C36/LJ-PME-r.



**Figure 6-7** Bending modulus of five POPC bilayer membranes, containing different amounts of DAG or TG.

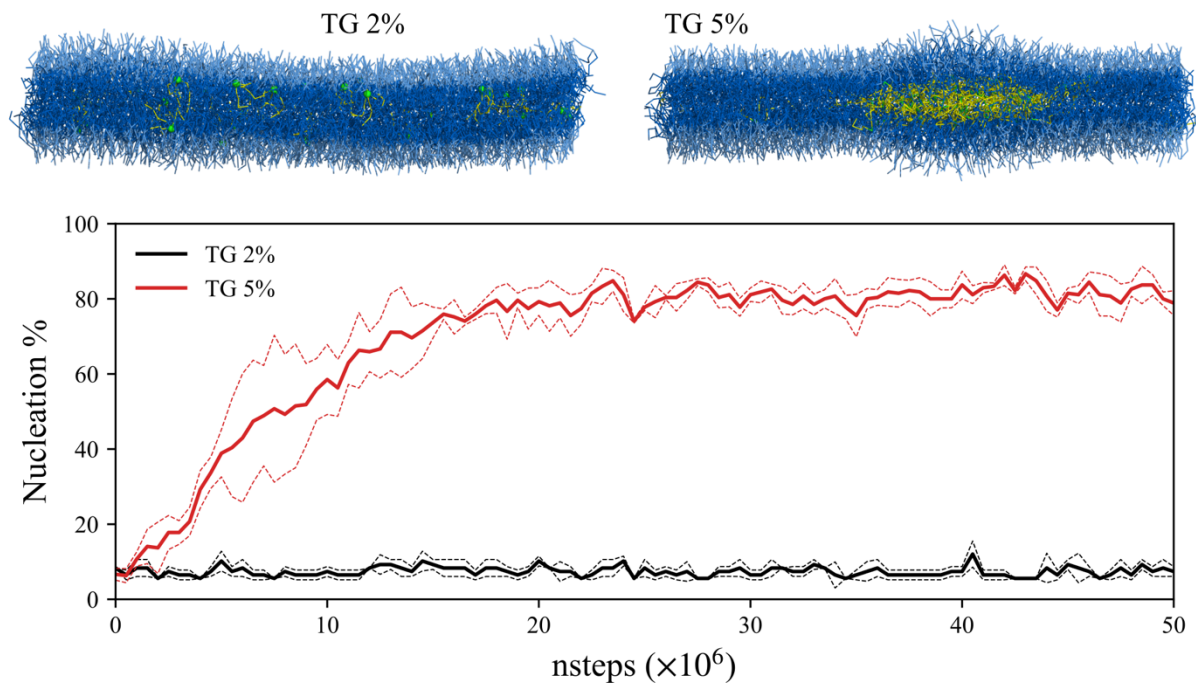
The number of POPC and DAG or TG molecules in each system is as follows: POPC (POPC 4050), DAG 10% (POPC 3200 + DAG 320), DAG 20% (POPC 3200 + DAG 640), DAG 30% (POPC3200 + DAG 960), TG 6% (POPC 1984 + TG 128). The errors represent the standard errors.





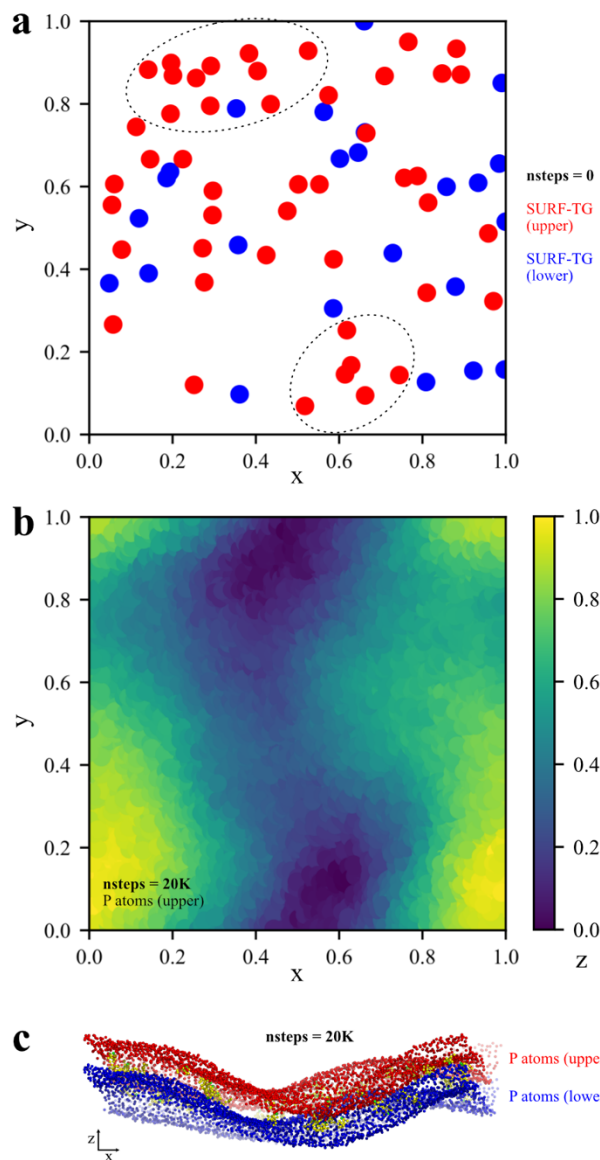
**Figure 6-8** Buckled POPC+DAG membrane (DAG 30%).

a) The snapshot of the MD simulation. Water is in red and DAG in blue. b) The height field,  $z(x, y)$ , of the upper leaflet, averaged over the trajectory. The averaged Z position of the upper phosphorous atoms was zeroed. c) The DAG distribution,  $p(x, y)$ , in the upper leaflet, averaged over the trajectory.



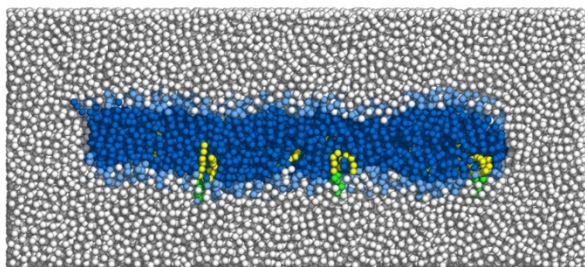
**Figure 6-9** CG simulations of a symmetric bilayer and TG nucleation.

The last snapshots (50M steps) of two membranes and nucleation % are shown. The dashed lines indicate the standard deviation of the quantity from three replicas for each concentration. In the snapshots, PL head group and acyl chains are shown in light blue and dark blue, respectively. The TG glycerol moiety and TG acyl chains are shown in green and yellow, respectively.



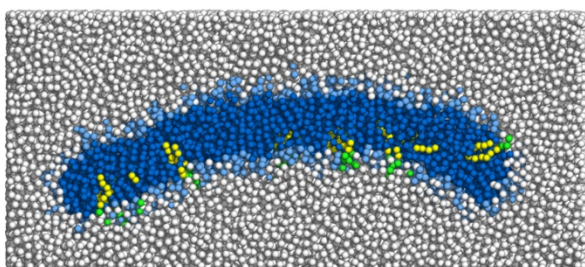
**Figure 6-10** CG simulation of an asymmetric bilayer and membrane bending toward the lower leaflet.

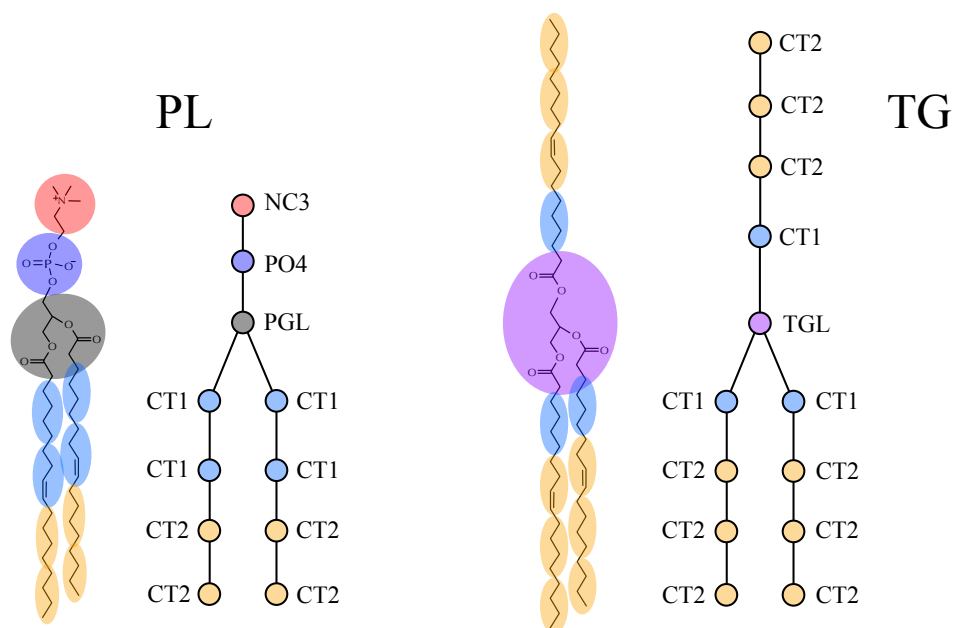
a) The initial distribution of SURF-TG in the upper leaflet (red) and in the lower leaflet (blue). The upper leaflet contains more SURF-TG than the lower leaflet; the scaled X, Y coordinates were used. The regions where the local density of SURF-TG is high in the upper leaflet are circled. b) The height of the phosphorus atoms of the upper leaflet at 20K steps. The scaled X, Y, and Z coordinates were used. c) The CG-MD snapshot at 20K time steps; the upper phosphorus atoms are shown in red and the lower phosphorus atoms in blue. TG glycerol moiety and acyl chains are shown in green and yellow, respectively.



**Figure 6-11** Initial (top) and 28.3 ns (bottom) snapshots of a bilayer membrane.

Clipped in the XZ plane. POPC head group and tails and DOPE head group and tails are shown in light blue and dark blue and green and yellow, respectively. Waters are shown

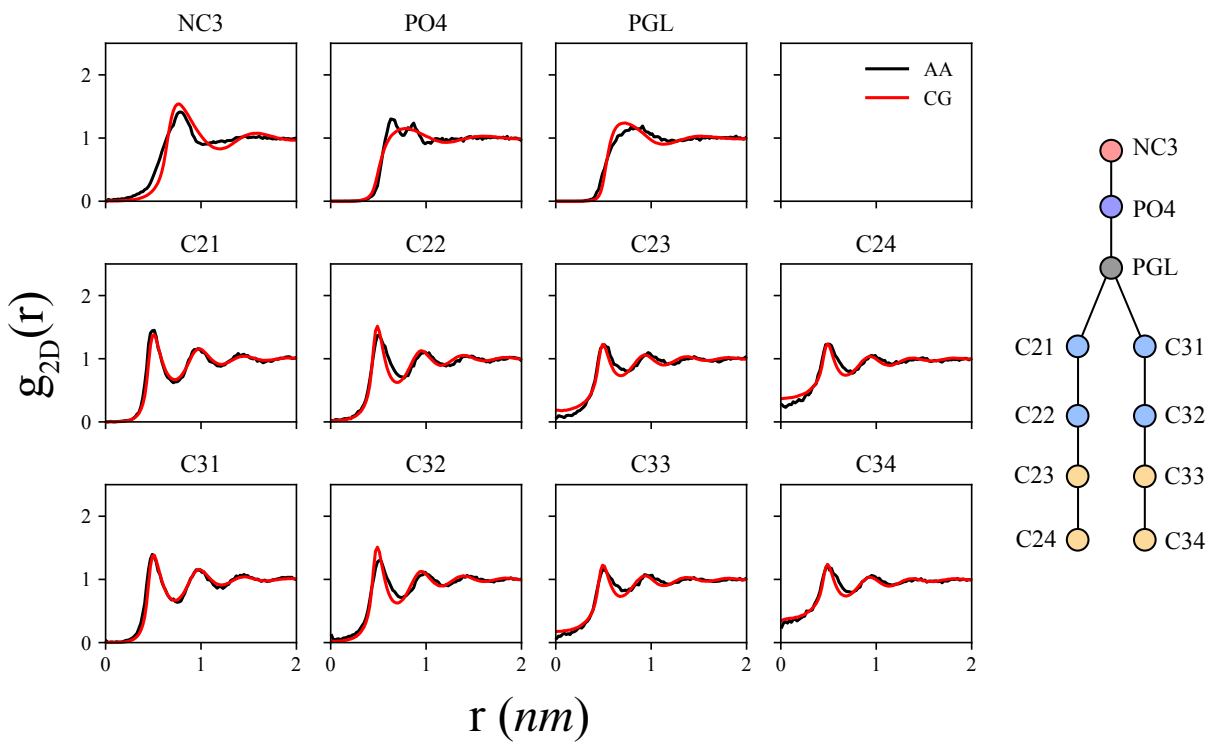




Atom 1	Atom 2	A [kcal/mol]	$\sigma$ [nm]
NC3	NC3	0	0.95
PO4	PO4	0	0.70
PGL	PGL	0	0.75
CT1	CT1	1.5	0.68
CT2	CT2	0.8	0.69
CT1	CT2	1.0	0.685
TGL	TGL	1.5	0.8
PGL	TGL	1.0	0.775

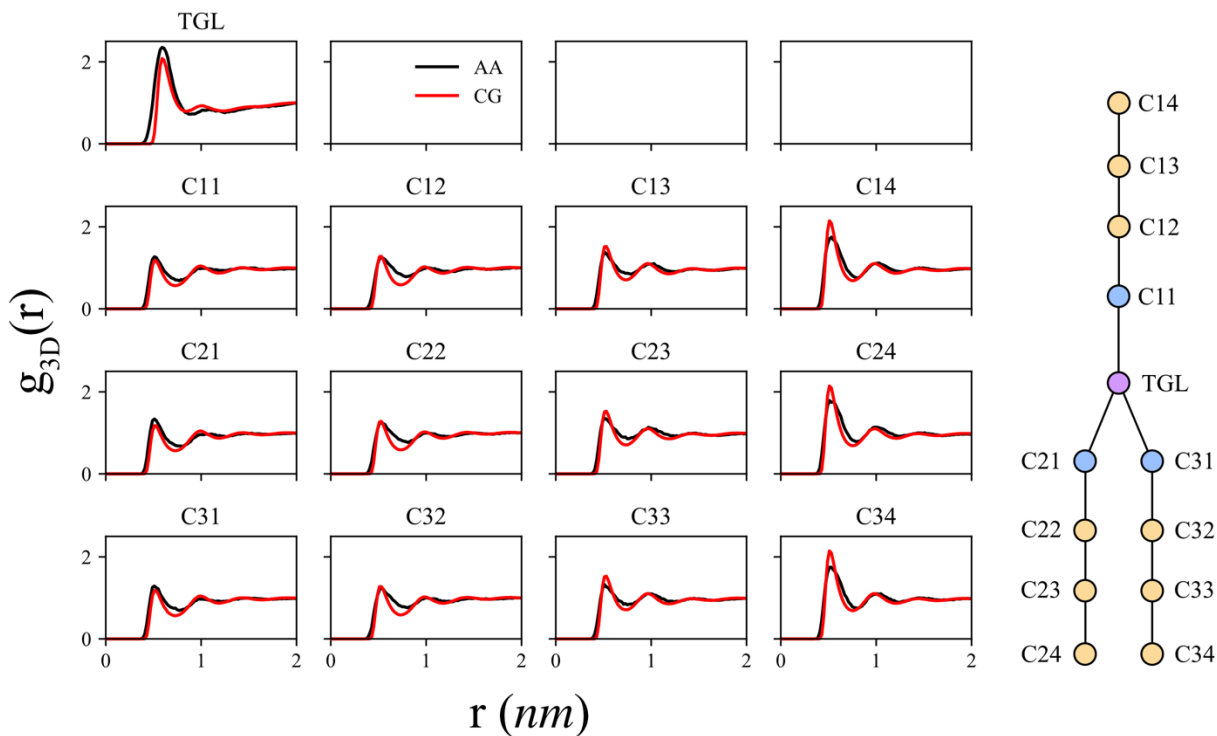
**Figure 6-12** Description of the CG model.

The potentials used for this model are the Gaussian function as the attraction,  $-A \exp(-Br^2)$  where  $B$  was set to  $2 \text{ nm}^{-2}$ , and the LJ repulsive part as the repulsion,  $4\epsilon (\sigma/r)^{12}$  where  $\epsilon$  was set to  $0.0028 \text{ kcal/mol}$ . The parameters for atom  $i$  and atom  $j$ , where  $i \neq j$ ,  $A_{ij} = \sqrt{A_i A_j}$  and  $\sigma_{ij} = 0.5 (\sigma_i + \sigma_j)$ , unless otherwise specified. The mass of NC3, PO4, PGL, CT1, CT2, TGL, and PGL are approximately as-is in the mapping, which are 87, 95, 157, 55.9, 55.9, and 215 *amu*, respectively. The equilibrated bond distance and force constant are  $0.5 \text{ nm}$  and  $500 \text{ kcal/mol} \cdot \text{nm}^2$ , respectively. The CG topology and force field are available at [https://github.com/ksy141/SK\\_CGFF.git](https://github.com/ksy141/SK_CGFF.git).



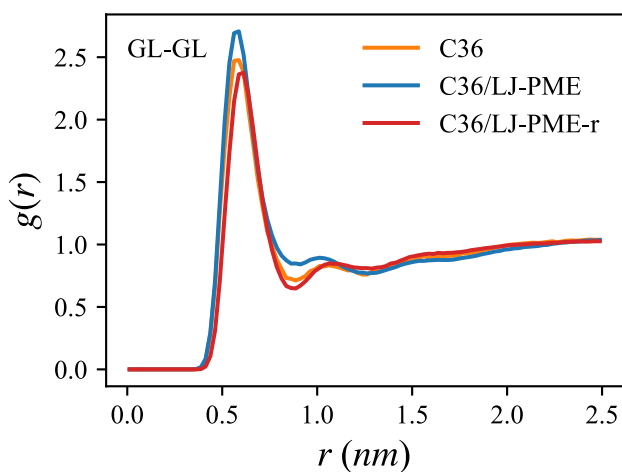
**Figure 6-13** Comparison of the 2-dimensional radial distribution functions computed from the CG (red) and CG mapped atomistic (black) trajectories for DOPC.

The AA simulation was carried out with C36 using a cutoff distance of 1.2 nm.

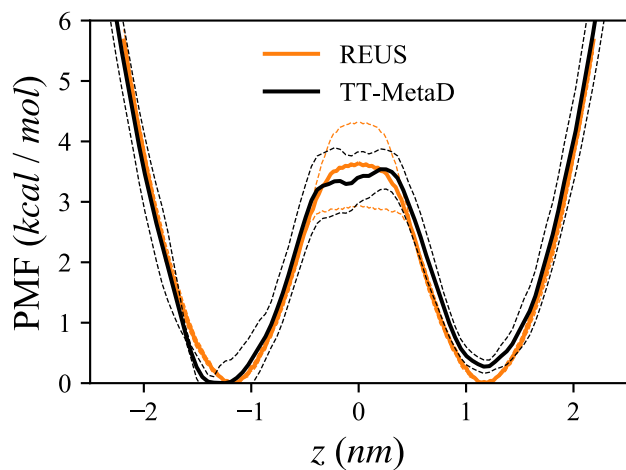


**Figure 6-14** Comparison of the radial distribution functions computed from the CG (red) and CG mapped atomistic (black) trajectories for TG.

The AA simulation was carried out using C36 with a cutoff distance of 1.2 nm.

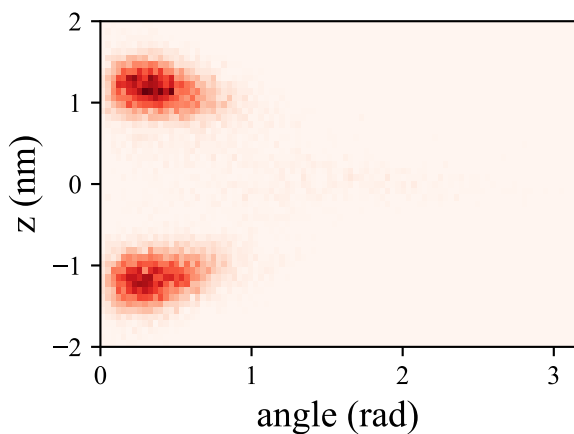


**Figure 6-15** Comparison of the radial distribution functions of the TG glycerol moiety computed from CG mapped atomistic trajectories of three different TG models.



**Figure 6-16** PMF obtained with REUS (orange) and TT-MetaD (black).

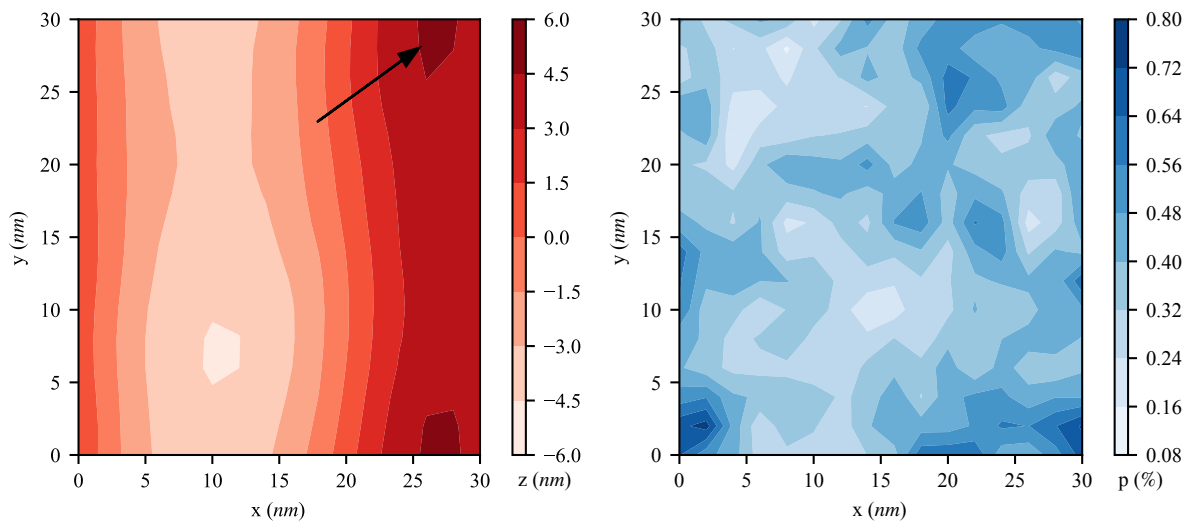
The simulations were performed with C36 using a cutoff distance of 1.2 nm. The dashed lines indicate the standard deviation. Five equal-length blocks and four replicas were used for REUS and TT-MetaD, respectively, to estimate the standard deviation. The REUS PMF was reflected with respect to the membrane center. Related to Fig. 6-3 of the main text.



**Figure 6-17** Orientation and position of TG molecules in a bilayer membrane.

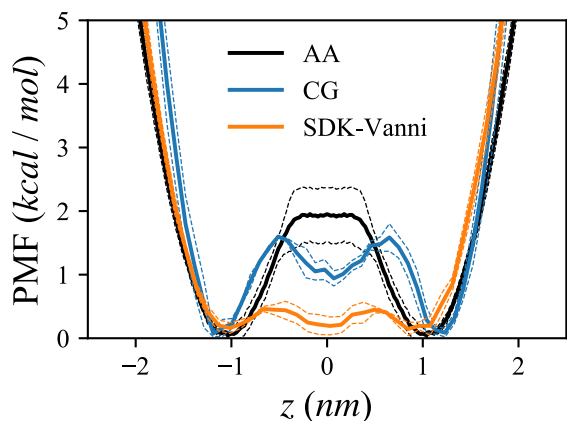
The Z position of 0 represents the membrane center. The angle is between the Z-axis and the positional vector of the TG glycerol moiety from the center of the mass of the TG acyl chains.





**Figure 6-18** Height field (left) and DAG distribution (right) of the lower leaflet of the DAG 30% membrane.

An arrow in the left figure indicates an exemplary motion of a DAG molecule diffusing into the negative curvature region during 500 ns. The total length of simulation was 1  $\mu$ s. Related to Fig. 6-8 of the main text.



**Figure 6-19** TG flip-flop PMF comparison.

The simulations were carried out with the C36/LJ-PME-r force field (black), the CG force field developed in this manuscript (blue), and the Vanni's modified SDK force field (orange). In the CG simulations, three replicas were used to estimate the standard deviation, shown in the dashed lines.

## **Chapter 7 Membrane rigidity as a key factor governing initial stages of lipid droplet formation**

### **ABSTRACT**

Lipid droplets (LDs) are neutral lipid storage organelles surrounded by a phospholipid (PL) monolayer. LD biogenesis from the ER is driven by phase separation of neutral lipids, overcoming surface tension and membrane deformation. However, the biophysics of the initial steps of LD formation are poorly understood. Here, we use a highly tunable, phenomenological coarse grained (CG) model to study triacylglycerol (TG) nucleation in a bilayer membrane. We show that PL rigidity has a profound influence on TG lensing and membrane remodeling: When membrane rigidity increases, TG clusters remain more plane with high anisotropy but a minor degree of phase nucleation. This finding was confirmed by free energy sampling simulations that calculate potential of mean force (PMF) as a function of the degree of nucleation and anisotropy. We also show that asymmetric tension, controlled by the number of PLs on each membrane leaflet, determines the budding direction. A TG lens buds in the direction of the monolayer containing excess PLs to allow for better PL coverage of TGs, consistent with reported experiments. Finally, two governing mechanisms of the LD growth, Ostwald ripening and merging, are observed. Taken together, we study the interplay between two thermodynamic terms at the initial LD phases, TG bulk free energy and membrane remodeling energy.

## INTRODUCTION

Lipid droplets (LDs) are ubiquitous organelles that store lipid. LDs are considered an oil-in-water emulsion in a cell with their core consisting of neutral lipids such as triacylglycerol (TG) or sterol esters, surrounded by a phospholipid (PL) monolayer (20, 25, 44, 47). During LD formation from the endoplasmic reticulum (ER), cells package neutral lipids with a PL monolayer. Lipid droplet assembly complexes (LDACs), and in particular the protein seipin, determine where LDs form and facilitate the process (38).

Biophysically, LD emergence can be considered in the context of classical nucleation theory (45). The driving force of TG nucleation is the bulk energy of TG, stemming from the hydrophobic interactions of TGs' three acyl chains and polar interactions between TGs' glycerol moieties. As LDs grow they inflict a few energy penalty due to raising surface tension ( $\sim 1$  mN/m), proportional to the LD surface area (34). In addition, TG lensing in the ER membrane leads to deformation of the membrane. The energy penalty due to membrane deformation is much more dominant than the surface tension term during the initial phases of LD formation when the phase boundary between the forming LD and cytoplasm is small. However, as the LD surface expands with LD growth, surface tension becomes dominant (35). Therefore, one may expect that an initial TG lens is flat to reduce membrane deformation, and it becomes more and more spherical to reduce the surface tension penalty as the LD grows. Such a process has been predicted in theory (302, 303) and shown in the molecular dynamics (MD) simulations (304).

Due to their limited time and length scales, studying LD biogenesis with all-atom (AA) simulations is not viable. For instance, TG and diacylglycerol (DAG) molecules do not nucleate in bilayers during 1  $\mu$ s at the higher concentrations than the critical (69). Therefore, in this study, we study initial LD biogenesis with a tunable, highly coarse-grained PL model developed by

Grime and Madsen and a TG model derived from it (305). Using the CG model, we aim to understand the initial LD formation in the regime where the membrane deformation penalty is more significant than the surface tension penalty. We report the interplay between the shape of a TG blister and PL rigidity, Ostwald ripening, the tension-dependent budding behavior, and the calculation of TG nucleation PMF.

## METHODS

### *CG lipid model*

A solvent-free, phenomenological CG model for PL and TG was used (305). Each PL and TG molecule consists of four CG beads (Fig. 7-1a). Despite a linear model, it correctly represents the number of acyl chains that each molecule has and the relative effect of hydrophobic interactions. PL has two acyl chains, therefore there are two tail atoms in the CG model and TG has three acyl chains and there are three tail atoms in the CG model. No CG bead carried any charge. All CG pairs except bonded interact with each other with the following pair potential,

$$F(r) = \begin{cases} A \cos\left(\frac{\pi}{2r_0}r\right) & r \leq r_0 \\ -B \cos\left(\frac{\pi}{2} - \frac{\pi}{r_c - r_0}(r_c - r)\right) & r_0 < r \leq r_c \\ 0 & r_c < r \end{cases}$$

where  $r_c$  is  $2r_0$ . The repulsion is a sine-based soft-core repulsion and much softer than a hard-core repulsion such as Lennard-Jones, which allows a larger integration timestep. In addition, the absence of electrostatic interactions, the relatively short cutoff distance (1.5 nm) of nonbonded interactions (considering the removal of charges), and the low resolution enable fast and efficient calculations and large length and time scales. The repulsion parameter,  $A$ , was chosen as  $25 k_B T$ ,

consistent with the original PL model (305). The attraction occurs between the following pairs of atom types: PGL-PGL, T-T, TGL-TGL, and TGL-T. The attraction parameter,  $B$ , was set to 1  $k_B T$  except for the TGL-TGL pair, which was chosen as 1.1  $k_B T$ . The higher attraction in the TGL-TGL pair was motivated by the previous paper that shows a sharper radial distribution function between the TG glycerol moieties than any other PL or TG pairs from the CG mapped atomistic trajectories (69). The other pair is purely repulsive by setting the attraction parameter to 0.  $r_0$  was set to 7.5 Å except for the interaction between the head and head groups, which was set to 4.5 Å. The potential and force become 0 at the cutoff distance ( $r_c$ ), therefore there is no need to apply a switching or shifting function. A bonded interaction is harmonic with an equilibrium distance of 7.5 Å and a spring constant of 25  $k_B T/\text{Å}^2$ . Each PL molecule has two harmonic angles (Fig. 7-1a) with an equilibrium angle of 180° and a spring constant of 0.5  $k_B T$  (soft PL) or 2  $k_B T$  (stiff PL). TG molecules do not have angle potentials. The mass of each CG bead was chosen as 200 g/mol. We note that this is a purely phenomenological model, and therefore the quantities calculated here are not likely related to any real underlying all-atom system directly.

### *AA simulation*

A 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer membrane, consisting of 64 molecules in each leaflet, was constructed using the CHARMM-GUI membrane builder (55, 56). The production run was conducted for 200 ns by GROMACS 2018 (54) with the CHARMM36 force field (61). The Lennard-Jones interaction was force-switched between 1.0 nm to 1.2 nm. Simulations were evolved with a 2-fs timestep. The particle mesh Ewald algorithm (119) was used to evaluate the long-range electrostatic interactions with a real distance cutoff of 1.2 nm. A bond involving a hydrogen atom was constrained using the LINCS algorithm (223). The Nose-Hoover

thermostat was used with a target temperature of 310 K and with a coupling time constant of 1 ps (225, 226). Semi-isotropically coupled pressure was controlled with the Parrinello-Rahman barostat with a target pressure of 1 bar and with a compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$  and a coupling time constant of 5 ps (120).

### *CG simulation*

The CG simulations were carried out using LAMMPS (29 Oct 2020) with tabulated CG potentials (261). Simulations were evolved with a 50-fs timestep. The Langevin thermostat was used with a target temperature of 310 K and with a coupling constant of 100 ps (306). In a flat bilayer simulation, the Nose-Hoover barostat with Martyna-Tobias-Klein correction was used with a target pressure of 0 atm in the XY dimension and with a coupling constant of 250 ps (120, 262, 263). The pressure in the X and Y dimensions was coupled. The cutoff distance of nonbonded interaction was set to 1.5 nm, where both the force and potential become 0. Biased simulations (described below) were performed with the external plugin PLUMED2.6 (135). The initial structures of CG simulations were prepared with MDAnalysis (62). Simulation details are provided in Table 7-1.

**Table 7-1** Description of CG simulations.

CG Simulations	Lipid composition	Nsteps	Ensemble	Angle parameter	To compute
PL bilayer	128 PL	1M	NPT	0.5 k <sub>B</sub> T, 2.0 k <sub>B</sub> T	APL, $K_A$ , O.P.1, O.P.2, density profile
PL bilayer	5000 PL	30M	NPT	0.5 k <sub>B</sub> T, 2.0 k <sub>B</sub> T	Bending modulus
Bulk TG	512 TG	1M	NPT	N/A	density
Bulk TG + water	512 TG	1M	NVT	N/A	interfacial tension
PL + TG bilayer	4900 PL + 100 TG, 4700 PL + 300 TG	100M	NPT	2.0 k <sub>B</sub> T	nucleation % and anisotropy
PL + TG bilayer	1764 PL + 36 TG, 1692 PL + 108 TG	50M, 120M	NPT + biased	0.5 k <sub>B</sub> T, 2.0 k <sub>B</sub> T	nucleation PMF
PL + TG sphere	15910 PL + 982 TG, 16848 PL + 1044 TG, 17808 PL + 1084 TG,	200M	NVT	0.5 k <sub>B</sub> T, 2.0 k <sub>B</sub> T	shape, budding, Ostwald ripening

### *Nucleation percentage and anisotropy*

We mainly calculated two quantities in this study: the nucleation percentage and anisotropy. The former was defined as the number of TG molecules in the largest cluster divided by the total number of TG molecules in the system. If two TGL atoms are within 2 nm (chosen by visual inspection of CG trajectories), those two TG molecules are considered to be in one cluster.

Anisotropy ( $k$ ) describes the shape of a TG lens and is computed in the following procedure: First, we find TG molecules that belong to the largest cluster. Second, we calculate the moment of inertia tensor of those TG molecules from the center of mass of the TG lens. Third, the moment of inertia tensor is diagonalized. Finally, we calculate the anisotropy as:

$$k = \frac{3}{2} \frac{\lambda_1^4 + \lambda_2^4 + \lambda_3^4}{(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)^2} - \frac{1}{2},$$

where  $\lambda_1, \lambda_2$ , and  $\lambda_3$  are the eigenvalues of the inertia tensor. Anisotropy ranges from 0 to 0.25, where 0 represents a sphere, and 0.25 does a plane.

### *Area compressibility*

Bilayer area compressibility was calculated using the following equation

$$K_A = k_B T \frac{\langle A \rangle}{\langle A^2 \rangle - \langle A \rangle^2},$$

where A is the area of a membrane.

### *Order parameters*

The PL order parameters were calculated with the following equation,  $S_{CD} = 0.5 \times |3 \cos^2 \theta - 1|$ , where the angle ( $\theta$ ) is between the Z-axis and the position vector of a tail atom to a glycerol group (PGL). If the positional vector is from the first tail atom, which is closer to the PGL atom, the calculated quantity is referred to as O.P.1, and if the vector is from the second tail atom, the quantity is referred to as O.P.2. For the AA trajectory, a CG-mapped trajectory was first obtained (Fig. 7-5), followed by the calculation of the order parameters.

### *TG nucleation PMF*

Well-tempered metadynamics simulations (307) were carried out to compute the TG nucleation PMF using the same procedure described in (308). The Gaussian hill was deposited every 500 steps at the height of 0.48 kcal/mol and a width of 10. A biasfactor of 50 was used. The biased collective variable was the sum of the coordination number of the TGL atoms (TG glycerol atoms; see Fig. 7-1a) that belong to the largest TG cluster. A sharp switching function was defined to compute the coordination number,



$$M = \begin{cases} 1 & r \leq r_0 \\ (y - 1)^2(1 + 2y) & r_0 < r \leq r_1 \\ 0 & r_1 < r \end{cases}$$

, where  $y = \frac{r-r_0}{r_1-r_0}$ ,  $r_0 = 1.95 \text{ nm}$ , and  $r_1 = 2.00 \text{ nm}$ . The above switching function is plotted in Fig. 7-6. By defining the sharp function that switches to 0 at 2.0 nm, this collective variable has a consistent cutoff distance with the nucleation percentage. The biased trajectories were reweighted with the following equation,

$$\langle A(\mathbf{q}) \rangle = \langle A(\mathbf{q}) e^{\beta [V(s(\mathbf{q},t) - c(t))]} \rangle_V$$

, where  $A$  and  $\mathbf{q}$  represent the property of the interest and the coordinates of atoms, respectively (309, 310).  $V$  and  $s$  represent the biasing potential and collective variable. The time-dependent constant was calculated as  $c(t) = \frac{1}{\beta} \frac{\int ds e^{-\beta F(s)}}{\int ds e^{-\beta [F(s) + V(s,t)]}}$ . The final potential of mean force (PMF) was represented with the nucleation percentage and anisotropy. The PLUMED script used in biasing simulations is included in the Supporting Information.

## RESULTS

### *Physical properties of PL and TG*

First, we characterized the physical properties of a PL bilayer and a bulk TG. We prepared a PL bilayer, consisting of 64 PL molecules in each leaflet, to compute the area per lipid (APL), area compressibility, order parameter (Table 7-2), and number density profile (Fig. 7-7). Two angle potential parameters, 0.5 k<sub>B</sub>T and 2.0 k<sub>B</sub>T, were tested. Both soft and stiff PLs had a liquid-disordered phase at 310 K. We also set up a larger PL bilayer, consisting of 2500 PL molecules in each leaflet, to calculate bending modulus. We first note how angle parameters modulate the

physical properties of bilayers. As PL stiffness increased, PLs became more rigid, therefore reducing the APL and increasing the bending modulus, area compressibility, and order parameters (Table 7-2). This implies that the initial LD formation would be harder, and a TG blister would be more plane with a higher angle potential parameter, which will be discussed later. The same relation between PL rigidity and an angle potential parameter has been observed in the other linear CG models such as the Brannigan-Philips-Brown model (267) and the Cooke-Kremer-Deserno model (311). The density profile (Fig. 7-7) also showed good agreement with the CG-mapped atomistic results of a POPC bilayer.

A 4-bead TG model was derived from the PL model (Fig. 7-1a). We characterized the properties of a bulk TG system containing 216 TG molecules. We determined the volume of each TG molecule and the interfacial tension at the water interface in our implicit solvent model as  $1.30 \pm 0.0 \text{ nm}^3$  and  $26.1 \pm 0.3 \text{ mN/m}$ , comparable to experimental data of those quantities, with  $1.64 \text{ nm}^3$  (116) and  $32 \text{ mN/m}$  (33), respectively. Our model presented here is purely phenomenological and is not directly linked to any underlying all-atom systems. A discussion on CG representability and transferability issues can be found in (312).

**Table 7-2** Physical properties of the POPC bilayer from the AA trajectory and the PL bilayers from the CG trajectories with two different angle potential parameters. Standard errors are given in parentheses.

Angle parameter	CG		AA
	0.5 $k_B T$	2.0 $k_B T$	
APL [ $\text{\AA}^2$ ]	66.2 (0.0)	59.7 (0.1)	65.4 (0.2)
$K_A$ [ $mN/m$ ]	151.7 (4.4)	224.3 (8.3)	221.4 (16.3)
$K_C$ [ $k_B T$ ]	16.6 (0.3)	35.7 (1.0)	31.1 (2.3) <sup>a)</sup>
O.P.1	0.76 (0.02)	0.80 (0.01)	0.75 (0.01)
O.P.2	0.70 (0.02)	0.75 (0.02)	0.75 (0.02)

a) Values from ref (69)

### *TG concentration-dependent nucleation*

Experiments measuring the solubility of triolein in a POPC bilayer reported that  $\sim 2.4\%$  mol TG can be dispersed in the membrane before phase nucleation occurs (37). If bilayers contain fewer TG molecules than this critical concentration, TG is dissolved in PL, whereas TG forms a distinct phase at the critical concentration or above. To model the TG concentration-dependent nucleation behavior, we simulated bilayers with two different TG concentrations, 2% mol and 6% mol TG. Consistent with the experimental data, TG underwent nucleation at 6% mol (Fig. 7-1b). In this case, the TG blister had anisotropy close to 0.25, representing a flat structure, to minimize membrane deformation (Figs. 7-1b and 7-1d). However, for TG concentration of 2% mol, TGs did not nucleate but remained dissolved in the PL phase. (Figs. 7-1b and 7-1c). The simulations discussed here were run with an angle potential parameter of 2.0  $k_B T$ . The same behaviors of TG dissolution at the bilayer containing 2% mol TG and TG nucleation at the bilayer containing 6%

mol TG were observed with an angle parameter of 0.5 k<sub>B</sub>T. By evaluating the nucleation PMF at those two different concentrations and with two different angle parameters, we will later discuss how TG concentrations and angle parameters change the free energy minimum and morphology of TG lenses.

### *Ostwald ripening and PL rigidity-dependent lens shape*

To study the mechanism of LD growth and PL rigidity-dependent lens shapes, a bilayer membrane sphere containing 6% mol TG with a diameter of 40 nm was simulated. We performed CG-MD simulations with two different angle potential parameters, 0.5 k<sub>B</sub>T and 2.0 k<sub>B</sub>T. In both cases, we found that the final structure had one big TG lens between the PL leaflets. Examination of the simulations revealed that this is explained by two distinct mechanisms of LD growth, Ostwald ripening and coalescence of oil phases. In the simulation with stiff PLs, Ostwald ripening was observed (Figs. 7-2a and 7-2b). Two TG lenses were generated, and both lenses grew by attracting the neighboring TG molecules up to 30 M steps. However, after that, the smaller cluster shrank and eventually dissolved, while the larger cluster grew in size. Since the two oil lenses were far apart, this process was not due to merging lenses but due to Ostwald ripening. Ostwald ripening of LDs was recently shown experimentally (313).

In contrast, simulations with soft PLs showed coalescing of clusters as indicated by a sharp increase of the TG number in the largest cluster and a sharp decrease in the second largest cluster. We also simulated a vesicle containing a 1:1 ratio mixture of stiff and soft PLs. In this case, the largest and the third largest oil lenses merged, while the second largest TG cluster gradually

disappeared by Ostwald ripening (Fig. 7-8). However, we did not observe any sorting of PLs and both stiff and soft PLs were equally distributed.

How does PL stiffness impact the shape of LDs? The dominant term that hinders initial LD formation is the membrane deformation penalty. Therefore, we hypothesized that the shape of a TG lens depends on PL rigidity. We found significant differences in anisotropy and the shape of a TG cluster in the vesicular simulations. While the system with stiff PLs showed a flat oil blister, the system with soft PLs had a spherical oil blister (Fig 7-2). Given the same bulk energy per volume, this can be understood as a TG's response to a high membrane deformation penalty due to the PL's high rigidity. We also confirmed that PL rigidity changed the nucleation percentage (Fig. 7-2c). With soft PLs, the equilibrated nucleation percentage increased. Interestingly, a vesicle that contained a 1:1 ratio of soft and stiff PLs presented the nucleation percentage and anisotropy values between those in single-component PL vesicles.

#### *Asymmetric tension and budding*

Two spherical systems with a diameter of 40 nm and an angle parameter of  $2.0 k_B T$  were set up by varying the number of PLs in the inner leaflet while fixing the number of PLs in the outer leaflet. With this approach we imposed asymmetric tension between the monolayers. Consistent with the recent experimental studies (274, 275), we confirmed asymmetric PL density controls the budding direction (Fig. 7-3). A TG blister budded toward the leaflet that better offers the PL coverage of TGs. In the first simulation, where the ratio of the number of PLs in the inner leaflet to that in the outer leaflet is 0.69, a TG lens budded to the outer leaflet representing the cytosolic side of an ER membrane. In the other simulation, where the ratio is 0.89, the budding direction

was reversed into the inside of the vesicle, representing the ER lumen. In contrast, a TG lens remained in a bilayer membrane in the simulation where the ratio is 0.79.

### *TG nucleation PMF*

To estimate the free energy of TG nucleation as a function of the degree of nucleation and anisotropy, we carried out well-tempered metadynamics simulations. Because the degree of nucleation (or equivalently the nucleation percentage) is not a continuous function for a given system and therefore cannot be biased, we instead biased the sum of the coordination number of the TGL atoms in the largest cluster. A high correlation between the degree of nucleation and the sum of the coordination number in the largest cluster was achieved (Fig. 7-4 top and second figures from top) by using a sharp switching function (Fig. 7-6). However, if one uses a broad switching function, a high correlation between the nucleation percentage and the coordination number is not guaranteed. A broad switching function can result in formation of several clusters, each having a dense aggregation of TGs, which is not preferred.

Small bilayers containing 2% mol and 6% mol TG were simulated for these calculations. In the case of 6% mol TG, two different angle potential parameters, 0.5  $k_B T$  and 2.0  $k_B T$ , were used. The coordination number (Fig. 7-4 top) sampled most of the region several times, except the very low and high values. Accordingly, the degree of TG nucleation was widely sampled in the biased simulations (Fig. 7-4 second figure from top). We reweighted the trajectory to calculate the TG nucleation PMF as a function of TG nucleation percentage and anisotropy (see Methods).

We first compared the results of the biased trajectories with those of the unbiased trajectories. Consistent with the unbiased trajectory that did not have TG nucleation in a bilayer

containing 2% mol TG, the PMF of the same bilayer indicated a free energy minimum at the low nucleation percentage. The PMF also increased with the nucleation percentage (Fig. 7-4a). In contrast, the bilayer containing 6% mol TG had a free energy minimum at the high nucleation %. Furthermore, denoted by red lines (Fig. 7-4 third figures from top) and red markers (Fig. 7-4 bottom figures), the equilibrated values of the nucleation percentage and anisotropy from the unbiased simulations agreed well with the free energy minima in the calculated PMFs. Overall, our PMFs showed good agreement with the unbiased trajectories, therefore confirmed the robustness and convergence of our biased simulations.

We next investigated the influence of the angle potential parameter. Soft PLs enhanced the nucleation percentage and decreased anisotropy, which is confirmed in our PMFs by comparing the free energy minima (Fig. 7-4b and 7-4c). The bilayer with an angle parameter of  $0.5 k_B T$  had an increased nucleation percentage and decreased anisotropy compared to the bilayer with an angle parameter of  $2.0 k_B T$ . Also, we observed that the contour of 10 kcal/mol became more extended toward the low anisotropy region in the simulation with soft PLs than the simulation with stiff PLs.

Finally, we compared TG nucleation with argon gas nucleation. In argon gas nucleation, the whole range of anisotropy from 0 to 0.25 was sampled (308), while in TG nucleation, anisotropy is only limited to the high values. This suggests different thermodynamic terms rule nucleation. Argon gas nucleation has the surface tension energy penalty, which makes a cluster tend to become spherical. In TG nucleation, membrane deformation rather than surface tension works against TG nucleation, which makes a TG blister sample only the plane region (high anisotropy region). However, we expect that the lower anisotropy region will be more sampled when a system becomes bigger.

## DISCUSSION

In this study, we investigated the biophysics of LD emergence with a highly tunable, phenomenological CG model (305). We characterized the physical properties of a 4-bead PL model (head - glycerol - tail - tail), including the APL, bilayer area compressibility, bending modulus, order parameters, and density profile. Although we used a generic model, the CG results compared well with the AA results of a POPC bilayer membrane, which is the primary component of PLs in a mammalian cell (7). Two spring constants of harmonic angles,  $0.5 \text{ k}_B\text{T}$  and  $2.0 \text{ k}_B\text{T}$ , were used to control PL rigidity. Based on the pair potentials between PL atoms, we made a 4-bead TG model (glycerol - tail - tail - tail) that showed the concentration-dependent behavior. If its concentration is 5% or above, TG nucleates a lens.

LD biogenesis is driven by neutral lipids' bulk energy and is opposed by surface tension and membrane deformation energy. While the LDs are small, the membrane deformation energy is dominant. However, as the LDs grow, the surface tension energy takes over as it is proportional to the surface area. A characteristic length of this transition is predicted to be  $10 \text{ nm} - 20 \text{ nm}$  (35, 302). In our simulations, LDs are smaller than this characteristic length, therefore we mostly demonstrate the interplay between TG lensing and membrane deformation energy, in which the latter was controlled via PL rigidity.

In the simulations of a vesicle with a diameter of  $40 \text{ nm}$ , we showed PL density-dependent budding phenomena, which is consistent with the recent experimental papers (274, 275). In a simulation where there was an excess PL in the inner leaflet, LDs budded to the center of the sphere, whereas LDs budded to the outside of the sphere if there was an excess PL in the outer leaflet. Therefore, in a closed, tunable-sized bilayer system, the balance of PLs between the inner and outer leaflets determines the budding directionality. In cells, the ER bilayer is an open system



due to vast amounts of ER, scramblases or flippases (314), and de novo PL synthesis; Yet, there are still apolipoprotein B-free luminal LDs (315, 316). Using this model, a possible explanation would be local membrane environment that has excess PLs in the luminal leaflet or stiff PLs in the cytosolic leaflet, driven by unknown reasons.

In the vesicular simulations that do not show budding, two mechanisms contribute to the formation of one large TG lens between the PL leaflets: Lens coalescence and Ostwald ripening. Ostwald ripening was shown in our trajectories where only one of two distanced TG lenses grew and the other became gradually dissolved. The experimental evidence of Ostwald ripening of LDs was reported in (313). Although not directly related to lens coalescence in a bilayer membrane, Fsp27-mediated LD coalescence at the LD contact site was shown in (317).

We also studied the correlation between PL rigidity and the shape of a TG lenses. With reduced PL rigidity, LDs became more spherical, and the degree of nucleation increased. To support our conclusions, we calculated the TG nucleation PMF in a flat bilayer. By biasing the sum of the coordination number of TG glycerol (TGL) atoms in the largest cluster and reweighting the biased trajectory, we were able to compute the PMF as a function of nucleation percentage and anisotropy. Consistent with the vesicular simulations, the free energy minimum is located at the lower anisotropy and higher nucleation percentage with reduced PL rigidity.

Taken together, our findings demonstrate membrane deformation plays a critical role in the initial formation and shape of LDs. How can cells reduce the membrane deformation penalty for LD emergence? Interestingly, TG itself can significantly reduce membrane rigidity. It was recently shown experimentally (269) and computationally (69) that TG reduces bending modulus of a bilayer membrane. An amphipathic helix that creates curvature can be also significant to reduce the bending modulus of the bilayer membrane (275). In addition, the LD assembly complex binds

TG and might aid in the nucleation reaction (38). Consistent with this notion, TG was observed to accumulate on membrane-inserted helices of seipin, increasing their concentration and thus aiding TG oil phase nucleation. (42, 43).

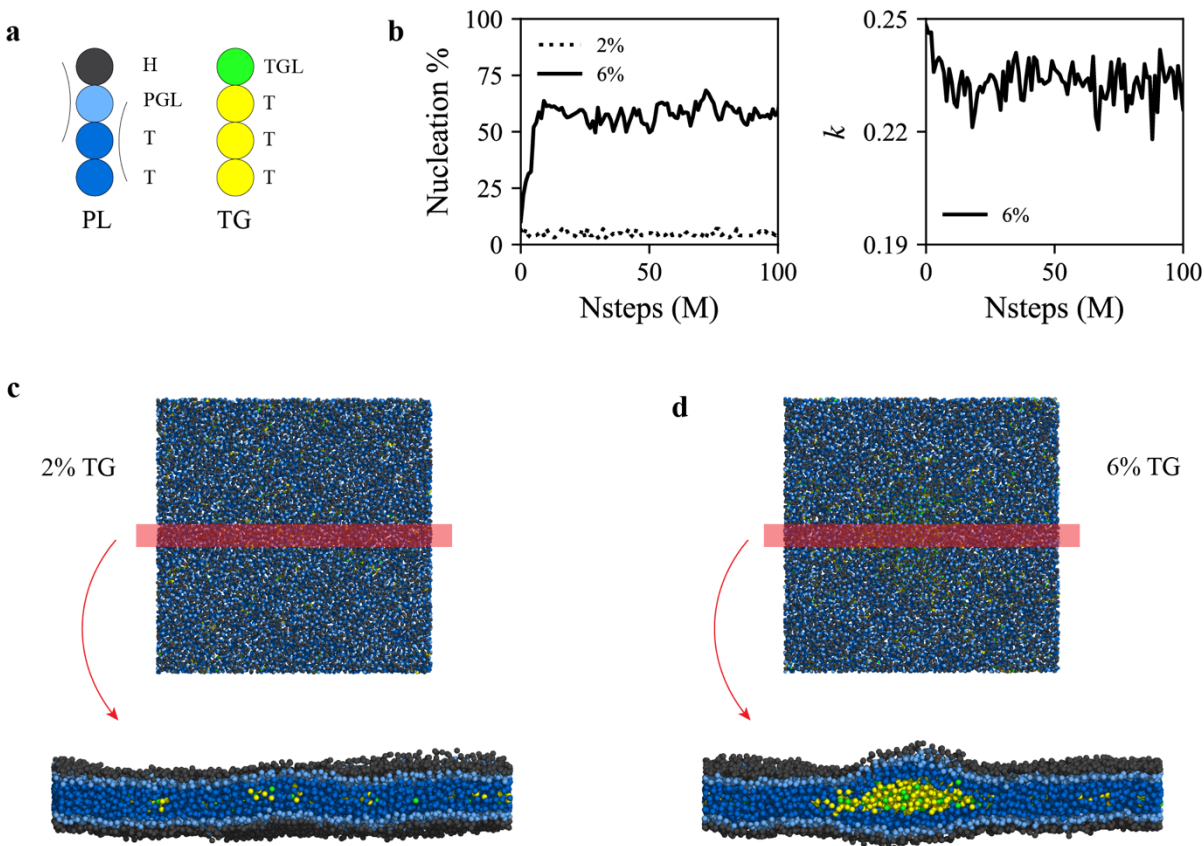
## **CONCLUSIONS**

In a bilayer membrane, TG nucleation is driven by its bulk energy. However, membrane deformation incurs an energy penalty on nucleation at the initial phases of LD formation. Our CG simulations demonstrated the competing effects of TG lensing and membrane deformation. We showed high membrane rigidity reduces the nucleation percentage and increases anisotropy of a TG lens, confirmed in large-scale vesicular simulations and the calculations of the TG nucleation free energy. In addition, two distinct mechanisms that govern LD growth, Ostwald ripening and coalescence of oil lenses, were shown in vesicular simulations. Finally, the LD budding direction was controlled by the number of PLs in the inner and outer leaflets. Taken together, we provide a better understanding of LD formation at the initial steps, validate the reported experiments, and conclude that membrane rigidity serves as a key factor in the formation and shape of a TG lens.

## **ACKNOWLEDGEMENTS**

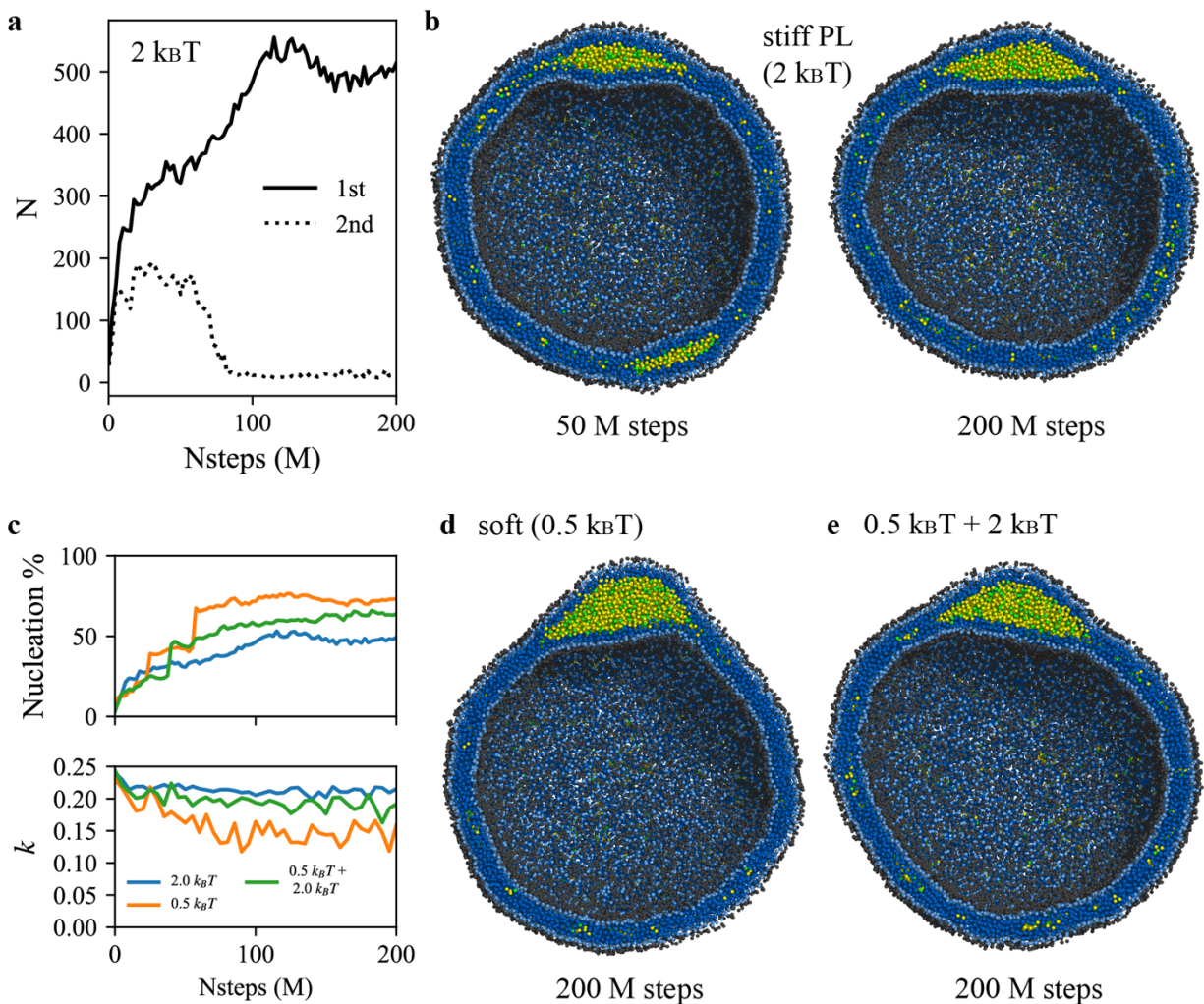
This research was supported by a grant from the National Institute of General Medical Science (NIGMS) of the National Institutes of Health (NIH), grant R01-GM063796. The computer simulations were performed on the midway2 and midway3 supercomputer at the University of Chicago and the Bridges2 supercomputer at the Pittsburgh Supercomputing Center (PSC) through allocation MCA94P017 with resources provided by the Extreme Science and Engineering

Discovery Environment (XSEDE) supported by NSF grant ACI-1548562. We thank Federico Giberti, Jeeyun Chung, and Alex Pak for their useful perspectives and discussions.



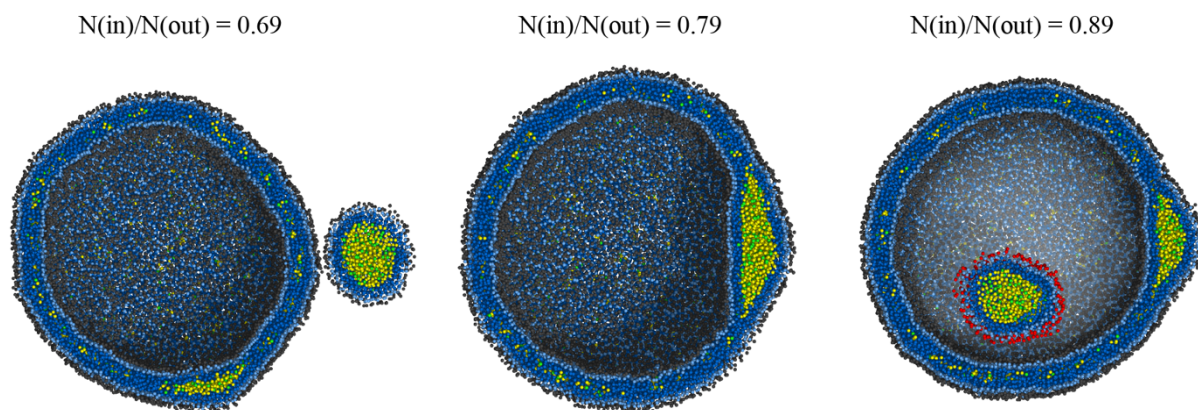
**Figure 7-1** TG concentration-dependent nucleation.

(a) Illustration of the PL and TG models used in this study. Each CG type is written next to each CG bead. Arcs represent harmonic angle potentials. The same color code is used in the rest of the study. (b) Nucleation % (left) and anisotropy (right) for bilayers containing 2% mol (dotted) or 6% mol (solid) TG. For visual clarity, anisotropy of the 2% mol TG bilayer is not shown. The last snapshots (100 M steps) of the bilayers containing (c) 2% mol or (d) 6% mol TG are shown. The top view (top) and the side view (bottom) of the interior (clipped) structure. Simulations discussed here were run with an angle potential parameter of  $2.0 k_B T$ .



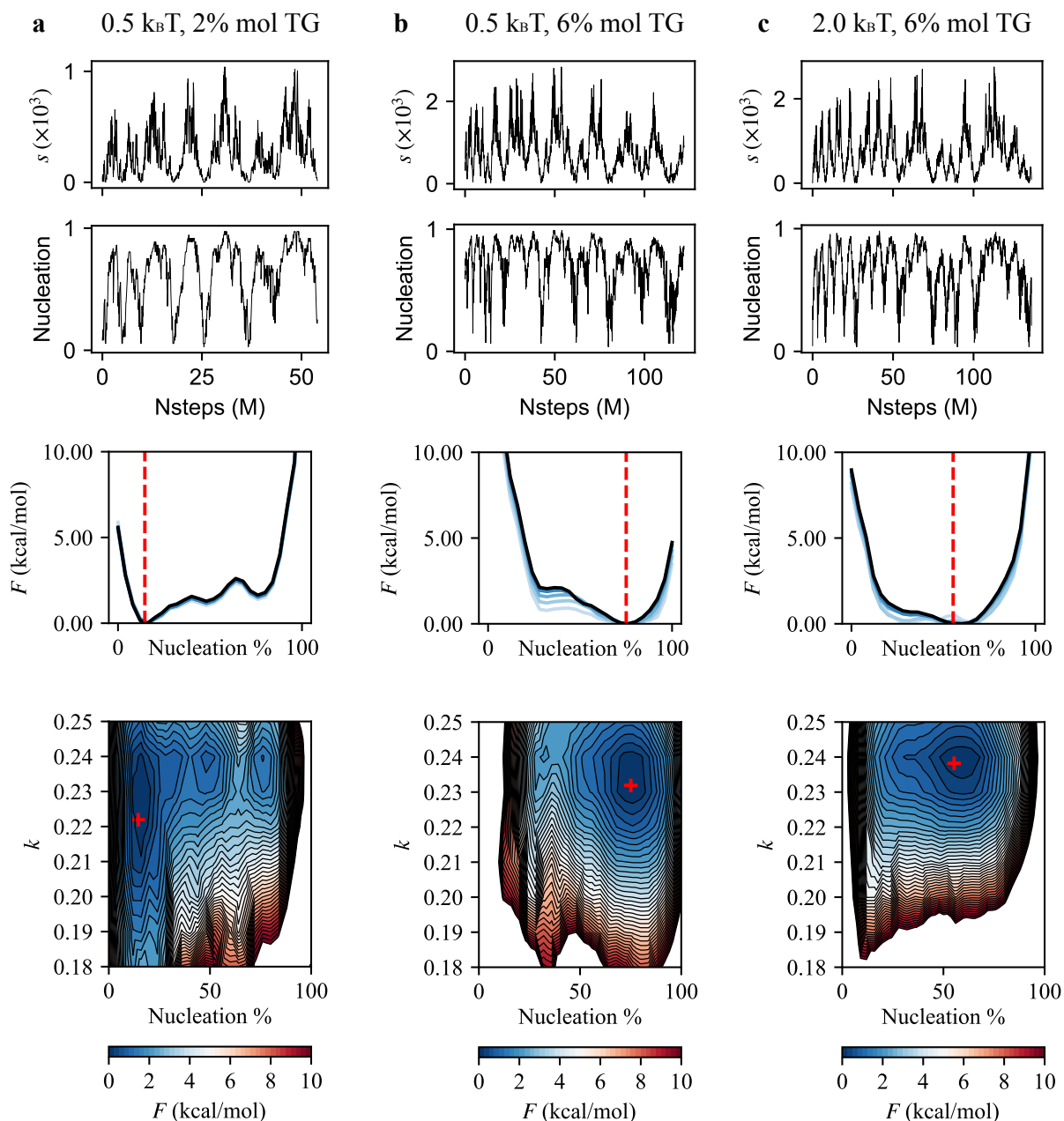
**Figure 7-2** Ostwald ripening and PL rigidity-dependent LD shape.

(a) The number of TGs in the first (solid) and second (dotted) largest cluster. Simulations were run with an angle parameter of  $2 \text{ k}_B T$ . (b) The interior view of the same simulation with (a) at 50 M steps (left) and 200 M steps (right). (c) Nucleation % and anisotropy with simulation times. Three systems were prepared. The first (blue) and second (orange) systems have angle parameters of  $2 \text{ k}_B T$  and  $0.5 \text{ k}_B T$ , respectively. The third system (green) consists of a 1:1 ratio mixture between stiff and soft PLs. The interior view at 200 M steps of the (d) second and (e) third system.



**Figure 7-3** PL density-dependent budding.

The interior view of two spheres that initially had asymmetric tension. LDs bud toward the monolayer that exhibits better PL coverage of a TG cluster. The initial ratio of the PL number in the inner leaflet to that in the outer leaflet is shown. For visual clarity, the PL head group of the budded LD is shown in red in the right figure. Simulations were run with an angle potential parameter of  $2.0 \text{ k}_B\text{T}$ .



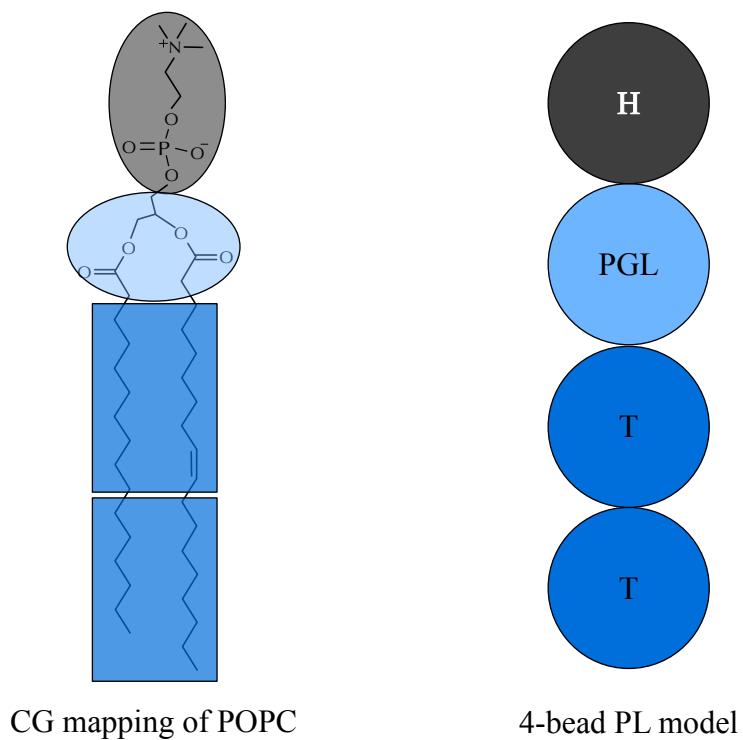
**Figure 7-4** TG nucleation PMF.

From top to bottom: (first figure) the biased collective variable, the sum of the coordination number of TGL atoms in the largest cluster, in the biased trajectory. (second) The degree of nucleation calculated in the biased trajectory. (third) The calculated PMF as a function of nucleation % by reweighting the biased simulation. The red dashed line indicates the equilibrated nucleation % value that was obtained from the unbiased simulation of the same bilayer. The blue lines indicate how PMFs evolve with simulation times to show convergence. The lighter the color, the less simulation frames were used for calculating the PMF. (last) The calculated PMF as a function of nucleation % and anisotropy.

**Figure 7-4, continued.**

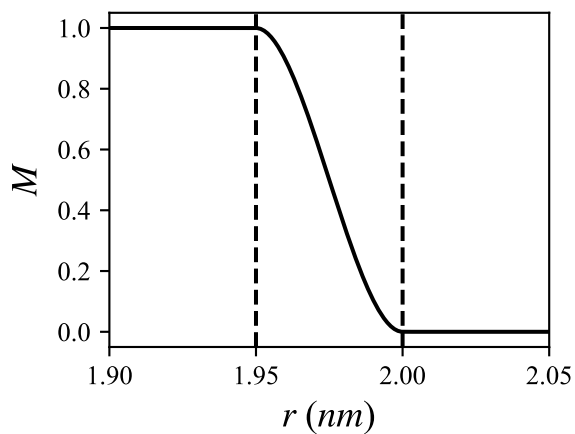
The red marker indicates the equilibrated values of nucleation % and anisotropy from the unbiased simulation of the same bilayer. (a) The simulation was run with an angle parameter of  $0.5 k_B T$  and contains 2% mol TG. (b) Angle parameter of  $0.5 k_B T$  in the 6% mol TG bilayer. (c) Angle parameter of  $2.0 k_B T$  in the 6% mol TG bilayer.

## SUPPORTING INFORMATION

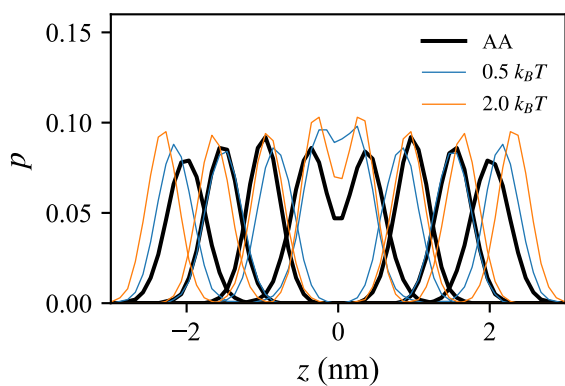


**Figure 7-5** CG mapping of a POPC molecule into a 4-bead PL model to calculate order parameters (Table 7-2) and distribution (Fig. 7-7).



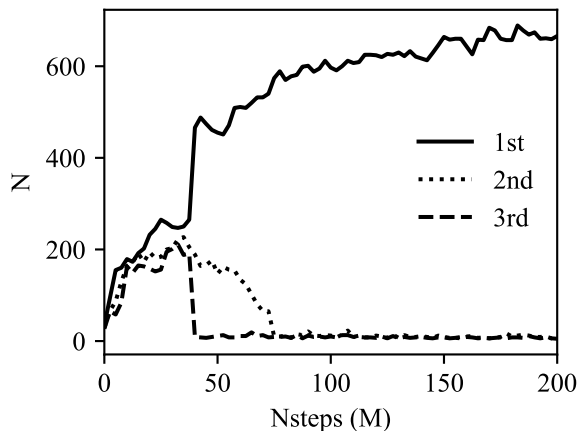


**Figure 7-6** A sharp switching function to compute the coordination number.



**Figure 7-7** Distribution (number) of CG beads in the bilayer normal.

The Z value of 0 represents the center of the membrane. The AA trajectory (black) was first CG-mapped according to the scheme shown in Fig. 7-5. The CG results of PL bilayers with angle potential parameters of  $0.5 k_B T$  and  $2.0 k_B T$  are shown in blue and orange, respectively.



**Figure 7-8** Vesicular system consisting of a 1:1 ratio mixture between soft and stiff PLs.

The number of TG molecules in the first (solid line), second (dotted line), and third largest cluster (dashed line). Merging between the first and third largest clusters happens at 45M steps, indicated by a sharp increase or decrease in the number of TGs. The second cluster gradually disappears by Ostwald ripening.

## PLUMED script for biasing simulations

```
MOLINFO STRUCTURE=molinfo.pdb

lq:    COORDINATIONNUMBER SPECIES={@mda:{name TGI}} SWITCH={CUBIC D_0=1.95 D_MAX=2.0} LOWMEM
cm:    CONTACT_MATRIX      ATOMS=lq           SWITCH={CUBIC D_0=1.95 D_MAX=2.0}
dfs:   DFSCUSTERING        MATRIX=cm           LOWMEM
clust1: CLUSTER_PROPERTIES CLUSTERS=dfs    CLUSTER=1    SUM

METAD ...
label=m
ARG=clust1.sum
HEIGHT=2.0
SIGMA=10
PACE=500
GRID_MIN=0
GRID_MAX=3000
GRID_WSTRIDE=500000
GRID_WFILE=grid.dat
TEMP=310
BIASFACTOR=50
CALC_RCT
... METAD

ss: CLUSTER_NATOMS CLUSTERS=dfs CLUSTER=1
PRINT ARG=clust1.sum,ss,m.bias,m.rbias STRIDE=500 FILE=colvar
```

## **Chapter 8 Seipin transmembrane segments function in triglyceride nucleation and droplet budding from the membrane**

### **ABSTRACT**

Lipid droplets (LDs) are organelles formed by the endoplasmic reticulum (ER) to store triacylglycerol (TG) and sterol esters. The ER protein seipin appears to be key for nucleation of TG and budding of a nascent droplet during LD biogenesis. Seipin forms a large, oligomeric ring of 10-12 subunits (depending on species), with each seipin monomer containing a luminal domain that inserts an evolutionarily conserved hydrophobic helix (HH) into the membrane. The luminal domain is flanked by two transmembrane (TM) segments and short, variable cytoplasmic tails. How the different parts of seipin function in TG nucleation and LD budding is still poorly understood. Here, we utilized all-atom (AA) and coarse-grained (CG) simulations of human seipin to study how seipin functions in the formation of LDs from the ER. AA simulations indicate that hydrophobic interactions between seipin TM segments and HH with TG are the main driving force of TG nucleation, in conjunction with hydrophilic interactions between the TG glycerol moiety and protein residues. We also find seipin TM segments control lipid diffusion and permeation. Simulating larger, growing LDs with CG models, we find that seipin TM segments convert a flat oil lens into a LD budding from the ER with a constricted neck structure. Taken together, our results suggest a model in which TM segments are critical for seipin functions by catalyzing TG nucleation, controlling lipid diffusion and permeation, setting the boundary of an oil lens, and creating a unique ER-LD neck structure.

## **SIGNIFICANCE STATEMENT**

Lipid droplets (LDs) are fat-storing organelles and metabolic hubs of the cell. Defects in LD biology or storage of excess LDs are related to metabolic diseases, such as obesity, lipodystrophy, or fatty liver disease. Seipin is a critical protein that governs the biogenesis of normal LDs, as the seipin deficiency results in very large, and aggregated, small LDs, as well as lipodystrophy. How seipin functions molecularly in LD emergence remains unclear. Here, we use molecular dynamics simulations to study seipin function in LD biogenesis. Our simulations reveal that human seipin catalyzes neutral lipid nucleation and constricts the neck of a budding LD, extruding it towards the cytoplasm.

## **INTRODUCTION**

The lipid droplet (LD) is a fat-storing organelle, surrounded by numerous coating proteins and a phospholipid (PL) monolayer (44, 47). LDs store excessive energy as highly reduced carbon triacylglycerol (TG) and can mobilize fatty acids for energy generation or membrane biosynthesis (20, 21). Due to their key role in metabolism, failure to control LD biogenesis leads to metabolic diseases, such as lipodystrophy. Overwhelming the capacity of cells to form LDs is thought to be crucial for the development of diseases linked to obesity, such as type II diabetes (16). Therefore, studying how LDs emerge from the endoplasmic reticulum (ER) bilayer is not only critical to understanding LD biology and but might also offer approaches to treat prevalent metabolic diseases (18).

Current models of LD biogenesis posit that lipid droplet assembly complexes (LDACs) in the endoplasmic reticulum (ER) bilayer determine LD formation sites and facilitate LD growth (38, 41). LDACs, consisting of seipin and lipid droplet assembly factor 1 (LDAF1) in humans or

seipin/Fld1, Ldb16 and Ldo in yeast, efficiently catalyze the initial stages of LD formation (38, 318, 319). Absence of seipin, effectively removing LDAF1 as well (38), changes LD number and morphology dramatically, resulting in aggregated, small LDs and/or few supersized LDs (202, 320-322). Therefore, investigating how seipin works is key to understanding LD biogenesis.

Human seipin is an undecamer, forming a ring-structure in the ER. Each subunit contains a luminal domain, flanked by two transmembrane (TM) segments and short cytoplasmic tails (38-41, 323). The luminal domain has a conserved hydrophobic helix (HH) thought to insert into the luminal leaflet of the ER membrane. It was suggested that the HH of human seipin, and in particular S165 and S166, are key tethering sites for TG (42, 43) and might provide a binding site of LDAF1 (38). Yeast seipin lacks the HH, which may explain why yeast seipin is not sufficient for function in LD formation (41, 318). It was also suggested that seipin TM segments are required for seipin function (38). Chimeric seipin proteins in which two TM segments were replaced with that of another ER protein FIT2, were not functional and unable to form LDs while they can form a seipin oligomer (38, 41, 324). This implies the crucial roles of the evolutionarily conserved TM segments in LD biogenesis.

In this study, we capitalized on new information on seipin TM segments to investigate their roles in TG nucleation and ER-LD bridge formation using all-atom (AA) and coarse-grained (CG) simulations. Using the AA simulation, we find TG-HH attractions as well as TG-TM attractions. The collective motions of TM segments increased PL permeation by opening the space between neighboring monomers while lipid-TM interactions decreased the rate of diffusion. In our CG simulations, we demonstrate that TG-TM and TG-HH interactions catalyze TG nucleation. We also showed that seipin TM segments convert a planar oil lens into the ER-LD neck structure, consistent with the experimental topology of nascent LDs (313, 325).

## RESULTS

Seipin TM segments are thought to be critical for seipin functions (38). The resolved structures, however, do not include TM segments likely because of their high flexibility (38, 40). Therefore, we modeled seipin structure with the residues ranging from Arg23 to Arg265 (Fig. 8-1) based on experimentally driven structural model (41) and our cryoelectron microscopy data that partially resolved the TM segments (Fig. 8-9). The detail of structural modeling is written in Methods.

How each part of seipin functions in LD biogenesis is not clearly known. To analyze the interactions between protein residues and lipids, we performed the AA simulation of human seipin in a 3-palmitoyl-2-oleoyl-D-glycero-1-phosphatidylcholine (POPC) bilayer containing 6% TG for 3  $\mu$ s, followed by grouping each lipid or protein residue as illustrated in Figs. 8-2a and 8-2b. The normalized coordination number by molecule or the coordination number per molecule ( $\|s\|$ ), which approximately indicates how much each protein residue prefers PL or TG (see Methods), was calculated (Fig. 8-2c). Such an analysis can be thought as the concentration-independent coordination number. The hydrophobic helix exhibited narrow spikes, indicating preferential aggregation with TG (Fig. 8-2d). S166 had the largest value in the analysis, consistent with other computational studies using CG simulations with the Shinoda-DeVane-Klein (SDK) or MARTINI force fields (42, 43). Although the modified parameters of TG have reduced charge distribution to reproduce the interfacial tension against water (69), the TG glycerol moiety can form hydrophilic interactions with protein residues in our AA simulation (e.g. S166). In contrast, the N- and C-terminus TM segments had weaker, however, broader attractions to TG. Because N- and C-terminal TM segments, and HH, have helical structures, the attraction map had a weakly-defined periodicity (Fig. 8-2d). For instance, V163, S166, and F170 faced the membrane center, increasing

the accessibility of TGs. In contrast, F164 and L168 faced the luminal interface, which prevented the interactions with TG.

We further compared protein residues' attractions to TG glycerol atoms or TG tail atoms by normalizing the coordination number by the number of atoms. While the coordination number per molecule ( $\langle |s| \rangle$ ) indicates a propensity of each protein residue for each lipid type, PL or TG (Figs. 8-2c and 8-2d), the coordination number per atom ( $\langle |s_A| \rangle$ ) provides a propensity for each atom type, in this case, a TG glycerol atom or TG tail atom (Fig 8-10). As expected, S166 had a strong interaction with TG glycerol moiety as they form a hydrophilic interaction (Fig. 8-10). The alignment of the insertion depths of S166 and TG glycerol moiety likely amplified the interaction. W257, which can form a hydrophilic interaction with TG glycerol moiety, had a high value as well (Fig. 8-10). However, we note that those results were normalized by the number of atoms. If we compare the coordination number of TG glycerol atoms and that of TG tail atoms, TG tail atoms will mostly have a higher value because there are 12 hydrophobic tail atoms and 1 glycerol moiety for each TG molecule. Therefore, while hydrophilic interactions at the hydrophobic phase are significant, the largest driving force of TG nucleation inside the seipin ring is provided by hydrophobic interactions of TG with seipin HH and N- and C-terminal TM segments.

The architecture of the seipin complex results in an unusual ring of TM helices, which could prevent the exchange of molecules between its interior and exterior. To study how molecules permeate through the dense TM region of the seipin ring, we analyzed the orientation of TM segments. The order parameter,  $S$ , was computed from the angle between each TM segment and the lower plane of the beta-sandwich region (Fig. 8-3a). If  $S$  is equal to 1, it represents the orientation of the TM segment vertical to the plane, and if  $S$  is -0.5, it is parallel to the plane. We also measured the distance ( $d$ ) between the center of the masses of the N and C-terminal TM

segments. Both N- and C-terminal TM segments showed a broad range of orientations (Fig. 8-3a). Especially, the N-terminal TM segment sampled a broader range of angles than the C-terminal TM segment, likely because the switch region near the C-terminal TM added some restraints on the movement. High flexibility of the TM segments was also supported by the calculation of the root-mean-square distance (RMSD) and root-mean-square fluctuation (RMSF) of  $\alpha$ -carbon atoms, shown in Figs. 8-11 and 8-12, respectively. When computing the RMSD of the  $\alpha$ -carbon atoms of the luminal domains, it leveled off to 0.2 nm. However, when computing the RMSD of the whole subunit, it had higher values with large error bars. Finally, the calculated RMSF showed high flexibility of both N- and C-terminal TM segments. The switch region itself was flexible (Phe220-Phe230) but restrained the movement of the C-terminal TM close to it.

To understand the collective motions of seipin, we aligned each subunit trajectory against its average structure and performed the principal component analysis (PCA) using the coordinates of  $\alpha$ -carbon atoms. Because the TM segments were most flexible in seipin, the dominating collective motions were related to the movements of seipin TM segments (Fig. 8-3b). For instance, the first principal motion with a variance of 34% was a swing-back-and-forth motion of N- and C-terminal TM segments together. The second principal motion with a variance of 18% was sliding of N- and C-terminal TM segments into the opposite directions. Interestingly, we observed that fluctuations and flexibility can open the space between monomers, significantly increasing lipid permeation through the protein-dense TM region (Fig. 8-3c). For instance, the orange monomer and purple monomer in the snapshots of Fig. 8-3c open to generate space in the cytosolic leaflet between them, which caused increased lipid influx (orange arrow). Similarly, the pink and dark green monomers had dispersed TM segments in the luminal leaflet, promoting lipid permeation (blue arrow). As TG is located closer to the membrane center than PL, TG permeation benefits



from opening of TM segments in either the cytosolic leaflet or luminal leaflet of the membrane. We also found the luminal leaflet had more limited PL permeability than the cytosolic leaflet because of the switch region. Thus, while seipin has a dense array of TM segments, their flexibility increases permeability of lipids in and out of the complex.

To understand how seipin influences the dynamics of lipids, we computed the position-dependent diffusion coefficient relative to the center of the mass of the luminal domain (Fig. 8-4). While all lipids had comparable diffusion coefficients in the protein-free region (7.5 nm - 10.0 nm), diffusion became slower near the TM segments and HH due to interactions with the protein. The slower diffusion near the TM region can also be seen in the lipid trajectories where the displacement became smaller near the TM segments (Fig. 8-3c). The decreased amount of diffusion coefficient is correlated with the contact area of protein. For instance, luminal PLs up to 7 nm from the seipin center had the lowest diffusion coefficients because of the HH and switch region in the luminal leaflet. In contrast, the cytosolic leaflet only contained the N- and C-terminal TM segments at the seipin boundary, leading to higher diffusion coefficients. The diffusion coefficients for TG were between those of cytosolic and luminal PLs because TG molecules close to the luminal leaflet can interact with the HH and switch region. In addition, strong attractions of TG with protein residues can further reduce the speed of diffusion (Fig. 8-2d). Due to confinement, the mean squared displacement (MSD) of the luminal PLs trapped inside the seipin ring, referred to as proteinized PLs, leveled off at later simulation times (Fig. 8-13). Such confinement can increase the bending modulus of this area, thereby working as a rigid base to ensure the direction of LD budding to the cytosolic side (326).

LD biogenesis is a microscopic/mesoscopic process with its time and length scales beyond those of AA simulations. For instance, our 3  $\mu$ s-long AA simulation demonstrated the recruitment

of TG inside the seipin complex. Yet, TG nucleation did not occur. To increase accessible time and length scales, we developed CG lipid and CG seipin models (Fig. 8-5). Linear, four-site models were used for lipids (305, 327). Every four protein residues were linearly mapped to one CG atom to match the resolution with CG lipids. Importantly, we placed 24 PL molecules inside the HH ring with the orientation consistent with other PL molecules in the luminal leaflet, referred to as proteinized PLs, because they were considered a part of a seipin oligomer. This is based on the AA simulation that demonstrated PLs inside the HH ring kept trapped, expectedly forming a rigid base. We constructed an elastic network model (ENM) by connecting a pair of seipin atoms via a harmonic bond with a constant spring constant ( $sc$ ) of 0.2 kcal/mol/Å<sup>2</sup> or 2.0 kcal/mol/Å<sup>2</sup> if the distance is less than 1.5 nm (Fig. 8-5b). To achieve the known stability of seipin in a bilayer membrane, nonbonded protein-lipid interactions were based on the lipid-lipid interactions with attraction scaling factors shown in Fig. 8-5c. Although it is difficult to quantitatively incorporate the AA simulation data into phenomenological models, higher scaling factors between TG-HH and TG-TM can be qualitatively justified by the high attractions of those pairs indicated in the analysis of the AA simulation (Fig. 8-2d).

Seipin is thought to catalyze TG nucleation, thereby decreasing the critical concentration (38, 39, 51, 202). To test this hypothesis, we performed the CG simulations of spherical bilayers with a diameter of 40 nm. The initial structures have evenly distributed 2% TG molecules. Because the TG concentration is below than the critical concentration (37), TG nucleation did not occur in the lipid system (Fig. 8-6). In contrast, the system that includes the seipin complex showed a nucleated TG lens inside the complex due to the attractions between TG-TM and TG-HH (Fig. 8-6).

To study the impact of the cage-like structure of the seipin oligomer and their TM segments in LD biogenesis, we simulated various geometries of seipin in the spherical bilayers containing 6% TG (Fig. 8-7). In the first model, we removed the TM segments, and the resulting model only contained the luminal domain. In the second model, we removed 6 continuous subunits from the seipin oligomer, and the resulting model contained the 5 subunits. As references, we also carried out simulations of the pure lipid system and the 11mer-containing system. The ENM with a spring constant of  $0.2 \text{ kcal/mol/\AA}^2$  was used. Because the TG concentration was above the critical concentration, TG nucleation occurred in those systems even in the system without seipin. However, the resulting morphologies of oil lenses of those systems were different as shown in the final snapshots and characterized by anisotropy. First, in the lipid system, a nucleated TG lens was flat and had high anisotropy to minimize the membrane deformation penalty (327). In contrast, in the 11-subunits model, a nucleated TG lens was located on top of the seipin luminal domain, surrounded by seipin TM segments. This resulted in a significant change in the shape of the oil lens from high anisotropy in the lipid system, minimizing the membrane deformation penalty, to low anisotropy in the seipin-containing systems. Given the nucleation percentages were comparable in those simulations, a change in anisotropy can be attributed to seipin TM, not to the amount of nucleated TG molecules. Importantly, the seipin TM segments constrained the area where TG can be in the bilayer, pushing excessive TG molecules to the LD. This results in the formation of the ER-LD neck structure, consistent with the experimental topology (313). An equilibrated diameter of the ER-LD structure (Fig. 8-14), approximated by a diameter of the circle formed by the end residues of N- and C-terminal TM segments, also agreed well with the experimentally measured data, which is in the range of 13 to 17 nm (313). We note the TM segments tilted away from the oligomeric center during LD growth, therefore, the diameter of the

seipin ring became increased with simulation times. We also simulated the 11-subunit model with a spring constant of 2.0 kcal/mol/Å<sup>2</sup> (Fig. 8-15). The higher spring constant reduced the diameter of the ER-LD neck structure, however, the nucleation percentage and morphology of the formed oil lens had little difference with those of the ENM with a spring constant of 0.2 kcal/mol/Å<sup>2</sup>.

The 5-subunits model can be considered a mixture of the lipid system and 11-subunit model because one end is occupied with seipin subunits while the other end is exposed to lipids. The resulting morphology of an oil lens was also between those results. The TG oil lens was elongated to the region where there was no seipin subunit, however, constrained in the region of seipin subunits, especially by their TM segments. The equilibrated anisotropy was also between the lipid and 11-subunits systems. Finally, we simulated the seipin model that only contained the luminal domain. Such a complex does not form a mobile focus in cells likely because it failed to form an oligomer, it was not stable in bilayers, or it was degraded (38). However, simulating this system can further provide valuable insights on the roles of the TM segments. The resulting oil lens showed little difference with the lipid system. The anisotropy was high, and the formation of the ER-LD neck structure was abolished.

To investigate the more advanced biogenesis steps, we simulated a larger spherical bilayer with a diameter of 60 nm, comparable to actual curvature of the ER tubule (328). The system contained 6% TG. We also constructed the heterogenous ENM (hENM) using the fluctuations obtained from the AA simulation of seipin in the bilayer membrane (329). The hENM had comparable fluctuations with the AA simulation in a bilayer (Figs. 8-8a and 8-8b). Consistent with the previous results, the seipin TM segments defined the oil boundary, facilitating the transport of TG into the LD (Fig. 8-8c). The equilibrated anisotropy was close to zero, indicating a spherical shape of the forming oil lens (Fig. 8-8d). Collectively, our tests demonstrated that the ring

geometry of seipin TM segments is key to defining the boundary of the forming oil lens and creating the unique ER-LD structure.

## DISCUSSION

LD biogenesis can be divided conceptually into three steps: TG nucleation, LD growth, and LD budding. Seipin is key to orchestrating this process by initiating LD biogenesis and allowing transfer of TG into a growing LD (38, 41-43, 202, 313, 325). Here, we investigated seipin-driven initial LD formation using AA and CG simulations. The AA simulations, carried out with modified parameters of TG that correctly reproduced the interfacial tension against water (69), demonstrated that seipin TM segments and HH attract TG. Consistent with prior studies (42), a broad range of the C-terminal TM residues attracted TG, whereas the HH had a narrow spike in the coordination analysis. Interestingly, we found most residues embedded in the hydrophobic phase of the bilayer prefer TG over PL. The preference of protein residues for TG is explained by the larger hydrophobicity of TG's three acyl chains compared to the two acyl chains of most PLs. In addition, although the partial charge distribution of the TG glycerol moiety is reduced with the modified parameters, it still forms a hydrophilic interaction with protein residues. A particular example was S166 having the largest coordination number per molecule with TG, consistent with the other computational reports that used the SDK or MARTINI force fields (42, 43).

We further partitioned protein-TG interactions into those caused by TG glycerol atoms or TG tail atoms. This analysis indicated that the TG glycerol moiety is more attracted to seipin residues than TG tail atoms; Overall, however, the contribution of TG tail atoms to the interaction was significantly larger than that of TG glycerol moiety because TG tail atoms outnumber TG glycerol atoms. Therefore, while TG's glycerol moiety in the membrane center can form

hydrophilic interactions with protein residues, the main driving force of TG nucleation is provided by hydrophobic interactions.

During LD growth, lipids or proteins such as *LiveDrop* migrate from the ER onto LD (66, 202). How lipids and proteins pass through the populated TM region is an intriguing question. In our AA simulation, we demonstrated that protein flexibility enabled various orientations of the TM region. The principal motion of seipin was the swing-back-and-forth motion of both N- and C-terminal TM segments as a bundle. This increases the space between TM segments of neighboring monomers, thereby increasing permeability to lipids. In contrast, the region where TM segments were packed showed limited lipid permeability. Given the low diffusion coefficients of lipids near the TM region, the opening of space may be important to facilitate lipid delivery to LDs. Also, given that proteins and peptides are usually larger than lipids, this may be more important for ER-to-LD targeting. For instance, the number of LDAF1 at the LD surface increases during LD growth, suggesting the protein continuously migrates from ER to the LD surface (38). Interestingly, as demonstrated in the CG simulations, the diameter of the ER-LD neck structure during the LD growth phase is larger than during the initial nucleation stage. The widely spread TM segments will promote lipid and protein influx to LDs. Furthermore, a recent study hypothesized that unfolding of the switch region during the LD growth, which would increase permeability in the luminal leaflet (41). The idea of stretching the switch region is partly supported by the high conformational fluctuations of this region.

While the AA simulation provided valuable insights into protein-lipid interactions, it did not show TG nucleation or LD growth stages because LD formation is a mesoscopic process. To overcome the limitations of time and length scales of AA simulations, we performed CG simulations in spherical bilayers. A bilayer containing 2% TG did not undergo TG nucleation

because the TG concentration was lower than the critical concentration for phase transition (37). However, seipin enabled TG nucleation in the same bilayer due to TG-TM and TG-HH attractions. In a bilayer containing 6% TG, above the critical concentration, both lipid and seipin-containing systems showed TG nucleation, however, with different morphologies of the forming oil lenses. While the lipid system demonstrated a flat oil lens to reduce membrane deformation (327), seipin TM segments set a boundary for the oil lens, effectively trapping TG inside the seipin and forming the ER-LD neck structure. Furthermore, the topology of the ER-LD contact was consistent with the recent studies with an agreement between the simulated and the experimentally measured diameters of the ER-LD neck structure (313).

To test the hypothesis in which seipin TM segments control the shape of the forming oil lens, we simulated the 6% mol TG bilayer with different topologies of seipin oligomer. When the seipin complex lacked the TM segments, the resulting oil lens was flat, and its anisotropy was high with little difference with the pure lipid system. When we removed the six continuous subunits from the seipin complex, the forming oil lens was elongated to the region absent of proteins. However, it was confined in the region of seipin subunits, especially by their TM segments. Those experiments demonstrate that the cage geometry of seipin TM segments constricts the LD area in the bilayer, promoting the unique ER-LD neck formation.

Collectively, our data reveal that seipin TM segments are key for LD biogenesis by nucleating TG, controlling lipid diffusion, and defining the boundary of the forming oil lens. Our study provides a better understanding of the roles of TM segments, critical for seipin functions.

## METHODS

### *Seipin structure*

We modeled a seipin structure (Arg23-Arg265) based on our cryoelectron microscopy data (Fig. 8-9). Our structure contained the luminal domain (Val60-His219), which was previously resolved (38, 40), and the partially resolved TM segments. However, due to the low resolution of the TM region, we were not able to identify residues in the TM segments. Instead, we used the orientation of the TM helices in our modeling. We assumed the residues that corresponded to the N- and C-terminal TM helical structures were Leu29-Gly50 and Ala235-Val258, respectively. The missing residues from Ser51 to Val60 were modeled with Modeller (330). The structure from Phe220 to Phe230, referred to as a switch region, was homology modeled using a yeast seipin structure as reference because this region is highly conserved and predicted to be folded similarly (41). The resulting structure of the switch region was helical, and its helicity was further supported by the PSIPRED (331), TMHMM (332), TMpred, and Phyre2 (333) servers. The missing residues, Pro231-Cys234, were modeled with Modeller (330).

### *AA simulation*

The seipin simulation in a POPC bilayer including 6% TG was carried out for 3  $\mu$ s. The equilibrated bilayer structure was taken from the previous work (69). Seipin has a HH ring at the center with a radius of  $\sim$ 2 nm in the luminal leaflet. We first placed 20 POPC molecules inside the seipin HH ring with their orientations consistent with other luminal POPC molecules using PACKMOL (60). Those PLs remained trapped inside the ring during the simulation, which we referred to as proteinized PLs in the CG model. We put the seipin complex at the membrane center



and removed any lipid molecules within 0.9 Å of seipin atoms. The equilibrium protocol provided by the CHARMM-GUI interface was used (56). Additionally, 100 ns of equilibration was carried out, restraining the positions of the backbone atoms of the luminal domain (Val60-His219) and the Z-positions of phosphorus atoms with a spring constant of 20 kJ/mol/nm<sup>2</sup>. The total numbers of POPC and TG molecules were 797 and 48, respectively.

The simulation was run by GROMACS 2020 (54) with a Lennard-Jones (LJ) cutoff-free version of C36 (57, 58). The modified TG parameters that reproduced the interfacial tension against water were used (69). Simulations were evolved with a 2-fs timestep. The long-range electrostatic and LJ interactions were evaluated with the Particle Mesh Ewald algorithm, with the real-space cutoff distance of 1.0 nm (119). Bond involving a hydrogen atom was constrained using the LINCS algorithm (223). A temperature of 310 K and a pressure of 1 bar were maintained with the Nose-Hoover thermostat and the Parrinello-Rahman barostat, respectively (120, 225, 226). The coupling time constants of 1 ps and 5 ps were used, respectively. A compressibility of  $4.5 \times 10^{-5}$  bar<sup>-1</sup> was used for semi-isotropic pressure coupling.

### *Coordination number analysis*

To study protein-lipid interactions, we first reduced the resolution of the AA simulation by mapping each POPC molecule into 11 molecular groups and each TG molecule into 13 groups. In this mapping scheme, each POPC molecule had choline head group, phosphate group, glycerol moiety and four tail groups for each acyl chain. Similarly, each TG molecule had glycerol moiety and four tail groups for each acyl chain. Each protein residue was mapped into one backbone atom and one side chain atom. For each amino acid, the coordination number between the side chain

atom and membrane atoms of PL or TG was calculated by  $s_M = \sum_M \sum_a [1 - (r_a/r_0)^6] / [1 - (r_a/r_0)^{12}]$ , where  $M$  represents PL or TG and  $a$  represents an atom belonged to  $M$ .  $r_0$  was set to 0.4 nm and  $r_a$  is the distance between the side chain atom and membrane atom (atom  $a$ ). The normalized coordination number by molecule or the coordination number per molecule ( $\|s\|$ ) was computed by dividing the coordination number by the number of molecules of PL or TG. The normalized coordination number by atom or the coordination number per atom ( $\|s_A\|$ ) was calculated by dividing the coordination number by the number of atoms of atom  $A$ .

### *Principal component analysis and diffusion coefficient*

Using the AA trajectory, we carried out the PCA and diffusion coefficient calculation with the MDAnalysis library (62, 334). For the PCA, each monomer trajectory was aligned with the initial monomer structure. The coordinates of the  $\alpha$ -carbon atoms were extracted from the aligned trajectory and used to compute the covariance matrix. The two primary collective motions with high variances were calculated. When calculating diffusion coefficient, we translated the system such that the center of the mass of the luminal domain of seipin was at the origin in each frame. Therefore, the diffusion coefficient of PL or TG reported here represents the relative diffusion coefficient with respect to the center of mass of the protein. The trajectory was divided into three trajectories, each of which was 1  $\mu$ s-long. In each trajectory, PL or TG molecules were categorized into three classes, based on the average XY-distance from the origin. The first class of lipids were located within 3.5 nm from the origin, slightly greater than the radius of the HH ring. The second class of lipids were located between 3.5 nm and 7.0 nm from the origin, where they mainly interacted with the TM segments. Finally, the lipids that were further than 7.0 nm were considered

lipids in the protein-free zone as they did not interact with the protein. The position-dependent diffusion coefficients were reported by calculating diffusion coefficients in each class. Three equal-length blocks were used to report the average and standard errors.

### *CG simulation*

We used a previously developed CG model for PL and TG with each molecule consisting of 4 beads (305, 327). An angle parameter of  $0.5 k_B T$  for PL was used in this study. A CG model for seipin was constructed by linearly mapping four amino acids into one CG bead. Such a resolution was chosen to match the resolution of lipids, preventing hydrophobic mismatch. We also placed 24 PL molecules inside the HH ring with their orientations consistent with the other PL molecules in the luminal leaflet. We referred to those as proteinized PLs because they were considered a part of a seipin oligomer. Three models were constructed with different elastic networks. The first two models utilized the ENM with a spring constant of  $0.2 \text{ kcal/mol } \text{\AA}^2$  or  $2 \text{ kcal/mol } \text{\AA}^2$  and with a distance cutoff of  $15 \text{ \AA}$ . The third model utilized the hENM that correctly represented the fluctuations of seipin in the underlying AA simulation (329). From the AA simulation, we first made a concatenated, aligned seipin monomer trajectory and obtained the hENM with a cutoff distance of  $12 \text{ \AA}$ . The hENM was applied to each monomer. A spring constant of  $0.1 \text{ kcal/mol } \text{\AA}^2$  was applied to the CG pairs that did not have hENM if the distance is less than  $11 \text{ \AA}$ . To achieve the known stability of seipin in a bilayer membrane, nonbonded protein-lipid interactions were based on the lipid-lipid interactions (305, 327). Protein atoms located in the hydrophobic phase interacted with the PL tail atoms with the equal attraction strength of the pair between PL tail and PL tail atoms or PL interfacial and PL interfacial atoms (scaling factor = 1). Protein atoms located at the membrane interface attracted PL interfacial atoms with a scaling factor of 1.5. The central

region of TM segments and two HH beads attracted TG atoms with a scaling factor of 1.5. Every CG bead carried a mass of 200 g/mol and no charge. Spherical bilayers with a diameter of 40 nm or 60 nm were constructed, containing 2% TG or 6% TG, followed by the placement of seipin and removal of lipids that had a close contact with seipin.

The CG simulations were run by LAMMPS (29 Oct 2020) with tabulated CG potentials (261). Simulations were evolved with a 50-fs timestep. Temperature was maintained at 310 K by the Langevin thermostat with a coupling constant of 100 ps (306). The cutoff distance of nonbonded interaction was 1.5 nm. The initial structures of CG simulations were prepared with the MDAnalysis library (62).

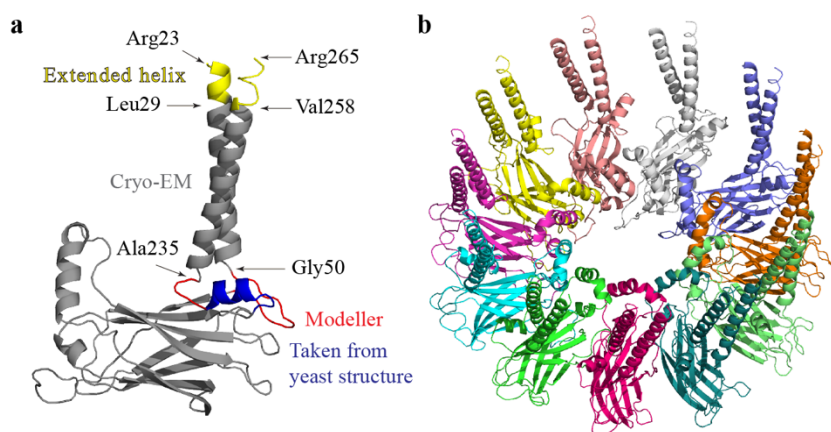
#### *Nucleation percentage and anisotropy*

We calculated the nucleation percentage and anisotropy as explained in (327). In short, the nucleation percentage was defined as the ratio of the number of TG molecules in the largest cluster to the number of TG molecules in a system. The distance cutoff of 2 nm was used for clustering TG molecules. The anisotropy was calculated by diagonalizing the moment of inertia tensor of the largest TG cluster. The anisotropy of 0 represents a spherical shape, and the anisotropy of 0.25 does a planar shape.

#### **ACKNOWLEDGEMENTS**

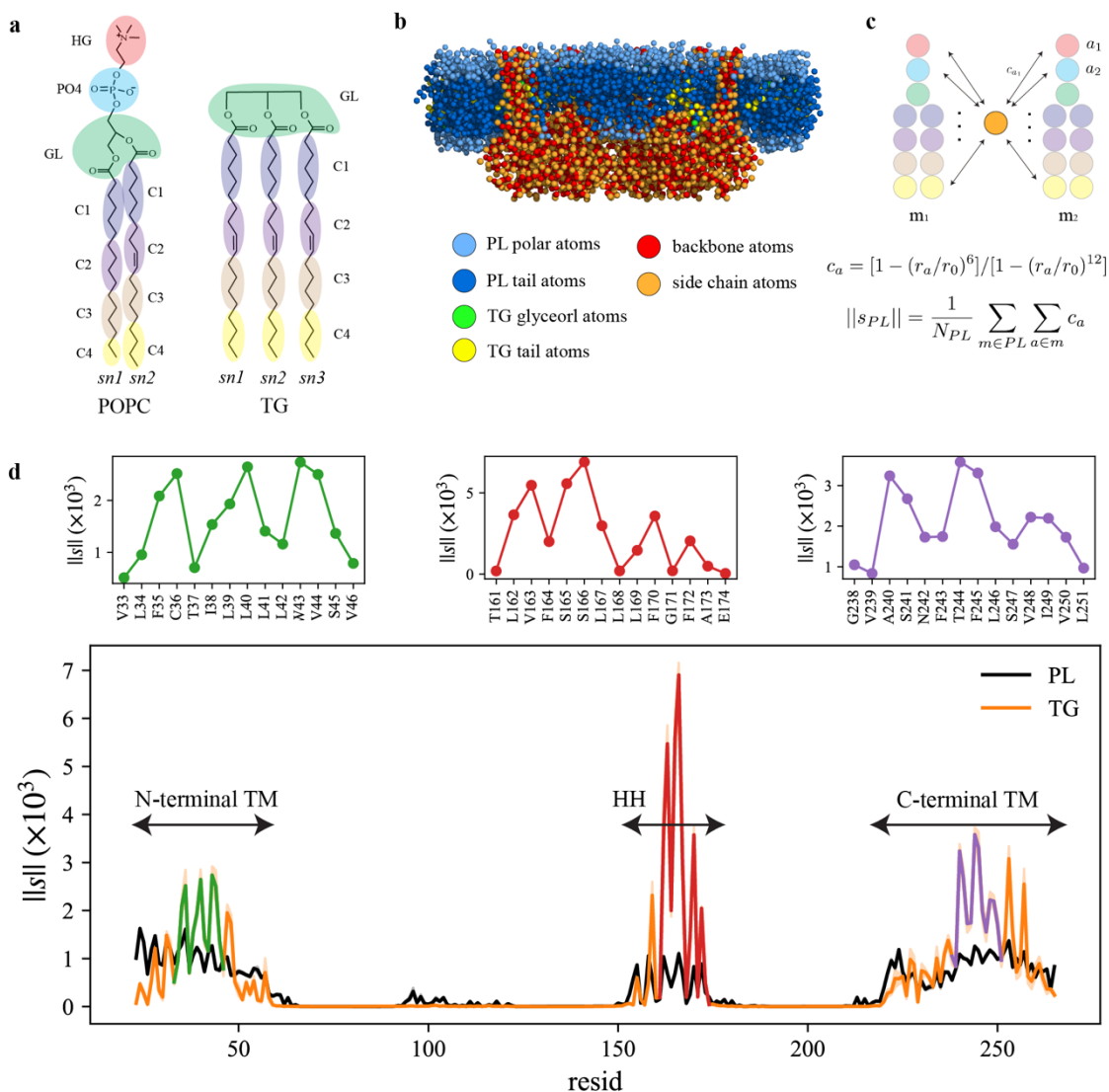
We thank Xudong Wu for performing cryoelectron microscopy imaging and critical discussion. This research was supported by a grant from the National Institute of General Medical Science (NIGMS) of the National Institutes of Health (NIH), grant R01-GM063796. The computer

simulations were performed on the Stampede2 supercomputer at the Texas Advanced Computing Center and the Bridges2 supercomputer at the Pittsburgh Supercomputing Center (PSC) through allocation MCA94P017 with resources provided by the Extreme Science and Engineering Discovery Environment (XSEDE) supported by NSF grant ACI-1548562. We also appreciate the computational resources on the Midway3 supercomputer at the University of Chicago. S.K. acknowledges Chenghan Li and Sriramvignesh Mani for critical discussion.



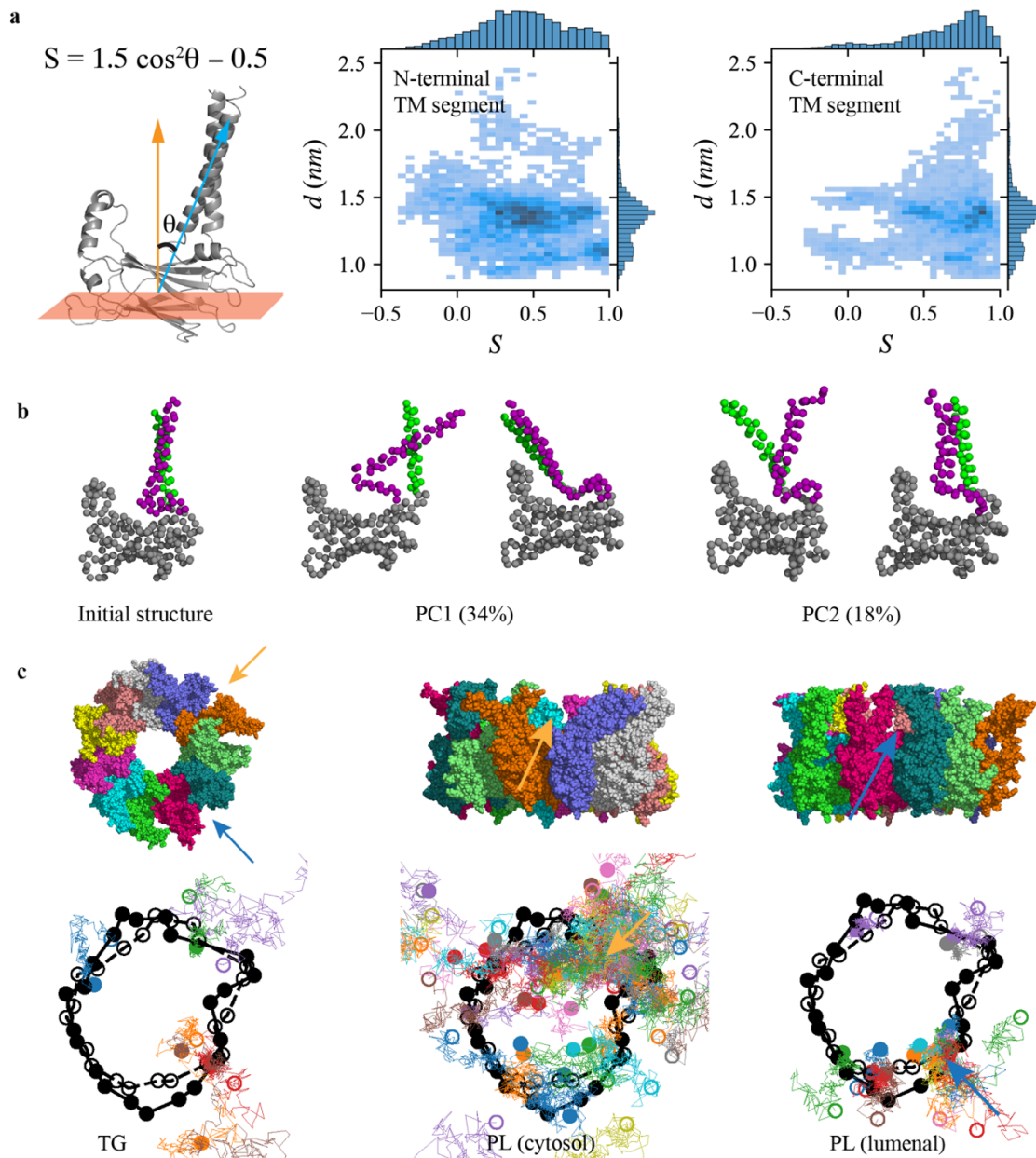
**Figure 8-1** Structural model of human seipin.

(a) Structure of a seipin subunit. The structure included in the cryoelectron microscopy data is shown in gray. Red loops were modelled with Modeller (330). The blue region was predicted using the yeast structure. The helical structures were extended (yellow). (b) Structure of the human seipin oligomer used in simulations. Each chain is shown with different colors.



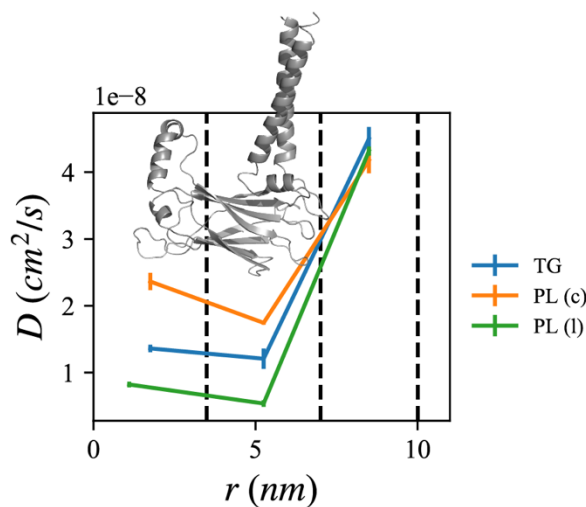
**Figure 8-2** Seipin hydrophobic helix and transmembrane segments attract TG.

(a) Molecular groupings of lipids. Each protein residue was mapped onto one side chain and one backbone atom. (b) Initial structure of the system at the reduced resolution. The snapshot was clipped in the XZ plane. (c) Illustration of the calculation of the coordination number per molecule. A side chain atom was depicted with orange circle and PL atoms with other colors. (d) Interaction plot of protein residues with PL (black) and TG (orange). The shaded area represents the standard error of the results of three equal-length blocks, each containing 1- $\mu$ s trajectory. The residues that had high interactions with TG in the N-terminal segment, HH, and C-terminal segment were shown in separate plots in the upper panel, colored with green, red, and purple lines, respectively.



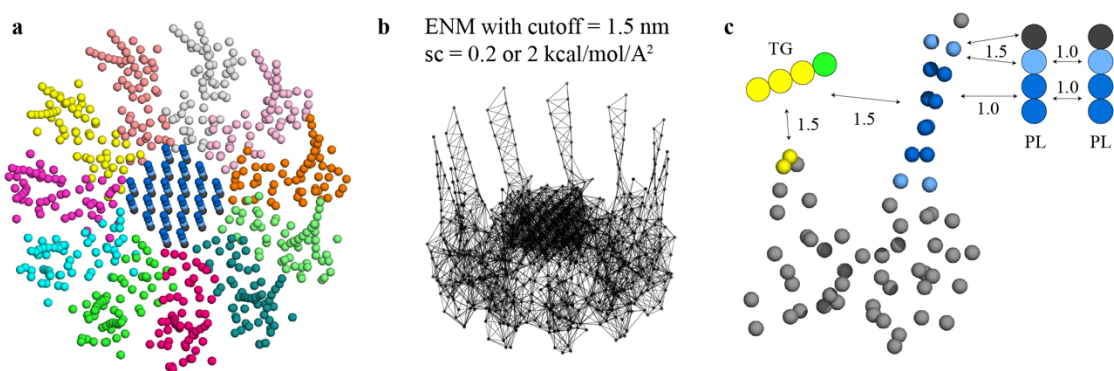
**Figure 8-3** Flexibility of seipin TM segments increases lipid permeability.

(a) Illustration of the angle ( $\theta$ ) of N- or C-terminal TM segments with the lower beta-sandwich plane and distance ( $d$ ) between the center of masses of N- and C-terminal segments. (b) Two principal collective motions with variances of 34% and 18%, respectively. (c) Snapshots of the last frame in top and side view (top). The XY positions of TGs, cytosolic PLs, and luminal PLs that permeate through the TM region are shown every 5 ns with colored lines (bottom). Black markers indicate the XY positions of N- and C-terminal TM segments. Filled circles indicate the positions in the last frame and unfilled circles in the first frame.



**Figure 8-4** Position-dependent diffusion coefficient. The center of the mass of the luminal domain of seipin is at origin.

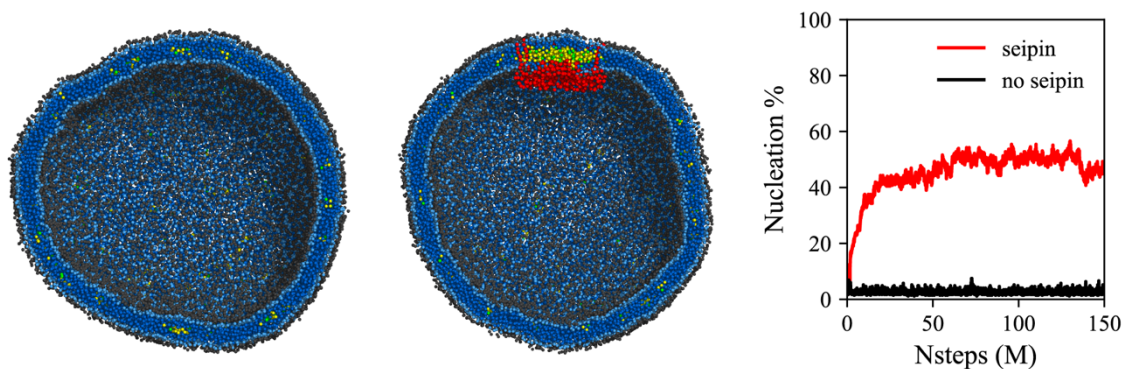
The first region (0-3.5 nm) contains the seipin hydrophobic helices and the second region (3.5 nm - 7.0 nm) the TM segments. The third region (7.0-10.0 nm) is protein-free. Diffusion coefficients of TG, cytosolic PL, and luminal PL are shown with blue, orange, and green lines, respectively. The error bar represents the standard error of the results of three equal-length blocks, each containing 1  $\mu$ s.



**Figure 8-5** CG models of seipin and lipids.

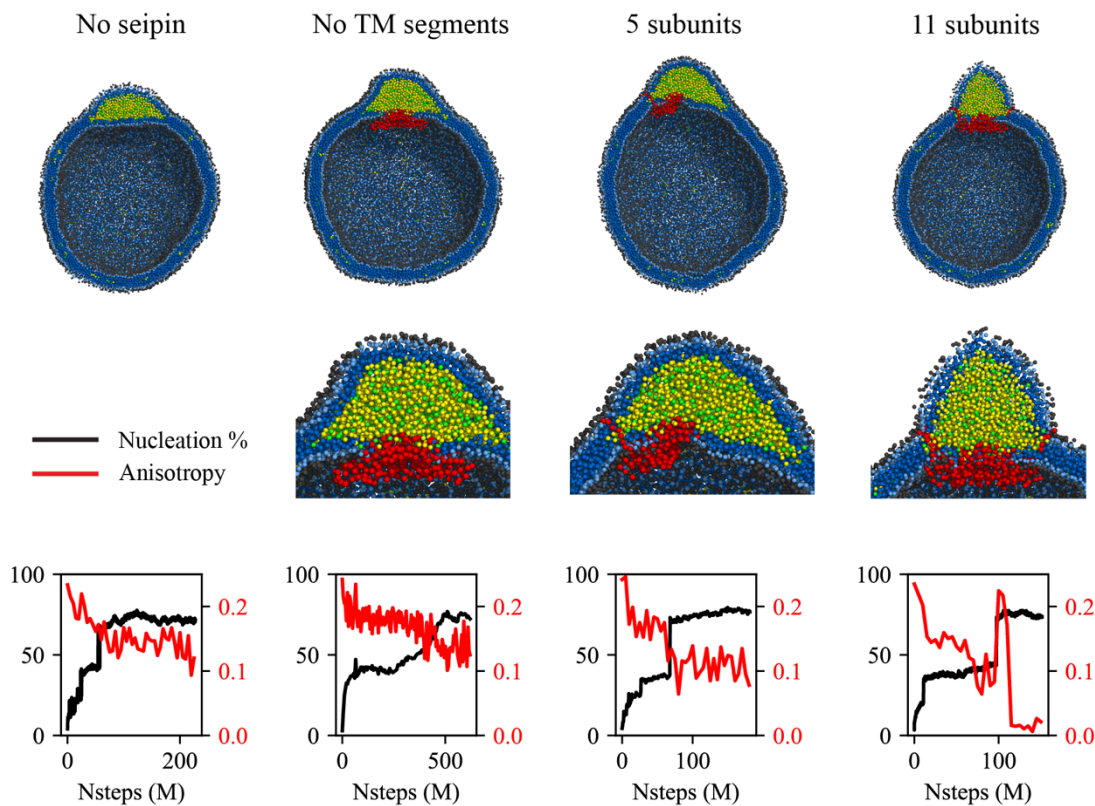
(a) CG model of human seipin oligomer. The CG atoms inside the hydrophobic helix ring represent PL atoms, referred to as proteinized PLs. (b) Elastic network model (ENM) with a spring constant of 0.2 or 2 kcal/mol/Å<sup>2</sup>. (c) Scaling factors of attraction parameters between seipin-PL and seipin-TG interactions. PL head, interfacial, and tail groups are shown with black, light blue, and dark blue, respectively. TG glycerol and tail groups are shown with green and yellow, respectively. Seipin atoms that attract PL tails are shown with dark blue, and those that attract PL interfacial atoms are shown with sky blue. Two seipin atoms in the HH, shown with yellow, and four seipin atoms in each TM segment, shown with dark blue, attract TG atoms.





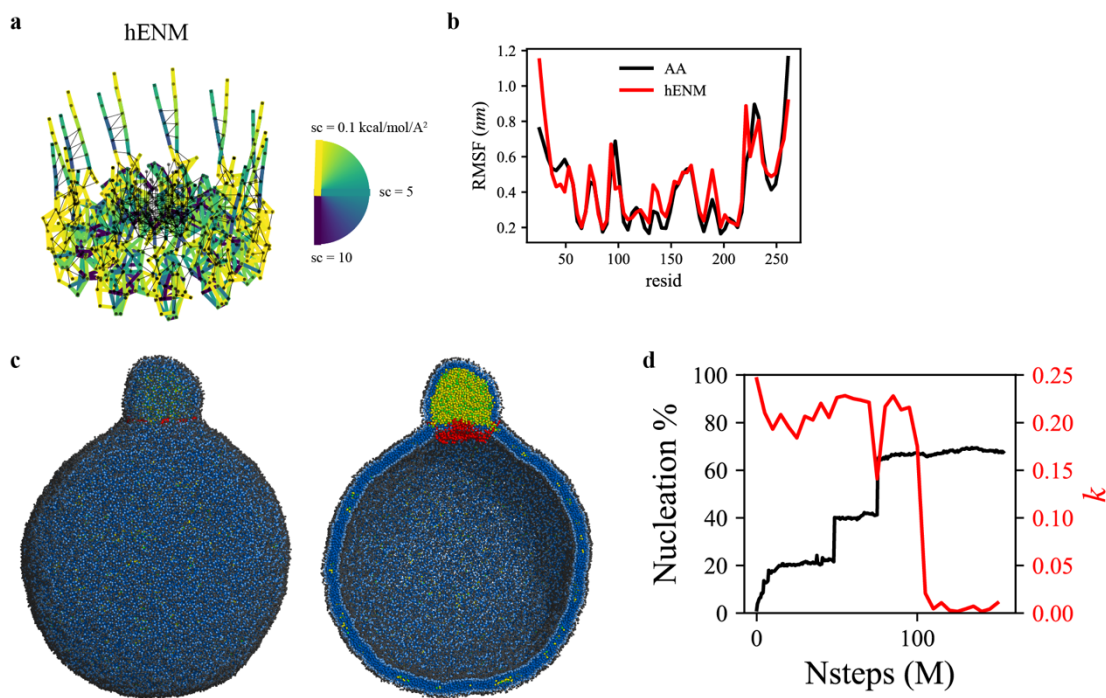
**Figure 8-6** Seipin lowers the critical concentration of TG nucleation.

CG simulations of bilayers containing 2% TG with a diameter of 40 nm were carried out. The clipped snapshots of the last frames of the pure lipid (left) and seipin-containing systems (center) are shown. The ENM model with a spring constant of  $0.2 \text{ kcal/mol/\AA}^2$  was used. PL head, interfacial, and tail atoms are shown with black, light blue, and dark blue, respectively. TG glycerol and tail atoms are shown with green and yellow, respectively. Seipin oligomer is indicated with red. The nucleation percentages of those systems with simulation steps are shown in right.



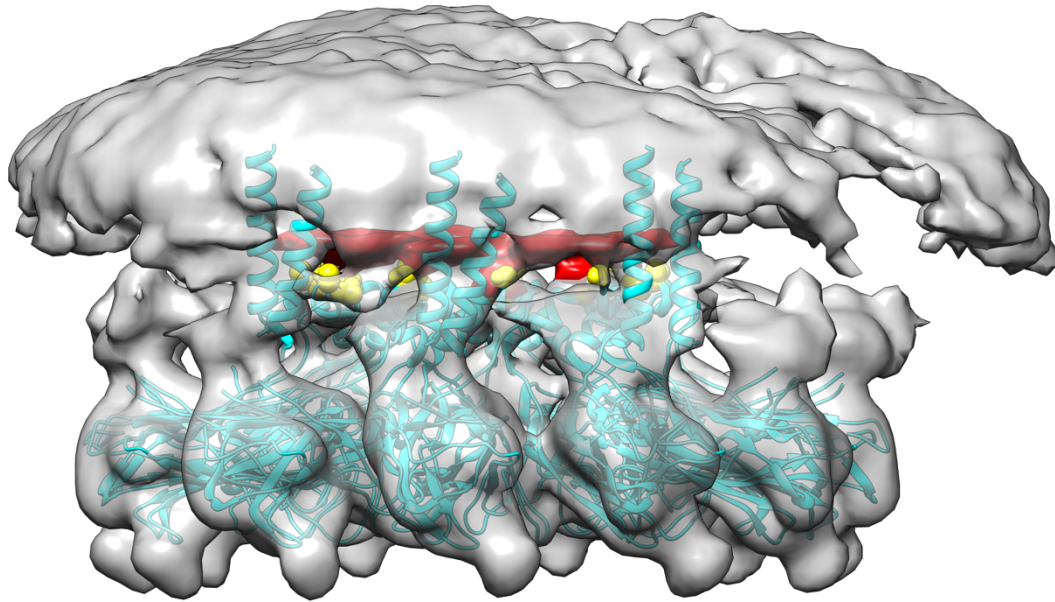
**Figure 8-7** Cage-like geometry of seipin TM segments is key to modulating the morphology of a forming oil lens.

The first row shows the clipped snapshots of the last frames, and the second row shows the close-up view of the seipin. The same color code is as in Fig. 8-6. The third row shows the nucleation percentage (black) and anisotropy (red). CG simulations of spherical bilayers containing 6% TG with a diameter of 40 nm were carried out. The ENM model with a spring constant of 0.2 kcal/mol/Å<sup>2</sup> was used.



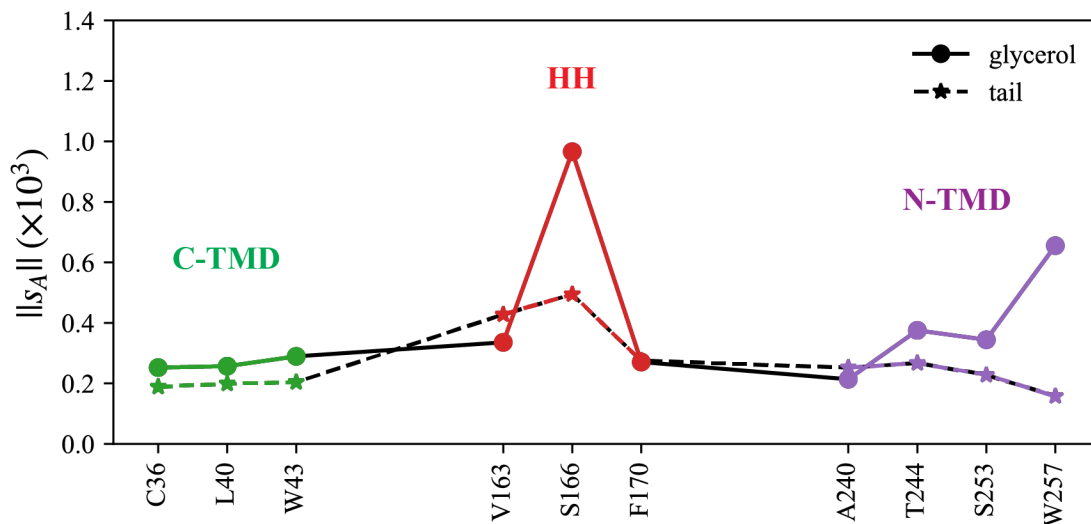
**Figure 8-8** CG-MD shows LD growth in a large bilayer with its curvature comparable to the actual ER tubule.

CG simulations of spherical bilayers containing 6% TG with a diameter of 60 nm were carried out. (a) Heterogeneous ENM (hENM) model of human seipin was constructed. Pairs of atoms in a subunit were connected with harmonic springs with their spring constants (sc) represented by their color. Additional harmonic springs (black lines) with a spring constant of 0.1 kcal/mol/Å<sup>2</sup> were added between pairs of atoms that were not included in the hENM with a distance cutoff of 11 Å to ensure connections between subunits. (b) RMSF of a seipin subunit was compared between the AA trajectory and hENM in a bilayer. (c) Exterior and interior view of the last frame. The color code is as in Fig. 8-6. (d) The nucleation percentage (black lines) and anisotropy (red lines) are shown.



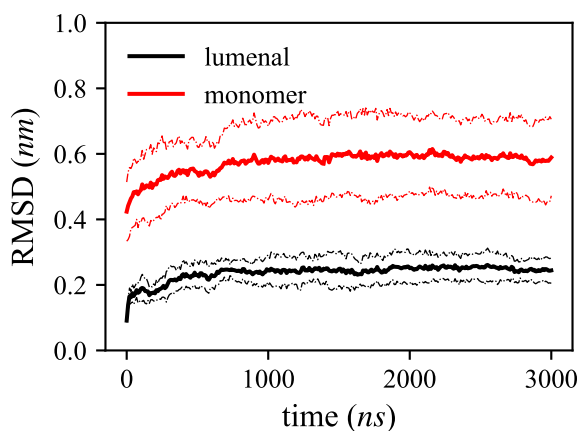
**Figure 8-9** Cryoelectron microscopy of human seipin.

The cyan model includes the seipin luminal domain and partially resolved transmembrane segments. The electron density of Ser165 and Ser166 is shown with yellow. The unidentified density that interacts with Ser165 and Ser166 is shown with red.

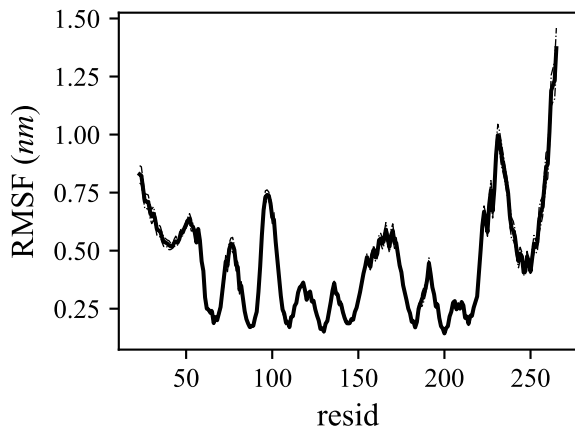


**Figure 8-10** Normalized coordination number by atom.

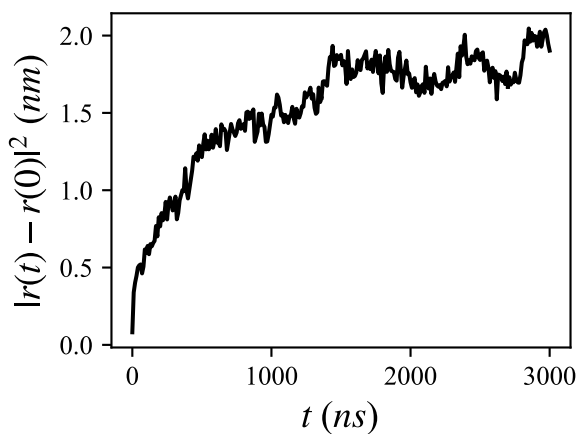
The interactions with TG glycerol moiety are shown with a continuous line and circle markers and those with TG tail atoms are shown with a dashed line and star markers. Related to Fig. 8-2.



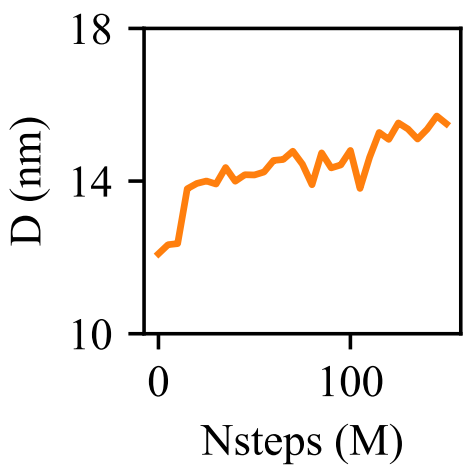
**Figure 8-11** Root-mean-square distance (RMSD) of the  $\alpha$ -carbon atoms of the luminal domain (black) or the whole subunit (red). The error bars represent the standard deviation of the RMSDs of 11 subunits.



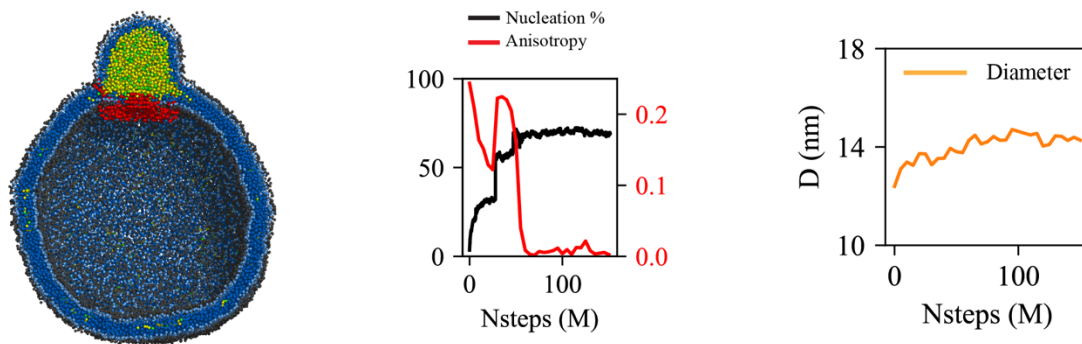
**Figure 8-12** Root-mean-square fluctuation (RMSF) of the  $\alpha$ -carbon atoms. The error bars represent the standard error of the RMSFs of three blocks.



**Figure 8-13** Mean squared distance of the 20 luminal PLs, trapped inside the hydrophobic helix.



**Figure 8-14** CG simulation of the bilayer containing 6% TG with a diameter of 40 nm. The ENM of seipin used a spring constant of  $0.2 \text{ kcal/mol/\AA}^2$ . The diameter of an oil lens is shown with simulation steps.



**Figure 8-15** CG simulation of the bilayer containing 6% TG with a diameter of 40 nm.

The ENM of seipin used a spring constant of  $2 \text{ kcal/mol/\AA}^2$ . The clipped snapshot is shown in left, the nucleation percentage (black) and anisotropy (red) are shown in middle, and the diameter of an oil lens (orange) is shown in right.

## **FUTURE DIRECTIONS**

Long considered tangential by cell biologists, LDs have recently gained the attention that they deserve (200). The amount of LD literature and the number of scientists who seek to answer the fundamental biophysics of LDs continue to increase. In this thesis, I have aimed to study a broad range of topics on LD biophysics, LD targeting, and LD biogenesis. However, the work is far from finished and many questions remain unanswered. This chapter will discuss possible future work, with a focus on computational, biological, and biophysical aspects of LDs.

First, the all-atom force field of TG needs to be developed to improve the quality of computational LD studies. As discussed in the Chapter 1, the current force fields used to describe TG fail to simultaneously reproduce three key properties: interfacial tension at the TG/water interface, the area per phospholipid of the LD surface, and LD surface tension. In Chapter 6, we have developed the new TG model that reproduces the interfacial tension at the TG/water interface by reducing the charge distribution of the TG glycerol moiety. The recent work by the Vanni group also has used the same approach to reproduce the interfacial tension, but more systematically using a gradient-based iterative optimization (335). However, simulations of LD surfaces using the new force fields predict smaller number of TG molecules whose glycerol moieties are exposed to water and hence a low area per phospholipid of the LD surface. To resolve this issue, one can imagine re-parameterizing the Lennard-Jones parameters between the water and TG glycerol moiety to incorporate polarization effects rather than changing the distribution of partial charges. This approach was used in the past to match the transfer free energy of a water molecule from water to hexadecane (137). Another approach to solve this issue would be to use a polarizable force field, which directly accounts for polarization effects.



Moreover, it would be beneficial to study how neutral lipids modulate the physical properties of LD cores and LD surfaces. Even though TG is used as a model neutral lipid in this thesis, cholesteryl ester (CE) is another major component of the LD core in cellular systems. Recent experiments showed that the phase of the LD core can change depending on the cellular state (336). Under starvation, TG in the LD core is preferentially hydrolyzed, increasing the CE concentration, which leads to a phase transition of the LD core from amorphous to liquid-crystalline phase. Also, the same study has demonstrated that the smectic liquid-crystalline phase has an onion-like structure with a lattice spacing of 3.5 nm (336). Unpublished MD simulations (carried out by Jay Braun and Jessica Swanson) compared the physical properties of the TG-rich LD and CE-rich LD: the former has an amorphous phase while the latter has a lipid liquid-crystalline phase, consistent with the experiments. The CE-rich LD simulations also indicated that PL monolayers are more packed and ordered than the TG-rich LD. Such dramatic changes in the LD core and LD surface alter the protein composition at the LD surface. For instance, recent experiments found that *LiveDrop*, known to target from the ER bilayer to the TG-rich LD (see Chapter 4), relocates to the ER under starvation because the principal component of the LD core becomes CE as TG is hydrolyzed (337).

The atomistic details of LD targeting mechanism remains largely unanswered. We have studied two LD targeting pathways, ER-to-LD (Chapter 4) and cytosol-to-LD (Chapter 5), with the model peptides, *LiveDrop* and the autoinhibitory motif of CCT, respectively. However, the conclusions that I drew from those studies were made for simplistic systems. The model systems could be improved to better mimic the cellular conditions. For instance, in cells, ER and LD are connected by seipin. Simulation of both ER and LD linked through seipin could allow us to observe how *LiveDrop* permeates through the seipin transmembrane-dense area. Furthermore, there are

many fundamental and valuable questions one can ask about CCT: a) how LD surfaces make the CCT's autoinhibitory motif released from its catalytic domain, b) how CCT exists as a dimer, and c) how CCT modulates the curvature (64, 218, 219, 338).

In addition to CCT and *LiveDrop*, there are many other proteins with unique amphipathic features that target LD surfaces. For conciseness, I will only discuss proteins that target from the cytosol via their amphipathic helices in this section. CGI-58 is a co-activator of adipose triglyceride lipase that has a strong binding affinity with LD surfaces (75). Its targeting motif is a short amphipathic helix with three tryptophan residues. A recent study has shown that mutating the tryptophan residues to alanine inhibits LD targeting (234, 235). The binding affinity calculation of the peptide and the mutated peptide (tryptophan to alanine) would support the importance of tryptophan residues in LD targeting as discussed in Chapter 4 and 5. Another example of LD surface targeting protein is ArfGAP1, although why it binds to the LD surface remains unclear. As discussed in Chapter 1, the amphipathic helix of ArfGAP1 has a small amount of charged residues in the polar face, which increases the sensitivity to the hydrophobic face. Therefore, ArfGAP1 binds to curved bilayers (339, 340). Given that the LD curvature is small because of its large size, it would be intriguing to address why ArfGAP1 sits at the LD surface. Finally, perilipin family is a class of protein that prevents the neutral lipid hydrolysis in LDs (23). Instead of hydrophobic residues that favor the LD monolayer (e.g., tryptophan), perilipin family has a long repetitive amphipathic helix that promotes LD targeting. Shortening the helix reduces the efficiency of LD targeting (84). Also, the perilipin family has specificity in LD targeting: some members target TG-rich LD surfaces while some target CE-rich LD surfaces (209). How perilipin amphipathic helices achieves specificity in LD targeting is an interesting subject.

There are many fundamental questions remained unanswered in LD biogenesis. The first question is how different cell types control the neutral lipid composition of the LD core. For instance, adipose cells typically have TG-rich LDs while macrophages have CE-rich LDs. It is worth studying whether the differences in the neutral lipid composition of the LD core can be attributed to different activation levels of neutral lipid synthesis or a different LD biogenesis mechanism.

The second aspect is on the binding partner of seipin. Human seipin has a binding partner called lipid droplet assembly factor 1 (LDAF1) that catalyzes the initial steps of LD formation (38). Together with LDAF1, human seipin defines the LD formation site. LDAF1 is predicted to have a membrane-embedded, double hairpin topology with the ends exposed to the cytosol (38). According to the unidentified density map in the cryo-electron microscopy, it is likely that the luminal hydrophobic helices of human seipin (especially S165 and S166) are in a proximity with the kink of LDAF1 (38). It is unclear how many LDAF1 subunits exist in each seipin oligomer. However, it is expected to be more than one subunit based on the cryo-electron microscopy. Therefore, the human seipin complex has a remarkably high density of transmembrane segments because of seipin and LDAF1.

Given that most hydrophobic residues embedded in the hydrophobic phase prefers TG over PL (see Chapter 4 and 8), one can expect that the high density of transmembrane segments facilitate TG nucleation. However, how the human seipin complex works in a molecular scale, how it facilitates TG nucleation and LD growth, how it interacts with the lipids, and how the interactions between the human seipin complex and lipids are different from those between the human seipin alone and lipids should be further addressed. In the absence of LD, LDAF1 is localized at the human seipin oligomer. However once LD forms, LDAF1 migrates towards the

growing LD surface. A study of how LDAF1 changes its conformation at two different surfaces and how LDAF1 migrates from the seipin complex to LD would be beneficial. The approach that was used to estimate LD accumulation energy in Chapter 4 could be helpful in answering these questions.

In yeast, the binding partner of seipin becomes more critical. While human seipin is sufficient for LD formation (i.e., without LDAF1, human seipin alone can make LDs), the binding partner of yeast seipin, called Ldb16, is necessary for LD biogenesis (318, 341). The difference between human and yeast seipin in their ability to concentrate TG can be found in their structural difference. Yeast seipin lacks the hydrophobic helices positioned toward the luminal leaflet that have the highest attraction with TG (Chapter 8) (41, 342). Therefore, yeast seipin needs its binding partner that can replace the missing hydrophobic helices to be functional. Comparison of human and yeast seipin should be followed after structural and functional understanding of Ldb16. Furthermore, one could study the critical roles of the evolutionarily conserved transmembrane segments of seipin. Specifically, it would be helpful to investigate why chimeric seipin proteins in which two transmembrane segments were replaced with another ER protein FIT2 were not functional despite the high density of transmembrane segments (38).

Finally, a bottom-up coarse-grained model of lipids and seipin would be worth pursuing. Top-down coarse-graining models successfully modeled TG nucleation and LD growth in Chapter 7 and 8. Although those models were inspired from the all-atom simulations and qualitatively used the analysis of protein-lipid interactions from the all-atom simulations, a rigorous and systematic bridge between coarse-grained and all-atom simulations would be of great importance.

## REFERENCES

1. L. L. M. van Deenen, Phospholipids and biomembranes. *Progress in the Chemistry of Fats and other Lipids* **8**, 1-127 (1966).
2. P. R. Cullis, M. J. Hope, C. P. S. Tilcock, Lipid polymorphism and the roles of lipids in membranes. *Chemistry and Physics of Lipids* **40**, 127-144 (1986).
3. H. Sprong, P. van der Sluijs, G. van Meer, How proteins move lipids and lipids move proteins. *Nat Rev Mol Cell Biol* **2**, 504-513 (2001).
4. G. van Meer, D. R. Voelker, G. W. Feigenson, Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* **9**, 112-124 (2008).
5. G. van Meer, A. I. de Kroon, Lipid map of the mammalian cell. *J Cell Sci* **124**, 5-8 (2011).
6. T. Harayama, H. Riezman, Understanding the diversity of membrane lipid composition. *Nature Reviews Molecular Cell Biology* **19**, 281-296 (2018).
7. Y. Yang, M. Lee, G. D. Fairn, Phospholipid subcellular localization and dynamics. *J Biol Chem* **293**, 6230-6240 (2018).
8. M. Doktorova, J. L. Symons, I. Levental, Structural and functional consequences of reversible lipid asymmetry in living membranes. *Nature Chemical Biology* **16**, 1321-1330 (2020).
9. J. E. Johnson, R. B. Cornell, Amphitropic proteins: regulation by reversible membrane interactions (review). *Mol Membr Biol* **16**, 217-235 (1999).
10. R. B. Cornell, S. G. Taneva, Amphipathic helices as mediators of the membrane interaction of amphitropic proteins, and as modulators of bilayer physical properties. *Curr Protein Pept Sci* **7**, 539-552 (2006).

11. G. Drin, B. Antonny, Amphipathic helices and membrane curvature. *FEBS Lett* **584**, 1840-1847 (2010).
12. B. Antonny, Mechanisms of membrane curvature sensing. *Annu Rev Biochem* **80**, 101-123 (2011).
13. J. Bigay, B. Antonny, Curvature, lipid packing, and electrostatics of membrane organelles: defining cellular territories in determining specificity. *Dev Cell* **23**, 886-895 (2012).
14. R. B. Cornell, Membrane lipid compositional sensing by the inducible amphipathic helix of CCT. *Biochim Biophys Acta* **1861**, 847-861 (2016).
15. G. Boden, Obesity and free fatty acids. *Endocrinol Metab Clin North Am* **37**, 635-646, viii-ix (2008).
16. A. S. Greenberg *et al.*, The role of lipid droplets in metabolic disease in rodents and humans. *J Clin Invest* **121**, 2102-2110 (2011).
17. G. Onal, O. Kutlu, D. Gozuacik, S. Dokmeci Emre, Lipid Droplets in Health and Disease. *Lipids Health Dis* **16**, 128 (2017).
18. N. L. Gluchowski, M. Becuwe, T. C. Walther, R. V. Farese, Jr., Lipid droplets and liver disease: from basic biology to clinical implications. *Nat Rev Gastroenterol Hepatol* **14**, 343-355 (2017).
19. S. Xu, X. Zhang, P. Liu, Lipid droplet proteins and metabolic diseases. *Biochim Biophys Acta Mol Basis Dis* **1864**, 1968-1983 (2018).
20. N. A. Ducharme, P. E. Bickel, Minireview: Lipid Droplets in Lipogenesis and Lipolysis. *Endocrinology* **149**, 942-949 (2008).
21. T. C. Walther, R. V. Farese, Jr., Lipid droplets and cellular lipid metabolism. *Annu Rev Biochem* **81**, 687-714 (2012).

22. H. Wang, M. V. Airola, K. Reue, How lipid droplets "TAG" along: Glycerolipid synthetic enzymes and lipid storage. *Biochim Biophys Acta Mol Cell Biol Lipids* **1862**, 1131-1145 (2017).
23. C. Sztalryd, D. L. Brasaemle, The perilipin family of lipid droplet proteins: Gatekeepers of intracellular lipolysis. *Biochim Biophys Acta Mol Cell Biol Lipids* **1862**, 1221-1232 (2017).
24. K. Bersuker *et al.*, A Proximity Labeling Strategy Provides Insights into the Composition and Dynamics of Lipid Droplet Proteomes. *Dev Cell* **44**, 97-112 e117 (2018).
25. J. A. Olzmann, P. Carvalho, Dynamics and functions of lipid droplets. *Nature Reviews Molecular Cell Biology* **20**, 137-155 (2019).
26. M. A. Roberts, J. A. Olzmann, Protein Quality Control and Lipid Droplet Metabolism. *Annual Review of Cell and Developmental Biology* **36**, 115-139 (2020).
27. Z. Cao, H. Y. Mak, Married at Birth: Regulation of Cellular Fat Metabolism by ER–Lipid Droplet Crosstalk. *Contact* **3**, 251525642093467 (2020).
28. N. Kory, R. V. Farese, Jr., T. C. Walther, Targeting Fat: Mechanisms of Protein Localization to Lipid Droplets. *Trends Cell Biol* **26**, 535-546 (2016).
29. K. Bersuker, J. A. Olzmann, Establishing the lipid droplet proteome: Mechanisms of lipid droplet protein targeting and degradation. *Biochim Biophys Acta Mol Cell Biol Lipids* **1862**, 1166-1177 (2017).
30. R. Dhiman, S. Caesar, A. R. Thiam, B. Schrul, Mechanisms of protein targeting to lipid droplets: A unified cell biological and biophysical perspective. *Semin Cell Dev Biol* **108**, 4-13 (2020).

31. M. J. Olarte, J. M. J. Swanson, T. C. Walther, R. V. Farese, Jr., The CYTOLD and ERTOLD pathways for lipid droplet-protein targeting. *Trends Biochem Sci* 10.1016/j.tibs.2021.08.007 (2021).
32. R. M. Venable, A. Kramer, R. W. Pastor, Molecular Dynamics Simulations of Membrane Permeability. *Chem Rev* **119**, 5954-5997 (2019).
33. M. A. Mitsche, L. Wang, D. M. Small, Adsorption of egg phosphatidylcholine to an air/water and triolein/water bubble interface: use of the 2-dimensional phase rule to estimate the surface composition of a phospholipid/triolein/water surface as a function of surface pressure. *J Phys Chem B* **114**, 3276-3284 (2010).
34. A. Chorlay, A. R. Thiam, Neutral lipids regulate amphipathic helix affinity for model lipid droplets. *J Cell Biol* **219** (2020).
35. K. Ben M'barek *et al.*, ER Membrane Phospholipids and Surface Tension Control Cellular Lipid Droplet Formation. *Dev Cell* **41**, 591-604 e597 (2017).
36. A. Bacle, R. Gautier, C. L. Jackson, P. F. J. Fuchs, S. Vanni, Interdigitation between Triglycerides and Lipids Modulates Surface Properties of Lipid Droplets. *Biophys J* **112**, 1417-1430 (2017).
37. J. A. Hamilton, D. M. Small, Solubilization and localization of triolein in phosphatidylcholine bilayers: a <sup>13</sup>C NMR study. *Proc Natl Acad Sci U S A* **78**, 6878-6882 (1981).
38. J. Chung *et al.*, LDAF1 and Seipin Form a Lipid Droplet Assembly Complex. *Dev Cell* **51**, 551-563 e557 (2019).
39. X. Sui *et al.*, Cryo-electron microscopy structure of the lipid droplet-formation protein seipin. *J Cell Biol* **217**, 4080-4091 (2018).



40. R. Yan *et al.*, Human SEIPIN Binds Anionic Phospholipids. *Dev Cell* **47**, 248-256 e244 (2018).
41. H. Arlt *et al.*, Seipin forms a flexible cage at lipid droplet formation sites. *bioRxiv* 10.1101/2021.08.05.455270, 2021.2008.2005.455270 (2021).
42. V. Zoni *et al.*, Seipin accumulates and traps diacylglycerols and triglycerides in its ring-like structure. *Proc Natl Acad Sci U S A* **118**, e2017205118 (2021).
43. X. Prasanna *et al.*, Seipin traps triacylglycerols to facilitate their nanoscale clustering in the endoplasmic reticulum membrane. *PLoS Biol* **19**, e3000998 (2021).
44. A. R. Thiam, R. V. Farese, Jr., T. C. Walther, The biophysics and cell biology of lipid droplets. *Nat Rev Mol Cell Biol* **14**, 775-786 (2013).
45. A. R. Thiam, L. Foret, The physics of lipid droplet nucleation, growth and budding. *Biochim Biophys Acta* **1861**, 715-722 (2016).
46. A. R. Thiam, M. Beller, The why, when and how of lipid droplet diversity. *J Cell Sci* **130**, 315-324 (2017).
47. T. C. Walther, J. Chung, R. V. Farese, Jr., Lipid Droplet Biogenesis. *Annu Rev Cell Dev Biol* **33**, 491-510 (2017).
48. C. L. Jackson, Lipid droplet biogenesis. *Curr Opin Cell Biol* **59**, 88-96 (2019).
49. M. F. Renne, Y. A. Klug, P. Carvalho, Lipid droplet biogenesis: A mystery "unmixing"? *Semin Cell Dev Biol* **108**, 14-23 (2020).
50. M. Henne, J. M. Goodman, H. Hariri, Spatial compartmentalization of lipid droplet biogenesis. *Biochim Biophys Acta Mol Cell Biol Lipids* **1865**, 158499 (2020).
51. A. R. Thiam, E. Ikonen, Lipid Droplet Nucleation. *Trends Cell Biol* **31**, 108-118 (2021).
52. E. Brini, C. Simmerling, K. Dill, Protein storytelling through physics. *Science* **370** (2020).

53. B. R. Brooks *et al.*, CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **30**, 1545-1614 (2009).
54. D. Van Der Spoel *et al.*, GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **26**, 1701-1718 (2005).
55. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859-1865 (2008).
56. J. Lee *et al.*, CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* **12**, 405-413 (2016).
57. Y. Yu *et al.*, CHARMM36 Lipid Force Field with Explicit Treatment of Long-Range Dispersion: Parametrization and Validation for Phosphatidylethanolamine, Phosphatidylglycerol, and Ether Lipids. *J Chem Theory Comput* **17**, 1581-1595 (2021).
58. Y. Yu *et al.*, Semi-automated Optimization of the CHARMM36 Lipid Force Field to Include Explicit Treatment of Long-Range Dispersion. *J Chem Theory Comput* **17**, 1562-1580 (2021).
59. A. D. Mackerell *et al.*, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. *The Journal of Physical Chemistry B* **102**, 3586-3616 (1998).
60. L. Martínez, R. Andrade, E. G. Birgin, J. M. Martínez, PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **30**, 2157-2164 (2009).

61. J. B. Klauda *et al.*, Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *The Journal of Physical Chemistry B* **114**, 7830-7843 (2010).
62. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, O. Beckstein, MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32**, 2319-2327 (2011).
63. R. B. Best *et al.*, Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone  $\phi$ ,  $\psi$  and Side-Chain  $\chi_1$  and  $\chi_2$  Dihedral Angles. *Journal of Chemical Theory and Computation* **8**, 3257-3273 (2012).
64. J. Lee, S. G. Taneva, B. W. Holland, D. P. Tieleman, R. B. Cornell, Structural basis for autoinhibition of CTP:phosphocholine cytidyltransferase (CCT), the regulatory enzyme in phosphatidylcholine synthesis, by its membrane-binding amphipathic helix. *J Biol Chem* **289**, 1742-1755 (2014).
65. F. Wilfling *et al.*, Triacylglycerol synthesis enzymes mediate lipid droplet growth by relocalizing from the ER to lipid droplets. *Dev Cell* **24**, 384-399 (2013).
66. M. J. Olarte *et al.*, Determinants of Endoplasmic Reticulum-to-Lipid Droplet Protein Targeting. *Dev Cell* **54**, 471-487 e477 (2020).
67. S. Kim, M. I. Oh, J. M. J. Swanson, Stressed Lipid Droplets: How Neutral Lipids Relieve Surface Tension and Membrane Expansion Drives Protein Association. *J Phys Chem B* **125**, 5572-5586 (2021).
68. S. Kim, J. M. J. Swanson, The Surface and Hydration Properties of Lipid Droplets. *Biophys J* **119**, 1958-1969 (2020).

69. S. Kim, G. A. Voth, Physical Characterization of Triolein and Implications for Its Role in Lipid Droplet Biogenesis. *J Phys Chem B* **125**, 6874-6888 (2021).
70. K. Vanommeslaeghe *et al.*, CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* 10.1002/jcc.21367, NA-NA (2009).
71. J. C. Phillips *et al.*, Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802 (2005).
72. J. V. Vermaas, D. J. Hardy, J. E. Stone, E. Tajkhorshid, A. Kohlmeyer, TopoGromacs: Automated Topology Conversion from CHARMM to GROMACS within VMD. *Journal of Chemical Information and Modeling* **56**, 1112-1116 (2016).
73. N. A. Ducharme, P. E. Bickel, Lipid droplets in lipogenesis and lipolysis. *Endocrinology* **149**, 942-949 (2008).
74. N. Krahnmer, R. V. Farese, Jr., T. C. Walther, Balancing the fat: lipid droplets and human disease. *EMBO Mol Med* **5**, 973-983 (2013).
75. N. Kory, A. R. Thiam, R. V. Farese, Jr., T. C. Walther, Protein Crowding Is a Determinant of Lipid Droplet Protein Composition. *Dev Cell* **34**, 351-363 (2015).
76. N. Jacquier *et al.*, Lipid droplets are functionally connected to the endoplasmic reticulum in *Saccharomyces cerevisiae*. *J Cell Sci* **124**, 2424-2437 (2011).
77. F. Wilfling *et al.*, Arf1/COPI machinery acts directly on lipid droplets and enables their connection to the ER for protein targeting. *Elife* **3**, e01607 (2014).
78. M. J. Olarte *et al.*, Determinants of Endoplasmic Reticulum-to-Lipid Droplet Protein Targeting. *Dev Cell* **54**, 471-487 (2020).

79. R. L. Meegalla, J. T. Billheimer, D. Cheng, Concerted elevation of acyl-coenzyme A:diacylglycerol acyltransferase (DGAT) activity through independent stimulation of mRNA expression of DGAT1 and DGAT2 by carbohydrate and insulin. *Biochemical and Biophysical Research Communications* **298**, 317-323 (2002).
80. J. M. Shockey *et al.*, Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. *Plant Cell* **18**, 2294-2313 (2006).
81. C. A. Harris *et al.*, DGAT enzymes are required for triacylglycerol synthesis and lipid droplets in adipocytes. *J Lipid Res* **52**, 657-667 (2011).
82. C. Prevost *et al.*, Mechanism and Determinants of Amphipathic Helix-Containing Protein Targeting to Lipid Droplets. *Dev Cell* **44**, 73-86 (2018).
83. A. R. Kimmel, C. Sztalryd, The Perilipins: Major Cytosolic Lipid Droplet-Associated Proteins and Their Roles in Cellular Lipid Storage, Mobilization, and Systemic Homeostasis. *Annu Rev Nutr* **36**, 471-509 (2016).
84. A. Copic *et al.*, A giant amphipathic helix from a perilipin that is adapted for coating lipid droplets. *Nat Commun* **9**, 1332 (2018).
85. D. Ajjaji *et al.*, Dual binding motifs underpin the hierarchical association of perilipins 1-3 with lipid droplets. *Mol Biol Cell* **30**, 703-716 (2019).
86. N. Kraemer *et al.*, Phosphatidylcholine synthesis for lipid droplet expansion is mediated by localized activation of CTP:phosphocholine cytidyltransferase. *Cell Metab* **14**, 504-515 (2011).

87. K. Hsieh *et al.*, Perilipin family members preferentially sequester to either triacylglycerol-specific or cholesteryl-ester-specific intracellular lipid storage droplets. *J Cell Sci* **125**, 4067-4076 (2012).
88. J. A. Hamilton, K. W. Miller, D. M. Small, Solubilization of triolein and cholesteryl oleate in egg phosphatidylcholine vesicles. *J Biol Chem* **258**, 12821-12826 (1983).
89. H. Khandelia, L. Duelund, K. I. Pakkanen, J. H. Ipsen, Triglyceride Blisters in Lipid Bilayers: Implications for Lipid Droplet Biogenesis and the Mobile Lipid Signal in Cancer Cell Membranes. *PLoS ONE* **5**, e12811 (2010).
90. K. Ben M'barek *et al.*, ER Membrane Phospholipids and Surface Tension Control Cellular Lipid Droplet Formation. *Dev Cell* **41**, 591-604 (2017).
91. P. Campomanes, V. Zoni, S. Vanni, Local accumulation of diacylglycerol alters membrane properties nonlinearly due to its transbilayer activity. *Communications Chemistry* **2**, 72 (2019).
92. A. Koivuniemi, M. Heikela, P. T. Kovanen, I. Vattulainen, M. T. Hyvonen, Atomistic simulations of phosphatidylcholines and cholesteryl esters in high-density lipoprotein-sized lipid droplet and trilayer: clues to cholesteryl ester transport and storage. *Biophys J* **96**, 4099-4108 (2009).
93. L. Yetukuri *et al.*, Composition and lipid spatial distribution of HDL particles in subjects with low and high HDL-cholesterol. *J Lipid Res* **51**, 2341-2351 (2010).
94. T. Vuorela *et al.*, Role of lipids in spheroidal high density lipoproteins. *PLoS Comput Biol* **6**, e1000964 (2010).
95. T. Murtola *et al.*, Low density lipoprotein: structure, dynamics, and interactions of apoB-100 with lipids. *Soft Matter* **7**, 8135-8141 (2011).

96. A. Koivuniemi, I. Vattulainen, Revealing structural and dynamical properties of high density lipoproteins through molecular simulations. *Soft Matter* **8**, 1262-1267 (2012).
97. O. H. Ollila, A. Lamberg, M. Lehtivaara, A. Koivuniemi, I. Vattulainen, Interfacial tension and surface pressure of high density lipoprotein, low density lipoprotein, and related lipid droplets. *Biophys J* **103**, 1236-1244 (2012).
98. A. Koivuniemi, M. Sysi-Aho, M. Oresic, S. Ollila, Interfacial properties of high-density lipoprotein-like lipid droplets with different lipid and apolipoprotein A-I compositions. *Biophys J* **104**, 2193-2201 (2013).
99. S. M. Gordon *et al.*, Identification of a novel lipid binding motif in apolipoprotein B by the analysis of hydrophobic cluster domains. *Biochim Biophys Acta Biomembr* **1859**, 135-145 (2017).
100. G. Henneré, P. Prognon, F. Brion, V. Rosilio, I. Nicolis, Molecular dynamics simulation of a mixed lipid emulsion model: Influence of the triglycerides on interfacial phospholipid organization. *Journal of Molecular Structure: THEOCHEM* **901**, 174-185 (2009).
101. V. V. Chaban, H. Khandelia, Distribution of neutral lipids in the lipid droplet core. *J Phys Chem B* **118**, 11145-11151 (2014).
102. V. V. Chaban, H. Khandelia, Lipid structure in triolein lipid droplets. *J Phys Chem B* **118**, 10335-10340 (2014).
103. J. W. Wagner, J. F. Dama, A. E. Durumeric, G. A. Voth, On the representability problem and the physical meaning of coarse-grained models. *J Chem Phys* **145**, 044108 (2016).
104. F. Aydin, R. Sun, J. M. J. Swanson, Mycolactone Toxin Membrane Permeation: Atomistic versus Coarse-Grained MARTINI Simulations. *Biophys J* **117**, 87-98 (2019).

105. R. Desikan, S. M. Patra, K. Sarthak, P. K. Maiti, K. G. Ayappa, Comparison of coarse-grained (MARTINI) and atomistic molecular dynamics simulations of  $\alpha$  and  $\beta$  toxin nanopores in lipid membranes. *Journal of Chemical Sciences* **129**, 1017-1030 (2017).
106. Z. Jarin, J. Newhouse, G. A. Voth, Coarse-grained Force Fields from the Perspective of Statistical Mechanics: Better Understanding the Origins of a MARTINI Hangover. *bioRxiv* **10.1101/2020.06.25.171363** (2020).
107. S. Jo, T. Kim, W. Im, Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS One* **2**, e880 (2007).
108. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859-1865 (2008).
109. E. L. Wu *et al.*, CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem* **35**, 1997-2004 (2014).
110. K. Gawrisch *et al.*, Energetics of a hexagonal-lamellar-hexagonal-phase transition sequence in dioleoylphosphatidylethanolamine membranes. *Biochemistry* **31**, 2856-2864 (1992).
111. W. Ding, M. Palaiokostas, W. Wang, M. Orsi, Effects of Lipid Composition on Bilayer Membranes Quantified by All-Atom Molecular Dynamics. *J Phys Chem B* **119**, 15263-15274 (2015).
112. J. B. Klauda *et al.*, Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B* **114**, 7830-7843 (2010).
113. A. J. Sodt, R. W. Pastor, Bending free energy from simulation: correspondence of planar and inverse hexagonal lipid phases. *Biophys J* **104**, 2202-2211 (2013).



114. L. Martinez, R. Andrade, E. G. Birgin, J. M. Martinez, PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J Comput Chem* **30**, 2157-2164 (2009).
115. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935 (1983).
116. T. H. Gouw, J. C. Vlugter, Physical Properties of Triglycerides. I. Density and Refractive Index. *Fette, Seifen, Anstrichmittel* **68**, 544-549 (1966).
117. M. J. Abraham *et al.*, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).
118. A. D. MacKerell *et al.*, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586-3616 (1998).
119. U. Essmann *et al.*, A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577-8593 (1995).
120. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**, 7182-7190 (1981).
121. B. Hess, P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput* **4**, 116-122 (2008).
122. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101 (2007).
123. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38 (1996).

124. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, O. Beckstein, MDAAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**, 2319-2327 (2011).
125. L. Vamparys *et al.*, Conical lipids in flat bilayers induce packing defects similar to that induced by positive curvature. *Biophys J* **104**, 585-593 (2013).
126. S. Vanni *et al.*, Amphipathic lipid packing sensor motifs: probing bilayer defects with hydrophobic residues. *Biophys J* **104**, 575-584 (2013).
127. R. Gautier *et al.*, PackMem: A Versatile Tool to Compute and Visualize Interfacial Packing Defects in Lipid Bilayers. *Biophys J* **115**, 436-444 (2018).
128. H. Cui, E. Lyman, G. A. Voth, Mechanism of membrane curvature sensing by amphipathic helix containing proteins. *Biophys J* **100**, 1271-1279 (2011).
129. H. Schindler, J. Seelig, Deuterium order parameters in relation to thermodynamic properties of a phospholipid bilayer. A statistical mechanical interpretation. *Biochemistry* **14**, 2283-2287 (1975).
130. C. Das, M. G. Noro, P. D. Olmsted, Simulation studies of stratum corneum lipid mixtures. *Biophys J* **97**, 1941-1951 (2009).
131. Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **314**, 141-151 (1999).
132. G. M. Torrie, J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **23**, 187-199 (1977).
133. A. Grossfield, WHAM: an implementation of the weighted histogram analysis method. (version 2.0.9).

134. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **13**, 1011-1021 (1992).
135. G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **185**, 604-613 (2014).
136. S.-J. Marrink, H. J. C. Berendsen, Simulation of water transport through a lipid membrane. *The Journal of Physical Chemistry* **98**, 4155-4168 (1994).
137. A. Kramer *et al.*, Interactions of Water and Alkanes: Modifying Additive Force Fields to Account for Polarization Effects. *J Chem Theory Comput* **15**, 3854-3867 (2019).
138. L. Ragni, A. Berardinelli, C. Cevoli, E. Valli, Assessment of the water content in extra virgin olive oils by Time Domain Reflectometry (TDR) and Partial Least Squares (PLS) regression methods. *Journal of Food Engineering* **111**, 66-72 (2012).
139. L. Cerretani *et al.*, Rapid FTIR determination of water, phenolics and antioxidant activity of olive oil. *European Journal of Lipid Science and Technology* **112**, 1150-1157 (2010).
140. E. Hatzakis, P. Dais, Determination of water content in olive oil by <sup>31</sup>P NMR spectroscopy. *J Agric Food Chem* **56**, 1866-1872 (2008).
141. R. Zimmermann *et al.*, Fat mobilization in adipose tissue is promoted by adipose triglyceride lipase. *Science* **306**, 1383-1386 (2004).
142. E. Smirnova *et al.*, ATGL has a key role in lipid droplet/adiposome degradation in mammalian cells. *EMBO Rep* **7**, 106-113 (2006).
143. M. P. Gaidhu, N. M. Anthony, P. Patel, T. J. Hawke, R. B. Ceddia, Dysregulation of lipolysis and lipid metabolism in visceral and subcutaneous adipocytes by high-fat diet: role of ATGL, HSL, and AMPK. *Am J Physiol Cell Physiol* **298**, C961-971 (2010).

144. A. Penno, G. Hackenbroich, C. Thiele, Phospholipids and lipid droplets. *Biochim Biophys Acta* **1831**, 589-594 (2013).
145. T. C. Walther, J. Chung, R. V. Farese, Jr., Lipid Droplet Biogenesis. *Annu Rev Cell Dev Biol* **33**, 491-510 (2017).
146. K. Tauchi-Sato, S. Ozeki, T. Houjou, R. Taguchi, T. Fujimoto, The surface of lipid droplets is a phospholipid monolayer with a unique Fatty Acid composition. *J Biol Chem* **277**, 44507-44512 (2002).
147. Y. Ding *et al.*, Identification of the major functional proteins of prokaryotic lipid droplets. *J Lipid Res* **53**, 399-411 (2012).
148. E. Currie *et al.*, High confidence proteomic analysis of yeast LDs identifies additional droplet proteins and reveals connections to dolichol synthesis and sterol acetylation. *J Lipid Res* **55**, 1465-1477 (2014).
149. D. L. Brasaemle, G. Dolios, L. Shapiro, R. Wang, Proteomic analysis of proteins associated with lipid droplets of basal and lipolytically stimulated 3T3-L1 adipocytes. *J Biol Chem* **279**, 46835-46842 (2004).
150. J. Bouchoux *et al.*, The proteome of cytosolic lipid droplets isolated from differentiated Caco-2/TC7 enterocytes reveals cell-specific characteristics. *Biol Cell* **103**, 499-517 (2011).
151. M. Poppelreuther *et al.*, The metabolic capacity of lipid droplet localized acyl-CoA synthetase 3 is not sufficient to support local triglyceride synthesis independent of the endoplasmic reticulum in A431 cells. *Biochim Biophys Acta Mol Cell Biol Lipids* **1863**, 614-624 (2018).

152. R. Zechner, FAT FLUX: enzymes, regulators, and pathophysiology of intracellular lipolysis. *EMBO Mol Med* **7**, 359-362 (2015).
153. J. A. Olzmann, P. Carvalho, Dynamics and functions of lipid droplets. *Nat Rev Mol Cell Biol* **20**, 137-155 (2019).
154. C. Prevost *et al.*, Mechanism and Determinants of Amphipathic Helix-Containing Protein Targeting to Lipid Droplets. *Dev Cell* **44**, 73-86 e74 (2018).
155. Y. Q. Chen *et al.*, AGPAT6 is a novel microsomal glycerol-3-phosphate acyltransferase. *J Biol Chem* **283**, 10048-10057 (2008).
156. C. A. Nagle *et al.*, Identification of a novel sn-glycerol-3-phosphate acyltransferase isoform, GPAT4, as the enzyme deficient in *Agpat6*<sup>-/-</sup> mice. *J Lipid Res* **49**, 823-831 (2008).
157. R. Coleman, R. M. Bell, Evidence that biosynthesis of phosphatidylethanolamine, phosphatidylcholine, and triacylglycerol occurs on the cytoplasmic side of microsomal vesicles. *J Cell Biol* **76**, 245-253 (1978).
158. H. Wang *et al.*, Seipin is required for converting nascent to mature lipid droplets. *Elife* **5** (2016).
159. A. Ruggiano, G. Mora, L. Buxo, P. Carvalho, Spatial control of lipid droplet proteins by the ERAD ubiquitin ligase Doa10. *EMBO J* **35**, 1644-1655 (2016).
160. Y. Ye, W. K. Tang, T. Zhang, D. Xia, A Mighty "Protein Extractor" of the Cell: Structure and Function of the p97/CDC48 ATPase. *Front Mol Biosci* **4**, 39 (2017).
161. H. J. Sharpe, T. J. Stevens, S. Munro, A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell* **142**, 158-169 (2010).

162. L. E. Hedin (2010) Intra- and intermolecular interactions in proteins: Studies of marginally hydrophobic transmembrane alpha-helices and protein-protein interactions. in *Department of Biochemistry and Biophysics* (Stockholm University, Sweden).
163. S. Raman *et al.*, Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins* **77 Suppl 9**, 89-99 (2009).
164. C. A. Rohl, C. E. Strauss, K. M. Misura, D. Baker, Protein structure prediction using Rosetta. *Methods Enzymol* **383**, 66-93 (2004).
165. D. T. Jones, Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* **23**, 538-544 (2007).
166. D. T. Jones, W. R. Taylor, J. M. Thornton, A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry* **33**, 3038-3049 (1994).
167. T. Nugent, D. T. Jones, Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics* **10**, 159 (2009).
168. K. D. Tsirigos, C. Peters, N. Shu, L. Kall, A. Elofsson, The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res* **43**, W401-407 (2015).
169. R. F. Alford *et al.*, An Integrated Framework Advancing Membrane Protein Modeling and Design. *PLoS Comput Biol* **11**, e1004398 (2015).
170. R. Bartz *et al.*, Lipidomics reveals that adiposomes store ether lipids and mediate phospholipid traffic. *J Lipid Res* **48**, 837-847 (2007).
171. C. Chitraju *et al.*, Lipidomic analysis of lipid droplets from murine hepatocytes reveals distinct signatures for nutritional stress. *J Lipid Res* **53**, 2141-2152 (2012).

172. J. F. Dama, G. Rotskoff, M. Parrinello, G. A. Voth, Transition-Tempered Metadynamics: Robust, Convergent Metadynamics via On-the-Fly Transition Barrier Estimation. *J Chem Theory Comput* **10**, 3626-3633 (2014).
173. K. E. Norman, H. Nymeyer, Indole localization in lipid membranes revealed by molecular simulation. *Biophys J* **91**, 2046-2054 (2006).
174. W. M. Yau, W. C. Wimley, K. Gawrisch, S. H. White, The preference of tryptophan for membrane interfaces. *Biochemistry* **37**, 14713-14718 (1998).
175. C. Landolt-Marticorena, K. A. Williams, C. M. Deber, R. A. Reithmeier, Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins. *J Mol Biol* **229**, 602-608 (1993).
176. J. P. Segrest, H. De Loof, J. G. Dohlman, C. G. Brouillette, G. M. Anantharamaiah, Amphipathic helix motif: classes and properties. *Proteins* **8**, 103-117 (1990).
177. N. Kraemer *et al.*, Protein correlation profiles identify lipid droplet proteins with high confidence. *Mol Cell Proteomics* **12**, 1115-1126 (2013).
178. X. D. Gao, H. Tachikawa, T. Sato, Y. Jigami, N. Dean, Alg14 recruits Alg13 to the cytoplasmic face of the endoplasmic reticulum to form a novel bipartite UDP-N-acetylglucosamine transferase required for the second step of N-linked glycosylation. *J Biol Chem* **280**, 36254-36262 (2005).
179. H. A. Saka *et al.*, Chlamydia trachomatis Infection Leads to Defined Alterations to the Lipid Droplet Proteome in Epithelial Cells. *PLoS One* **10**, e0124630 (2015).
180. D. Gorlich, S. Prehn, E. Hartmann, K. U. Kalies, T. A. Rapoport, A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation. *Cell* **71**, 489-503 (1992).

181. B. Schrul, R. R. Kopito, Peroxin-dependent targeting of a lipid-droplet-destined membrane protein to ER subdomains. *Nat Cell Biol* **18**, 740-751 (2016).
182. P. Bradley, K. M. Misura, D. Baker, Toward high-resolution de novo structure prediction for small proteins. *Science* **309**, 1868-1871 (2005).
183. D. E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **32**, W526-531 (2004).
184. Y. Song *et al.*, High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735-1742 (2013).
185. V. Yarov-Yarovoy, J. Schonbrun, D. Baker, Multipass membrane protein structure prediction using Rosetta. *Proteins* **62**, 1010-1025 (2006).
186. P. Barth, J. Schonbrun, D. Baker, Toward high-resolution prediction and design of transmembrane helical protein structures. *Proc Natl Acad Sci U S A* **104**, 15682-15687 (2007).
187. P. Barth, B. Wallner, D. Baker, Prediction of membrane protein structures with complex topologies using limited constraints. *Proc Natl Acad Sci U S A* **106**, 1409-1414 (2009).
188. J. Koehler Leman, B. K. Mueller, J. J. Gray, Expanding the toolkit for membrane protein modeling in Rosetta. *Bioinformatics* **33**, 754-756 (2017).
189. J. Koehler Leman, M. B. Ulmschneider, J. J. Gray, Computational modeling of membrane proteins. *Proteins* **83**, 1-24 (2015).
190. H. Viklund, A. Elofsson, OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics* **24**, 1662-1668 (2008).



191. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33-38 (1996).
192. R. B. Best *et al.*, Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J Chem Theory Comput* **8**, 3257-3273 (2012).
193. M. Javanainen, H. Martinez-Seara, Efficient preparation and analysis of membrane and membrane protein systems. *Biochim Biophys Acta* **1858**, 2468-2482 (2016).
194. A. Laio, M. Parrinello, Escaping free-energy minima. *Proc Natl Acad Sci U S A* **99**, 12562-12566 (2002).
195. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, Multidimensional free-energy calculations using the weighted histogram analysis method. *Journal of Computational Chemistry* **16**, 1339-1350 (1995).
196. B. Roux, The calculation of the potential of mean force using computer simulations. *Computer Physics Communications* **91**, 275-282 (1995).
197. A. Grossfield (WHAM: an implementation of the weighted histogram analysis method.
198. R. Sun, J. F. Dama, J. S. Tan, J. P. Rose, G. A. Voth, Transition-Tempered Metadynamics Is a Promising Tool for Studying the Permeation of Drug-like Molecules through Membranes. *J Chem Theory Comput* **12**, 5157-5169 (2016).
199. F. Wilfling, J. T. Haas, T. C. Walther, R. V. Farese, Jr., Lipid droplet biogenesis. *Curr Opin Cell Biol* **29**, 39-45 (2014).
200. R. V. Farese, Jr., T. C. Walther, Lipid droplets finally get a little R-E-S-P-E-C-T. *Cell* **139**, 855-860 (2009).

201. L. Caillon *et al.*, Triacylglycerols sequester monotopic membrane proteins to lipid droplets. *Nat Commun* **11**, 3944 (2020).
202. H. Wang *et al.*, Seipin is required for converting nascent to mature lipid droplets. *Elife* **5**, e16582 (2016).
203. E. R. Rowe *et al.*, Conserved Amphipathic Helices Mediate Lipid Droplet Targeting of Perilipins 1-3. *J Biol Chem* **291**, 6664-6678 (2016).
204. L. Yue *et al.*, Differential dephosphorylation of CTP:phosphocholine cytidylyltransferase upon translocation to nuclear membranes and lipid droplets. *Mol Biol Cell* **31**, 1047-1059 (2020).
205. I. C. Northwood, A. H. Tong, B. Crawford, A. E. Drobnies, R. B. Cornell, Shuttling of CTP:Phosphocholine cytidylyltransferase between the nucleus and endoplasmic reticulum accompanies the wave of phosphatidylcholine synthesis during the G(0) --> G(1) transition. *J Biol Chem* **274**, 26240-26248 (1999).
206. M. Gimenez-Andres, A. Copic, B. Antonny, The Many Faces of Amphipathic Helices. *Biomolecules* **8** (2018).
207. H. Cui, E. Lyman, G. A. Voth, Mechanism of Membrane Curvature Sensing by Amphipathic Helix Containing Proteins. *Biophysical Journal* **100**, 1271-1279 (2011).
208. V. Monje-Galvan, J. B. Klauda, Preferred Binding Mechanism of Osh4's Amphipathic Lipid-Packing Sensor Motif, Insights from Molecular Dynamics. *J Phys Chem B* **122**, 9713-9723 (2018).
209. K. Hsieh *et al.*, Perilipin family members preferentially sequester to either triacylglycerol-specific or cholesteryl-ester-specific intracellular lipid storage droplets. *Journal of Cell Science* **125**, 4067-4076 (2012).

210. B. Antonny, I. Huber, S. Paris, M. Chabre, D. Cassel, Activation of ADP-ribosylation factor 1 GTPase-activating protein by phosphatidylcholine-derived diacylglycerols. *J Biol Chem* **272**, 30848-30851 (1997).
211. J. Bigay, P. Gounon, S. Robineau, B. Antonny, Lipid packing sensed by ArfGAP1 couples COPI coat disassembly to membrane bilayer curvature. *Nature* **426**, 563-566 (2003).
212. K. Bersuker *et al.*, A Proximity Labeling Strategy Provides Insights into the Composition and Dynamics of Lipid Droplet Proteomes. *Developmental Cell* **44**, 97-112.e117 (2018).
213. R. B. Cornell, N. D. Ridgway, CTP:phosphocholine cytidyltransferase: Function, regulation, and structure of an amphitropic enzyme required for membrane biogenesis. *Prog Lipid Res* **59**, 147-171 (2015).
214. R. B. Cornell, Membrane Lipids Assist Catalysis by CTP: Phosphocholine Cytidylyltransferase. *Journal of Molecular Biology* **432**, 5023-5042 (2020).
215. R. B. Cornell, Regulation of CTP:phosphocholine cytidyltransferase by lipids. 2. Surface curvature, acyl chain length, and lipid-phase dependence for activation. *Biochemistry* **30**, 5881-5888 (1991).
216. R. B. Cornell, Regulation of CTP:phosphocholine cytidyltransferase by lipids. 1. Negative surface charge dependence for activation. *Biochemistry* **30**, 5873-5880 (1991).
217. M. Ramezanpour, J. Lee, S. G. Taneva, D. P. Tieleman, R. B. Cornell, An auto-inhibitory helix in CTP:phosphocholine cytidyltransferase hijacks the catalytic residue and constrains a pliable, domain-bridging helix pair. *J Biol Chem* **293**, 7070-7084 (2018).
218. R. B. Cornell, I. C. Northwood, Regulation of CTP:phosphocholine cytidyltransferase by amphitropism and relocalization. *Trends in Biochemical Sciences* **25**, 441-447 (2000).

219. S. M. Davies, R. M. Epand, R. Kraayenhof, R. B. Cornell, Regulation of CTP: phosphocholine cytidyltransferase activity by the physical properties of lipid membranes: an important role for stored curvature strain energy. *Biochemistry* **40**, 10522-10531 (2001).
220. S. Jo, J. B. Lim, J. B. Klauda, W. Im, CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophysical Journal* **97**, 50-58 (2009).
221. E. L. Wu *et al.*, CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *Journal of Computational Chemistry* **35**, 1997-2004 (2014).
222. W. L. Jorgensen, Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *Journal of the American Chemical Society* **103**, 335-340 (1981).
223. B. Hess, P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *Journal of Chemical Theory and Computation* **4**, 116-122 (2008).
224. G. Bussi, G. A. Tribello, "Analyzing and Biasing Simulations with PLUMED" in *Methods in Molecular Biology*. (Springer New York, 2019), 10.1007/978-1-4939-9608-7\_21, pp. 529-578.
225. S. Nosé, A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **81**, 511-519 (1984).
226. W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **31**, 1695-1697 (1985).
227. H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, J. R. Haak, Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684-3690 (1984).

228. J. C. Phillips *et al.*, Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys* **153**, 044130 (2020).
229. S. Fan *et al.* (2019) PMDA - Parallel Molecular Dynamics Analysis. in *Proceedings of the 18th Python in Science Conference (SciPy)*.
230. A. Grossfield, version 2.0.9. WHAM: an implementation of the weighted histogram analysis method.
231. S. J. Dunne, R. B. Cornell, J. E. Johnson, N. R. Glover, A. S. Tracey, Structure of the membrane binding domain of CTP:phosphocholine cytidylyltransferase. *Biochemistry* **35**, 11975-11984 (1996).
232. S. Taneva, J. E. Johnson, R. B. Cornell, Lipid-induced conformational switch in the membrane binding domain of CTP:phosphocholine cytidylyltransferase: a circular dichroism study. *Biochemistry* **42**, 11768-11776 (2003).
233. W.-M. Yau, W. C. Wimley, K. Gawrisch, S. H. White, The Preference of Tryptophan for Membrane Interfaces. *Biochemistry* **37**, 14713-14718 (1998).
234. A. Gruber *et al.*, The N-terminal region of comparative gene identification-58 (CGI-58) is important for lipid droplet binding and activation of adipose triglyceride lipase. *J Biol Chem* **285**, 12289-12298 (2010).
235. A. Boeszoermyeni *et al.*, Structure of a CGI-58 motif provides the molecular basis of lipid droplet anchoring. *J Biol Chem* **290**, 26361-26372 (2015).
236. A. J. Sodt, M. Logan Sandar, Klaus Gawrisch, R. W. Pastor, Edward Lyman, The Molecular Structure of the Liquid-Ordered Phase of Lipid Bilayers. *Journal of the American Chemical Society* **136**, 725-732 (2014).

237. R. M. Venable, F. L. H. Brown, R. W. Pastor, Mechanical properties of lipid bilayers from molecular dynamics simulation. *Chem Phys Lipids* **192**, 60-74 (2015).
238. G. Enkavi, M. Javanainen, W. Kulig, T. Rog, I. Vattulainen, Multiscale Simulations of Biological Membranes: The Challenge To Understand Biological Phenomena in a Living Substance. *Chem Rev* **119**, 5607-5774 (2019).
239. V. Corradi *et al.*, Emerging Diversity in Lipid-Protein Interactions. *Chem Rev* **119**, 5775-5848 (2019).
240. S. J. Marrink *et al.*, Computational Modeling of Realistic Cell Membranes. *Chemical Reviews* **119**, 6184-6226 (2019).
241. P. Campomanes, V. Zoni, S. Vanni, Local accumulation of diacylglycerol alters membrane properties nonlinearly due to its transbilayer activity. *Communications Chemistry* **2** (2019).
242. V. Zoni *et al.*, Pre-existing bilayer stresses modulate triglyceride accumulation in the ER versus lipid droplets. *Elife* **10** (2021).
243. V. V. Chaban, H. Khandelia, Distribution of Neutral Lipids in the Lipid Droplet Core. *The Journal of Physical Chemistry B* **118**, 11145-11151 (2014).
244. V. V. Chaban, H. Khandelia, Lipid Structure in Triolein Lipid Droplets. *The Journal of Physical Chemistry B* **118**, 10335-10340 (2014).
245. Z. Chen, R. P. Rand, The influence of cholesterol on phospholipid membrane curvature and bending elasticity. *Biophys J* **73**, 267-276 (1997).
246. M. M. Kozlov, S. Leikin, R. P. Rand, Bending, hydration and interstitial energies quantitatively account for the hexagonal-lamellar-hexagonal reentrant phase transition in dioleoylphosphatidylethanolamine. *Biophys J* **67**, 1603-1611 (1994).

247. R. P. Rand, N. L. Fuller, Structural dimensions and their changes in a reentrant hexagonal-lamellar transition of phospholipids. *Biophysical Journal* **66**, 2127-2138 (1994).
248. Alexander J. Sodt, Richard W. Pastor, Bending Free Energy from Simulation: Correspondence of Planar and Inverse Hexagonal Lipid Phases. *Biophysical Journal* **104**, 2202-2211 (2013).
249. A. N. Leonard *et al.*, Comparison of Additive and Polarizable Models with Explicit Treatment of Long-Range Lennard-Jones Interactions Using Alkane Simulations. *J Chem Theory Comput* **14**, 948-958 (2018).
250. S. E. Feller, R. W. Pastor, Constant surface tension simulations of lipid bilayers: The sensitivity of surface areas and compressibilities. *The Journal of Chemical Physics* **111**, 1281-1287 (1999).
251. J. Seelig, W. Niederberger, Deuterium-labeled lipids as structural probes in liquid crystalline bilayers. Deuterium magnetic resonance study. *Journal of the American Chemical Society* **96**, 2069-2072 (1974).
252. S. Marčelja, Chain ordering in liquid crystals. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **367**, 165-176 (1974).
253. H. Schindler, J. Seelig, Deuterium order parameters in relation to thermodynamic properties of a phospholipid bilayer. Statistical mechanical interpretation. *Biochemistry* **14**, 2283-2287 (1975).
254. J. F. Dama, G. Rotskoff, M. Parrinello, G. A. Voth, Transition-Tempered Metadynamics: Robust, Convergent Metadynamics via On-the-Fly Transition Barrier Estimation. *Journal of Chemical Theory and Computation* **10**, 3626-3633 (2014).

255. R. Sun, J. F. Dama, J. S. Tan, J. P. Rose, G. A. Voth, Transition-Tempered Metadynamics Is a Promising Tool for Studying the Permeation of Drug-like Molecules through Membranes. *Journal of Chemical Theory and Computation* **12**, 5157-5169 (2016).
256. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126**, 014101 (2007).
257. J. J. Madsen, J. M. A. Grime, J. S. Rossman, G. A. Voth, Entropic forces drive clustering and spatial localization of influenza A M2 during viral budding. *Proc Natl Acad Sci U S A* **115**, E8595-E8603 (2018).
258. W. Shinoda, R. Devane, M. L. Klein, Computer simulation studies of self-assembling macromolecules. *Current Opinion in Structural Biology* **22**, 175-186 (2012).
259. W. Shinoda, R. Devane, M. L. Klein, Coarse-grained molecular modeling of non-ionic surfactant self-assembly. *Soft Matter* **4**, 2454 (2008).
260. W. Shinoda, R. DeVane, M. L. Klein, Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants. *Molecular Simulation* **33**, 27-36 (2007).
261. S. Plimpton, Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics* **117**, 1-19 (1995).
262. G. J. Martyna, D. J. Tobias, M. L. Klein, Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **101**, 4177-4189 (1994).
263. W. Shinoda, M. Shiga, M. Mikami, Rapid estimation of elastic constants by molecular dynamics simulation under constant stress. *Physical Review B* **69** (2004).
264. R. S. Arnold, R. B. Cornell, Lipid regulation of CTP: phosphocholine cytidyltransferase: electrostatic, hydrophobic, and synergistic interactions of anionic phospholipids and diacylglycerol. *Biochemistry* **35**, 9917-9924 (1996).



265. R. B. Cornell, R. S. Arnold, Modulation of the activities of enzymes of membrane lipid metabolism by non-bilayer-forming lipids. *Chemistry and Physics of Lipids* **81**, 215-227 (1996).
266. G. Brannigan, F. L. Brown, Solvent-free simulations of fluid membrane bilayers. *J Chem Phys* **120**, 1059-1071 (2004).
267. G. Brannigan, P. F. Phillips, F. L. Brown, Flexible lipid bilayers in implicit solvent. *Phys Rev E Stat Nonlin Soft Matter Phys* **72**, 011915 (2005).
268. E. G. Brandt, A. R. Braun, J. N. Sachs, J. F. Nagle, O. Edholm, Interpretation of fluctuation spectra in lipid bilayer simulations. *Biophys J* **100**, 2104-2111 (2011).
269. A. Santinho, A. Chorlay, L. Foret, A. R. Thiam, Fat inclusions strongly alter membrane mechanics. *Biophys J* **120**, 607-617 (2021).
270. N. Kučerka *et al.*, Lipid Bilayer Structure Determined by the Simultaneous Analysis of Neutron and X-Ray Scattering Data. *Biophysical Journal* **95**, 2356-2367 (2008).
271. J. Pan, S. Tristram-Nagle, N. Kučerka, J. F. Nagle, Temperature Dependence of Structure, Bending Rigidity, and Bilayer Interactions of Dioleoylphosphatidylcholine Bilayers. *Biophysical Journal* **94**, 117-124 (2008).
272. E. Evans, W. Rawicz, B. A. Smith, Concluding remarks Back to the future: mechanics and thermodynamics of lipid biomembranes. *Faraday Discuss.* **161**, 591-611 (2013).
273. W. Rawicz, K. C. Olbrich, T. McIntosh, D. Needham, E. Evans, Effect of Chain Length and Unsaturation on Elasticity of Lipid Bilayers. *Biophysical Journal* **79**, 328-339 (2000).
274. A. Chorlay, A. R. Thiam, An Asymmetry in Monolayer Tension Regulates Lipid Droplet Budding Direction. *Biophys J* **114**, 631-640 (2018).

275. A. Chorlay *et al.*, Membrane Asymmetry Imposes Directionality on Lipid Droplet Emergence from the ER. *Dev Cell* **50**, 25-42 e27 (2019).
276. X. Sui *et al.*, Structure and catalytic mechanism of a human triacylglycerol-synthesis enzyme. *Nature* **581**, 323-328 (2020).
277. L. Wang *et al.*, Structure and mechanism of human diacylglycerol O-acyltransferase 1. *Nature* **581**, 329-332 (2020).
278. A. Santinho *et al.*, Membrane Curvature Catalyzes Lipid Droplet Assembly. *Curr Biol* **30**, 2481-2494 e2486 (2020).
279. M. Pinot *et al.*, Lipid cell biology. Polyunsaturated phospholipids facilitate membrane deformation and fission by endocytic proteins. *Science* **345**, 693-697 (2014).
280. X. Rong *et al.*, Lpcat3-dependent production of arachidonoyl phospholipids is a key determinant of triglyceride secretion. *Elife* **4** (2015).
281. T. Hashidate-Yoshida *et al.*, Fatty acid remodeling by LPCAT3 enriches arachidonate in phospholipid membranes and regulates triglyceride transport. *eLife* **4** (2015).
282. H. Barelli, B. Antonny, Lipid unsaturation and organelle dynamics. *Curr Opin Cell Biol* **41**, 25-32 (2016).
283. J. Augenreich *et al.*, The conical shape of DIM lipids promotes Mycobacterium tuberculosis infection of macrophages. *Proc Natl Acad Sci U S A* **116**, 25649-25658 (2019).
284. S. Izvekov, G. A. Voth, A Multiscale Coarse-Graining Method for Biomolecular Systems. *The Journal of Physical Chemistry B* **109**, 2469-2473 (2005).
285. S. Izvekov, G. A. Voth, Multiscale Coarse-Graining of Mixed Phospholipid/Cholesterol Bilayers. *Journal of Chemical Theory and Computation* **2**, 637-648 (2006).

286. W. G. Noid *et al.*, The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *The Journal of Chemical Physics* **128**, 244114 (2008).
287. W. G. Noid *et al.*, The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. *J Chem Phys* **128**, 244115 (2008).
288. S. Izvekov, G. A. Voth, Solvent-Free Lipid Bilayer Model Using Multiscale Coarse-Graining. *The Journal of Physical Chemistry B* **113**, 4443-4455 (2009).
289. M. S. Shell, The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J Chem Phys* **129**, 144108 (2008).
290. A. J. Pak, T. Dannenhoffer-Lafage, J. J. Madsen, G. A. Voth, Systematic Coarse-Grained Lipid Force Fields with Semiexplicit Solvation via Virtual Sites. *J Chem Theory Comput* **15**, 2087-2100 (2019).
291. M. Simunovic *et al.*, Protein-mediated transformation of lipid vesicles into tubular networks. *Biophys J* **105**, 711-719 (2013).
292. M. Simunovic, A. Srivastava, G. A. Voth, Linear aggregation of proteins on the membrane as a prelude to membrane remodeling. *Proc Natl Acad Sci U S A* **110**, 20396-20401 (2013).
293. A. Srivastava, G. A. Voth, A Hybrid Approach for Highly Coarse-grained Lipid Bilayer Models. *J Chem Theory Comput* **9**, 750-765 (2013).
294. M. Simunovic *et al.*, How curvature-generating proteins build scaffolds on membrane nanotubes. *Proc Natl Acad Sci U S A* **113**, 11226-11231 (2016).
295. Z. Jarin *et al.*, Unusual Organization of I-BAR Proteins on Tubular and Vesicular Membranes. *Biophys J* **117**, 553-562 (2019).

296. J. F. Dama *et al.*, The Theory of Ultra-Coarse-Graining. 1. General Principles. *Journal of Chemical Theory and Computation* **9**, 2466-2480 (2013).
297. A. Davtyan, J. F. Dama, A. V. Sinitskiy, G. A. Voth, The Theory of Ultra-Coarse-Graining. 2. Numerical Implementation. *Journal of Chemical Theory and Computation* **10**, 5265-5275 (2014).
298. J. F. Dama, J. Jin, G. A. Voth, The Theory of Ultra-Coarse-Graining. 3. Coarse-Grained Sites with Rapid Local Equilibrium of Internal States. *J Chem Theory Comput* **13**, 1010-1022 (2017).
299. J. Jin, G. A. Voth, Ultra-Coarse-Grained Models Allow for an Accurate and Transferable Treatment of Interfacial Systems. *J Chem Theory Comput* **14**, 2180-2197 (2018).
300. J. Jin, Y. Han, G. A. Voth, Ultra-Coarse-Grained Liquid State Models with Implicit Hydrogen Bonding. *J Chem Theory Comput* **14**, 6159-6174 (2018).
301. J. Jin, A. Yu, G. A. Voth, Temperature and Phase Transferable Bottom-up Coarse-Grained Models. *J Chem Theory Comput* **16**, 6823-6842 (2020).
302. F. Deslandes, A. R. Thiam, L. Foret, Lipid Droplets Can Spontaneously Bud Off from a Symmetric Bilayer. *Biophys J* **113**, 15-18 (2017).
303. M. Wang, X. Yi, Bulging and budding of lipid droplets from symmetric and asymmetric membranes: competition between membrane elastic energy and interfacial energy. *Soft Matter* **17**, 5319-5328 (2021).
304. V. Zoni *et al.*, To Bud or Not to Bud: A Perspective on Molecular Simulations of Lipid Droplet Budding. *Frontiers in Molecular Biosciences* **6** (2019).
305. J. M. A. Grime, J. J. Madsen, Efficient Simulation of Tunable Lipid Assemblies Across Scales and Resolutions. *arXiv* arXiv:1910.05362 (2019).

306. T. Schneider, E. Stoll, Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Physical Review B* **17**, 1302-1322 (1978).
307. A. Barducci, G. Bussi, M. Parrinello, Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters* **100** (2008).
308. G. A. Tribello, F. Giberti, G. C. Sosso, M. Salvalaglio, M. Parrinello, Analyzing and Driving Cluster Formation in Atomistic Simulations. *J Chem Theory Comput* **13**, 1317-1327 (2017).
309. M. Bonomi, A. Barducci, M. Parrinello, Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *Journal of Computational Chemistry* **30**, 1615-1621 (2009).
310. P. Tiwary, M. Parrinello, A Time-Independent Free Energy Estimator for Metadynamics. *The Journal of Physical Chemistry B* **119**, 736-742 (2015).
311. I. R. Cooke, K. Kremer, M. Deserno, Tunable generic model for fluid bilayer membranes. *Physical Review E* **72** (2005).
312. J. W. Wagner, J. F. Dama, A. E. P. Durumeric, G. A. Voth, On the representability problem and the physical meaning of coarse-grained models. *The Journal of Chemical Physics* **145**, 044108 (2016).
313. V. T. Salo *et al.*, Seipin Facilitates Triglyceride Flow to Lipid Droplet and Counteracts Droplet Ripening via Endoplasmic Reticulum Contact. *Dev Cell* **50**, 478-493 e479 (2019).
314. H. M. Hankins, R. D. Baldrige, P. Xu, T. R. Graham, Role of Flippases, Scramblases and Transfer Proteins in Phosphatidylserine Subcellular Distribution. *Traffic* **16**, 35-47 (2015).
315. C. A. Alexander, R. L. Hamilton, R. J. Havel, Subcellular localization of B apoprotein of plasma lipoproteins in rat liver. *Journal of Cell Biology* **69**, 241-263 (1976).

316. R. L. Hamilton, J. S. Wong, C. M. Cham, L. B. Nielsen, S. G. Young, Chylomicron-sized lipid particles are formed in the setting of apolipoprotein B deficiency. *Journal of Lipid Research* **39**, 1543-1557 (1998).
317. J. Gong *et al.*, Fsp27 promotes lipid droplet growth by lipid exchange and transfer at lipid droplet contact sites. *Journal of Cell Biology* **195**, 953-963 (2011).
318. C. W. Wang, Y. H. Miao, Y. S. Chang, Control of lipid droplet size in budding yeast requires the collaboration between Fld1 and Ldb16. *J Cell Sci* **127**, 1214-1228 (2014).
319. V. Teixeira *et al.*, Regulation of lipid droplets by metabolically controlled Ldo isoforms. *J Cell Biol* **217**, 127-138 (2018).
320. W. Fei *et al.*, Fld1p, a functional homologue of human seipin, regulates the size of lipid droplets in yeast. *J Cell Biol* **180**, 473-482 (2008).
321. K. M. Szymanski *et al.*, The lipodystrophy protein seipin is found at endoplasmic reticulum lipid droplet junctions and is important for droplet morphology. *Proc Natl Acad Sci U S A* **104**, 20890-20895 (2007).
322. V. T. Salo *et al.*, Seipin regulates ER-lipid droplet contacts and cargo delivery. *EMBO J* **35**, 2699-2716 (2016).
323. C. Lundin *et al.*, Membrane topology of the human seipin protein. *FEBS Lett* **580**, 2281-2284 (2006).
324. M. Becuwe *et al.*, FIT2 is an acyl-coenzyme A diphosphatase crucial for endoplasmic reticulum homeostasis. *J Cell Biol* **219** (2020).
325. V. T. Salo, M. Holtta-Vuori, E. Ikonen, Seipin-Mediated Contacts as Gatekeepers of Lipid Flux at the Endoplasmic Reticulum–Lipid Droplet Nexus *Contact* **3**, 251525642094582 (2020).

326. I. Schachter, C. Allolio, G. Khelashvili, D. Harries, Confinement in Nanodiscs Anisotropically Modifies Lipid Bilayer Elastic Properties. *J Phys Chem B* **124**, 7166-7175 (2020).
327. S. Kim, C. Li, R. V. Farese, T. C. Walther, G. A. Voth, Membrane rigidity as a key factor governing initial stages of lipid droplet formation. *submitted*.
328. P. Georgiades *et al.*, The flexibility and dynamics of the tubules in the endoplasmic reticulum. *Sci Rep* **7**, 16474 (2017).
329. E. Lyman, J. Pfandtner, G. A. Voth, Systematic Multiscale Parameterization of Heterogeneous Elastic Network Models of Proteins. *Biophysical Journal* **95**, 4183-4192 (2008).
330. A. Fiser, R. K. Do, A. Sali, Modeling of loops in protein structures. *Protein Sci* **9**, 1753-1773 (2000).
331. L. J. McGuffin, K. Bryson, D. T. Jones, The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404-405 (2000).
332. A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567-580 (2001).
333. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858 (2015).
334. P. de Buyl, tidynamics: A tiny package to compute the dynamics of stochastic and molecular simulations. *Journal of Open Source Software* **3**, 877 (2018).

335. P. Campomanes, J. Prabhu, V. Zoni, S. Vanni, Re-charging your fats: Charmm36 parameters for neutral lipids triacylglycerol and diacylglycerol. *bioRxiv* **10.1101/2021.09.29.462351** (2021).
336. J. Mahamid *et al.*, Liquid-crystalline phase transitions in lipid droplets are related to cellular states and specific organelle association. *Proc Natl Acad Sci U S A* **116**, 16866-16871 (2019).
337. S. Rogers *et al.*, Liquid-crystalline lipid phase transitions in lipid droplets selectively remodel the LD proteome. *bioRxiv* **10.1101/2021.08.30.458229** (2021).
338. M. J. Bogan, G. R. Agnes, F. Pio, R. B. Cornell, Interdomain and membrane interactions of CTP:phosphocholine cytidyltransferase revealed via limited proteolysis and mass spectrometry. *J Biol Chem* **280**, 19613-19624 (2005).
339. B. Mesmin *et al.*, Two lipid-packing sensor motifs contribute to the sensitivity of ArfGAP1 to membrane curvature. *Biochemistry* **46**, 1779-1790 (2007).
340. G. Drin *et al.*, A general amphipathic alpha-helical motif for sensing membrane curvature. *Nat Struct Mol Biol* **14**, 138-146 (2007).
341. C. W. Wang, Lipid droplet dynamics in budding yeast. *Cell Mol Life Sci* **72**, 2677-2695 (2015).
342. Y. A. Klug *et al.*, Mechanism of lipid droplet formation by the yeast Sei1/Ldb16 Seipin complex. *Nat Commun* **12**, 5892 (2021).