CHAPTER 1

INTRODUCTION

1.1 MOTIVATION

A big part of my research is simulating proteins in membranes using molecular dynamics (MD) simulations. As is well known, MD is notoriously slow and can simulate limited time (microsecond, with a few examples going into the millisecond range [1, 2], or if you have a million computers at your disposable, get a 0.1 second simulation [3]) and length (nanometer) scales, and thus, despite being able to model processes as complex and as specific as a ligand binding to a protein, any process that happens on a larger scale is beyond its scope. One of these processes is the oligomerization of membrane proteins in a multicomponent bilayer, a phenomenon that plays an important role in protein signaling and function (one of the examples of interest to me is CXCR4, a G-protein coupled receptor that binds to a chemokine and induces cell migration through chemotaxis [4, 5]). This clustering of GPCRs is affected by several factors, including lipid composition of the membrane [6], actin cyotskeleton [7], and other transmembrane proteins [8]. To model this clustering, we will use dissipative particle dynamics (DPD) which allows the modeling of a lipid bilayer with several proteins residents. Of the different factors affecting clustering, we decide to take lipid composition into consideration in this work, with the possibility of one type of lipid separating into a "lipid raft".

1.2 GOALS

We aim to study the effect of the presence of trans-membrane proteins on a multicomponent lipid bilayer and whether the proteins could trigger a phase separation in a bilayer that would rather slightly mix. The aim could be divided into the following sub-goals:

1. Define the "interesting" regime in a 2-component membrane, where a small perturbation can induce a phase transition in the system
2. Figure out a way to incorporate "rigid" transmembrane proteins using the soft forcefield of dissipative particle dynamics (DPD)
3. Find suitable metrics to quantify the phase transition or clustering of lipids around a protein

CHAPTER 2

METHODS

2.1 SETTING UP THE SYSTEM IN DPD

There are two main elements that we need to set up in the system, the first is the multi-component bilayer, and the second is the transmembrane protein.

2.1.1 MULTI-COMPONENT BILAYER

We use the "compositelamella" state in the DPD input file to generate a two-component bilayer with the following parameters:

State compositelamella
Polymers Lipid CoLipid
Normal 0 0 1
Centre 0.5
Thickness 5.0
Linearise 1
UpperFraction 0.5 0.5
Patches 1 1
Polymerise 0

This will create a patched bilayer that will either remain phase separated or mix with the other lipid, depending on the conservative parameters chosen. A more detailed study of the parameters and the phase transition can be found in Sec. 3.1.

In the table below we show the parameters used in the simulation of the bilayer, with H and T being the head and tail group of the lipid, W being water beads, and G and L being the head and tail group of the colipid respectively. The parameter denoted as X in the table is the cross interaction between T and L, the lipid and colipid tails, and will be the variable used to study the phase transition in a two-component membrane. The box size was 24 24 24 units.

```
bead    H
        0.5
        30
        4.5

bead    T
        0.5
        35    10
        4.5   4.5

bead    W
        0.5
        30    75    25
        4.5   4.5   4.5

bead    L
        0.5
        30    X     75    25
        4.5   4.5   4.5   4.5

bead    G
        0.5
        30    35    30    35    30
        4.5   4.5   4.5   4.5   4.5
```

Each lipid molecule is a double tail lipid with three head beads and eight tail beads, as shown below:

```
Polymer    Water    0.977986    " (W) "
Polymer    Lipid    0.011007    " (H H (* (T T T T)) H T T T T) "
Polymer    CoLipid    0.011007    " (G G (* (L L L L)) G L L L L) "
```

2.1.2 MEMBRANE PROTEINS

Given the nature of the proteins we are trying to mimic (GPCRs), which have seven trans-membrane helices that resemble a barrel , it would be appropriate to model them with a cylinder of beads. GPCRs have a hydrophobic trans-membrane region and an abundance of charged residues on the intra-cellular and extra-cellular sides (Fig. 2.1:left), as such, we model the GPCR as a cylinder with hydrophobic beads in the membrane and hydrophilic beads on the surface.

In order to create barrel-like trans-membrane proteins, we transform a cylinder of lipids in the bilayer into a new type of bead and then polymerise the beads in order to form a single polymer representing our protein. Tail groups (T and L) are transformed into the hydrophobic core, while head groups (H and G) are transformed into the hydrophilic extremities of the protein. The following DPD code performs this process:

```
Command SelectBeadTypeInCylinder    1    cylinder1 0 0 1 0.7 0.7 0.5 0.2667 0.0 2.0
Command SelectPolymerTypeHeadInCylinder    1    cp1    CoLipid 0 0 1 0.7 0.7 0.5 0.2667 0.0 2.0
Command PolymerisePolymersInTarget    1    cp1    12 1.5 1.0 128.0 0.5
Command ChangeNamedBeadType    1    cylinder1    TL
Command SetDPDBeadDissInt    1    TL    TL 4.5
Command SetDPDBeadConsInt    1    TL    TL 5
```

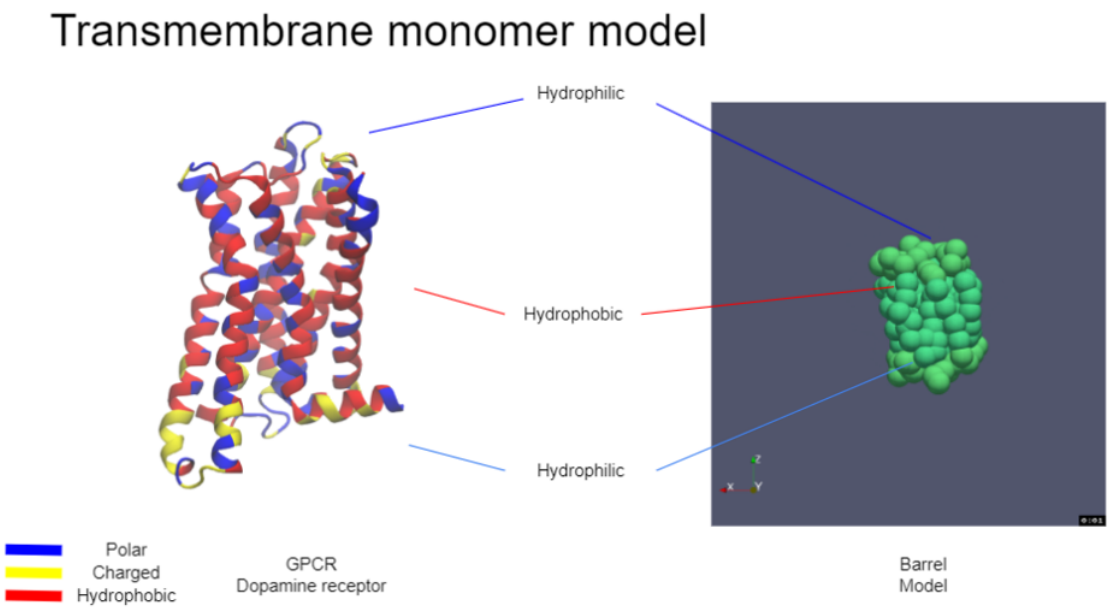

FIGURE 2.1
Trans-membrane monomer model with LEFT: structure of dopamine D2 GPCR where we can see the seven trans-membrane helices as well as the distribution of hydrophobic, charged, and polar residues. RIGHT: barrel model in DPD with hydrophilic beads on the intra and extra-cellular sides and hydrophobic beads at the core of the membrane.

Since the transformed beads are polymerised with an attractive conservative parameter, the lipid will shrink after the transformation. We set the conservative self interaction of the protein to 5 (lipid/colipid self interaction is 10) so that it is not favorable for the lipid to penetrate into the protein, and it works!

As for the interaction between the protein and the lipids, we set the command at 10000 steps so that the membrane and the protein had some time to equilibrate,

Command SetDPDBeadConsInt 10000 TL T Y

Where Y is a variable parameter that is further studied in the results section.

2.2 TROUBLESHOOTING

The sudden shrinkage of a cylindrical section of the lipid membrane seems the cause a decrease in density that induces a hole in the membrane, especially at higher values of repulsion between the protein and one/both of the lipid species. The hole uses the protein as a seed in most of the simulations where this is seen.

We thought of two ways to ameliorate this problem, either by increasing the box size (from 24 to 32) so bilayer has space to breath, or simply increasing the lipid fraction from 0.010507 to 0.011007 to compensate for the lost density in the lipid to protein transformation, as seen in Fig. 2.3. Due to time constraints, we decided to increase the lipid density without increasing the box size, so simulations would not take a long time to run.

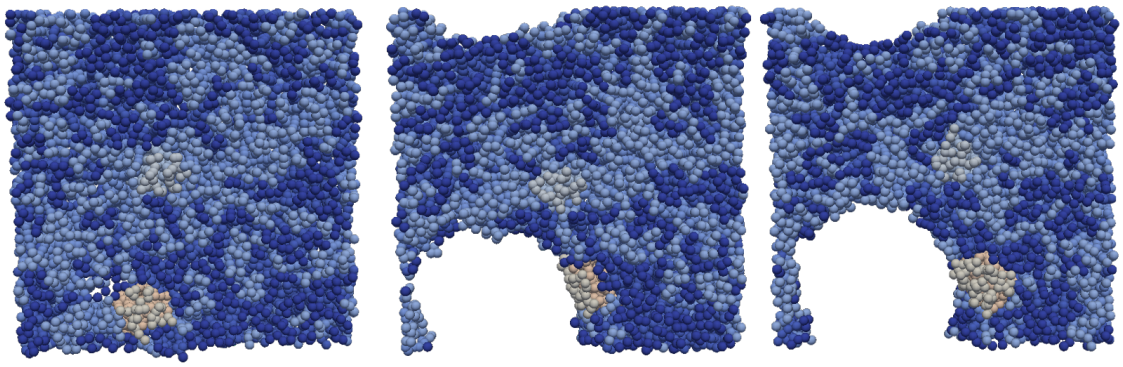

FIGURE 2.2
Hole in the membrane rapidly growing and using the protein as a starting point. The hydrophylic part of the protein and the lipid heads rotate toward the hole.

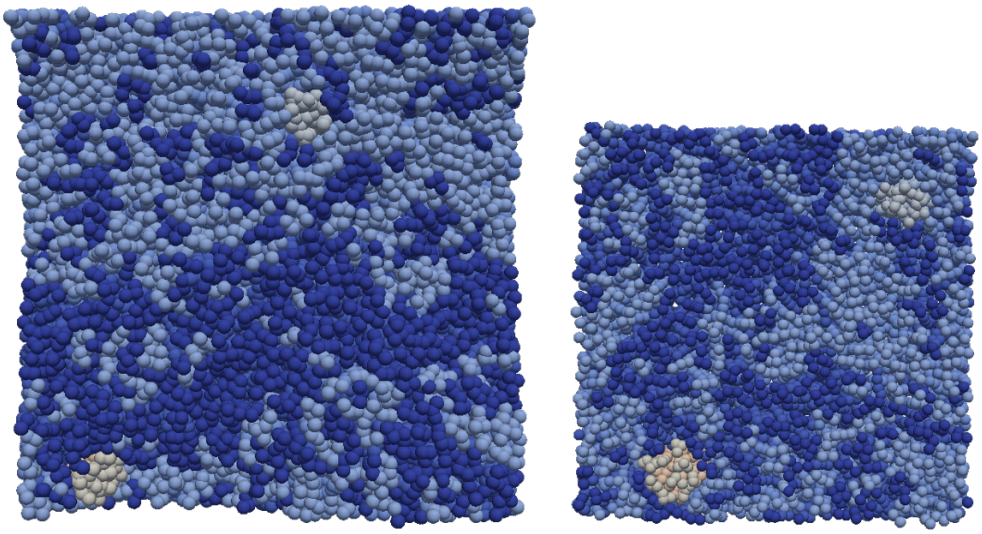

FIGURE 2.3
Two simulations with the same parameters as that shown in Fig. 2.2 but without holes in the bilayer. LEFT: increasing the box size from 24 to 32 units allows the membrane to breath. RIGHT: increasing lipid fraction from 0.010507 to 0.011007 compensates for the lost density due to lipid to protein transformation.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 THE INTERESTING REGIME IN A 2-COMPONENT BILAYER

We define "interesting" as the regime where a small perturbation (thermal fluctuations for example) or a not so small but not large perturbation (protein inserted into the membrane, dimerization of membrane proteins, ...) can lead to a change in behavior or a change in phase of the membrane. I could envision doing this by modifying different conservative force parameters within the simulation. These could be:

1. The two lipids have different self-interaction.
2. One of the lipid tails is more hydrophobic than the other.
3. The two lipids have the same self-interaction but different cross interaction.

We decided to go with option 3, changing the cross interaction of the different lipids, while not touching the self-interaction. We define the ratio between the cross-interaction of the lipid tail (T) and colipid tail (L) and self-interaction of L or T as:

$$\gamma = \frac{\text{Cross interaction between L and T}}{\text{Self interaction of L or T}}, \tag{3.1}$$

or if the self-interaction of L is different than that of T, one could define:

$$\gamma = \frac{(\text{Cross interaction between L and T})^2}{\text{Self interaction of L} \times \text{Self interaction of T}}. \tag{3.2}$$

Throughout this work, the self interactions of the two species of lipids will always be identical, and unless otherwise mentioned, the fractions of lipid and colipid are both at 0.5. For this section, We ran 5 replicas with 300,000 steps each, totaling a sum of 1,500,000 steps for every value of γ tested. At $\gamma = 1$, the cross interaction is equal to the self interaction and the two species cannot differentiate each other. At $\gamma = 1.1$, the colipid (red, Fig. 3.1) shows jagged interfaces but showing an almost phase separated membrane. Already at $\gamma = 1.2$, we can see a clear phase separation of the two species (Fig. A.1). I hypothesize that a small change in γ is sufficient since there are eight tail beads in the lipid tail, which amplifies the small shift in interaction.

To quantify the phase transition seen in Fig. 3.1, we use the radial distribution function (RDF) which will count how many lipid beads we have in a spherical shell around a given lipid bead (Even though the bilayer we are dealing with is planar in nature and a 2D RDF would have fit better. A side effect of using a 3D RDF is the second peak that we see in Fig. 3.2 around $r = 4$, which is simply reflecting the shape and thickness of the bilayer). We count the colipid head group (called G in the input file) after 200,000

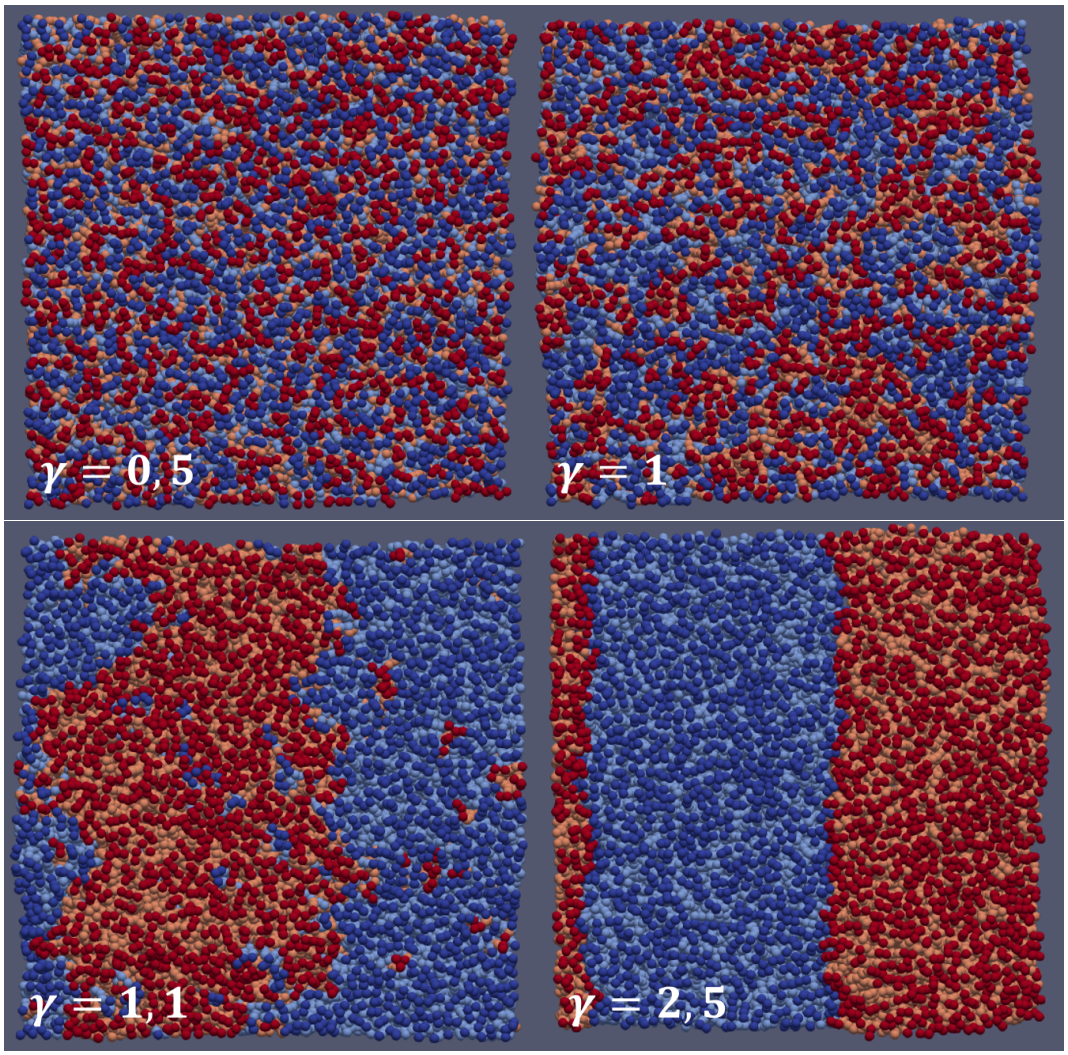

FIGURE 3.1
The different observed regimes in the simulation, varying from total mixing ($\gamma = 0.5$, top-left) to marginal mixing to total separation ($\gamma = 2.5$, bottom-right).

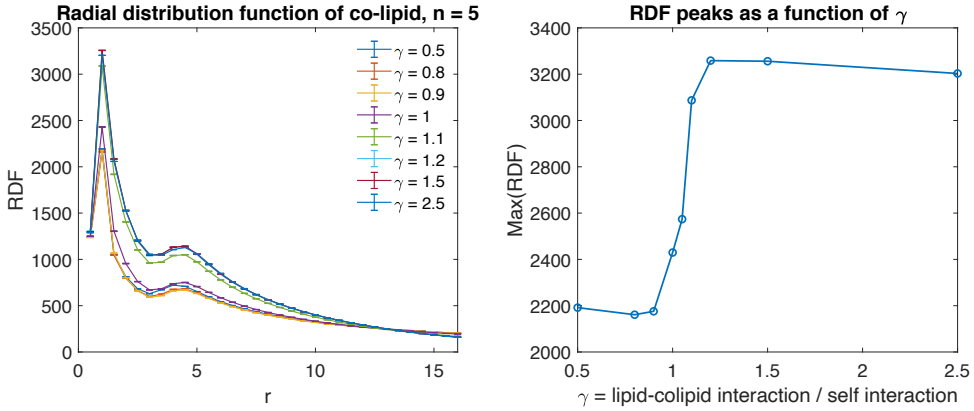

FIGURE 3.2
LEFT: RDFs for co-lipid for different γs. RIGHT: Peaks of RDF plotted as a function of γ showing the phase transition around $\gamma = 1.05$.

time steps have passed in every simulation so we capture the equilibrated distribution. As seen in Fig. 3.2, we can observe a clear transition using this metric, especially when we plot RDF peaks versus γ. The transition occurs in the narrow range $0.9 < \gamma < 1.2$. We choose the center of this range, $\gamma = 1.05$ which is at the heart of the transition for the following simulations.

3.2 EFFECT OF PROTEIN-LIPID INTERACTION AND NUMBER OF PROTEINS

We add membrane proteins into the bilayer with varying interactions with the lipid species and observe the effect the proteins have on the membrane. As mentioned in the methods sections, proteins are created from the lipids of the bilayers by polymerizing a cylinder and then setting the interactions of the new protein beads with themselves and with their lipid environment. The variable interactions are the conservative parameters between protein - T and protein - L (the tails of the lipid and colipid). We aim at having the proteins being repelled to one lipid less than the other, but at the same time we want the lipids to prefer each other rather than being inside the protein so in short: lipids prefer themselves, protein prefers one lipid over the other. Another variable we will be considering here is the number of proteins in the bilayer. We study the evolution of the bilayer by measuring the RDF every 50,000 steps using an analysis period of 10,000 (so we try to measure an "instantaneous" RDF rather an averaged one over an extended period of time). We ran 5 replicas with 1,000,000 steps each, totaling a sum of 5,000,000 steps for every combination of protein-lipid interaction and number of proteins.

The RDFs with proteins show a similar shape to that which we have seen in Fig. 3.2 with a first peak measuring the clustering of lipid heads and the second peak reflecting the thickness and shape of the membrane. We are interested in the first peak, and thus we will take the peaks and plot them as a function of time, as seen in Fig. 3.3.

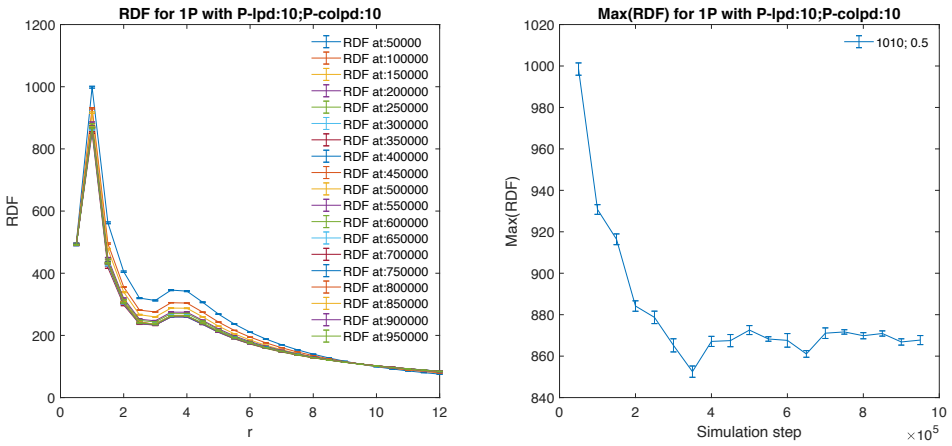

FIGURE 3.3
LEFT: RDFs for co-lipid head group G with one protein in the bilayer at different points of the simulation. RIGHT: Peaks of RDF plotted as a function of time showing the equilibrium state of the system.

For only one protein in the membrane, changing the interactions seems not to have an effect, as the RDFs are very close to one another (check Fig. A.2 in the appendix), even though in the simulation we do observe a small cluster of colipid around the protein (Fig. A.3), but the effect is too small to appear in the RDFs. In order to explore the effect of changing the interactions, we will compare using simulations with higher number of proteins. We start observing a slight effect in the RDFs at three proteins, with the effect becoming more and more pronounced at four and five as seen in Fig. 3.4 right. Note that the total number of beads is different when number of proteins is different since we are "transforming" lipid beads into

protein beads with the same total number of lipids conserved between simulations, so the more proteins there are, the less lipid molecules exist. At five proteins we can already observe a difference in the RDF, but how far are we from a real phase transition?

To quantify the RDF in a phase separated bilayer, we set γ to 1.2 (Fig. 3.2:right shows where that value of γ is on the transition), and the result is that the phase separated RDF shows equilibrated values around 970, as opposed to 780 for P-T:20;P-L:10 and 760 for P-T:10;P-L:10 as seen in Fig. 3.4:left. The three different states are shown in snapshots in Fig. 3.5, The increase in P-T:20;P-L:10 accounts for 10% of the full transition, so there is still room for improvement.

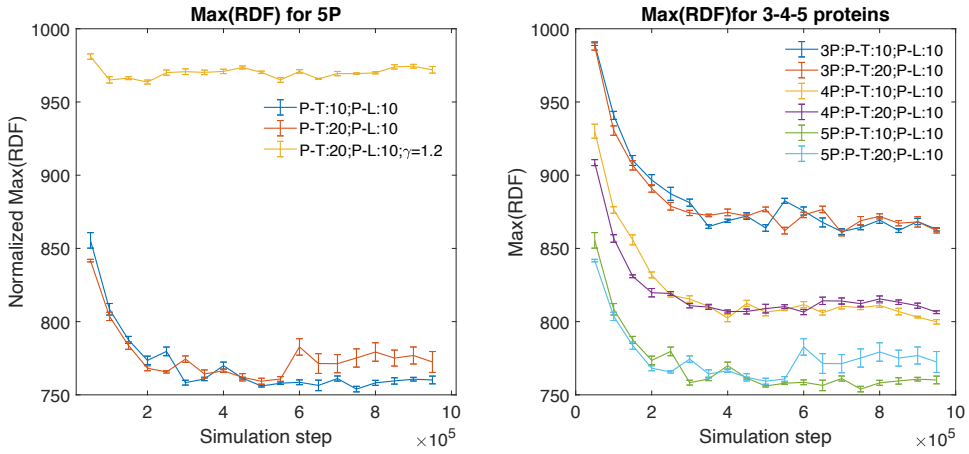

FIGURE 3.4
RDFs for co-lipid head group G with five proteins in the bilayer with a different set of interactions and a γ value that shows a phase separation (left) and compared with three and four proteins (right). P-T is the conservative interaction between the proteins and lipid tail, while P-L is that of the protein and the colipid tail.

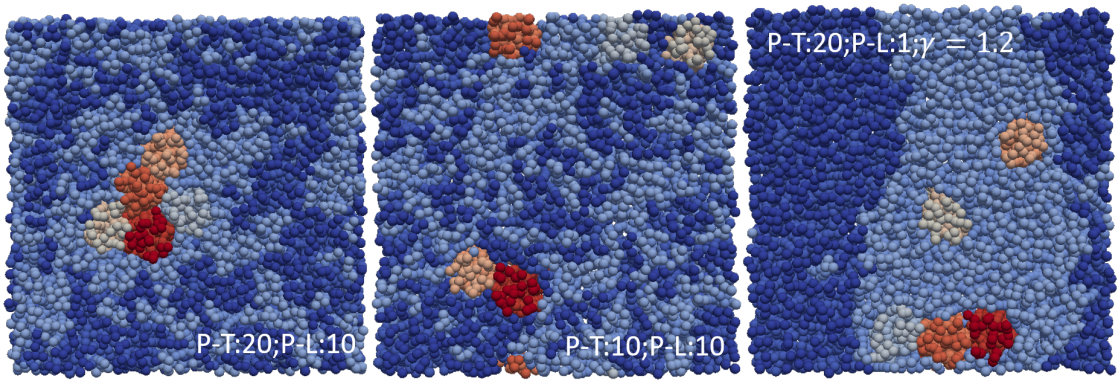

FIGURE 3.5
Snapshots of the simulation with five proteins for the three different set of parameters shown in Fig. 3.4:left with an oligomerization of the proteins as well as formation of a colipid domain around the oligomer (left), indifference toward any type of lipids (middle), and total phase separation (right).

3.3 EFFECT OF FRACTION OF COLIPID

We could think of two ways to see the phase transition more clearly, the first is by decreasing the fraction of the colipid (the lipid that the protein likes), and the other is by shifting γ to a higher value that is closer to the phase transition. In this section I discuss the effect of decreasing the fraction of colipid on the

bilayer. Instead of having the usual half-half distribution of lipids, we try having the colipid fraction (that the protein likes) at 0.25 while the lipid is at 0.75, while maintaining the same total amount of lipids. We want to see if the colipid will form domains around the protein. I call the colipid fraction "cf" for short.

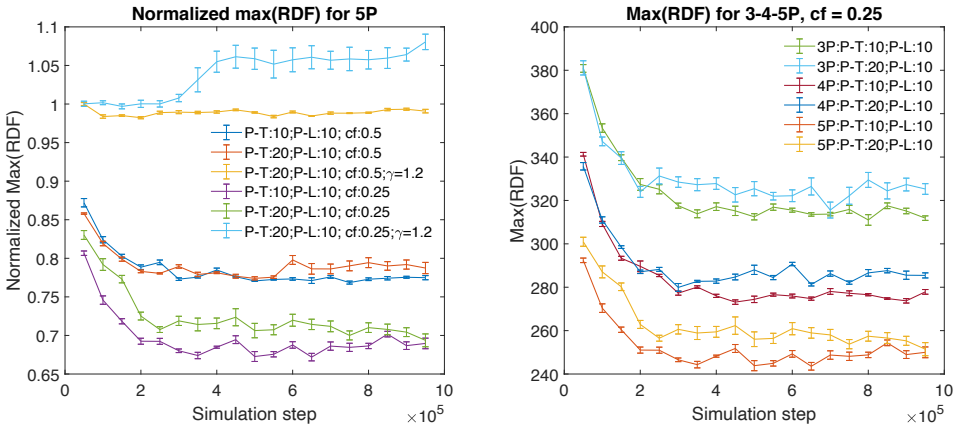

FIGURE 3.6

LEFT: Normalized RDF comparison between colipid fraction $cf = 0.5$ and $cf = 0.25$ for five proteins in the bilayer. RIGHT: RDF comparison between three, four, and five proteins for $cf = 0.25$.

Similar to Fig. 3.4:left, we compare the five proteins simulation to an idealized case when $\gamma = 1.2$ (total phase separation) in Fig. 3.6:left. Three striking differences can be seen between $cf = 0.5$ and $cf = 0.25$:

1. Early increase of light green over purple compared to red over blue, which only starts increasing in the second half of the simulation (where light green and red are the simulations when the proteins are more attracted to the colipid)
2. Increase in RDF for the control case for $cf = 0.25$ (light blue) over time compared to the decrease in the control case for $cf = 0.5$ (yellow). This is due to the colipid forming circular domains around the protein cluster rather than the rectangular domains at the beginning of the simulation.
3. Difference in the normalized "steady states" for the two values of cf (look at purple and blue for example), with the reason being the transformation of lipid to make protein molecules, if there is a absolute number of lipids molecules, the same number of beads "lost" to become proteins creates this effect

Another interesting effect for having a smaller fraction of colipid is that a smaller number of proteins can start showing a difference in the RDF (compare Fig. 3.6:right to Fig. 3.4:right), where even with three proteins in the bilayer, we can see an increase in the RDF over the control.

What about the case when a phase separation is forced by setting $\gamma = 1.2$, would we see any difference between a bilayer with or without proteins? Looking at the trajectories gives us a hint that proteins act as catalysts for a "rectangle" to disk transformation, as seen in Fig. 3.7, where the disks surrounding the protein appeared in all five runs, while the disks showed in only two of the five simulations for a proteinless bilayer.

This transformation was quantified by plotting the normalized peaks of the RDF as a function of time, where the normalization is the RDF of the phase separated state at $t = 50,000$ (since this was the first recorded RDF). The larger the number of proteins, the larger (and the earlier) the observed increase in the RDF (Fig. 3.8). While the faster increase of the RDF is significant, I believe the relative increase of the normalized RDF is an artifact of (1) the smaller number of lipids in the membrane as we increase the number of proteins, since we can observe the disk domain already at one protein, and (2) holes in the

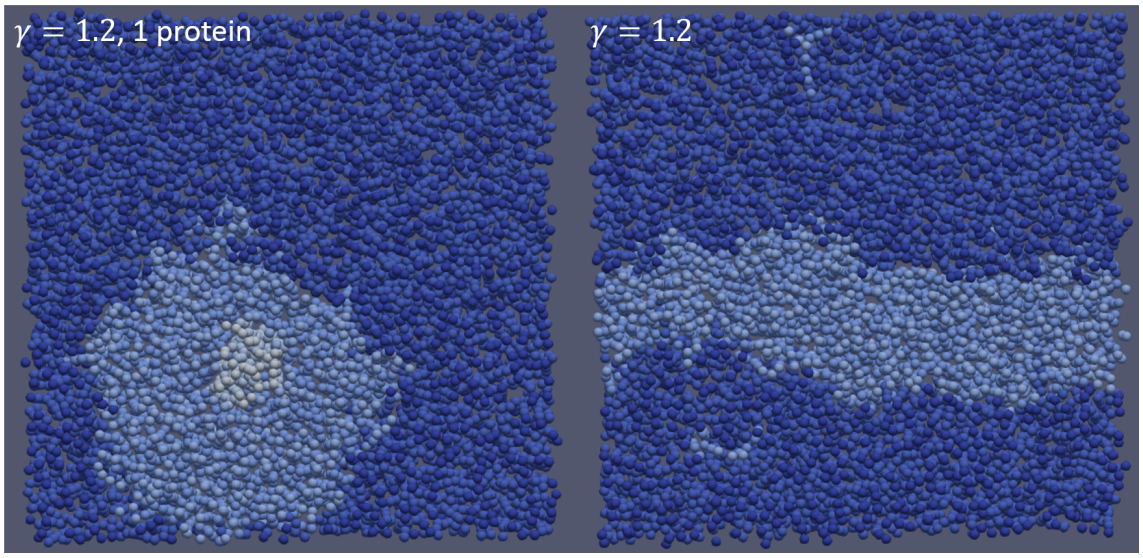

FIGURE 3.7
LEFT: Colipid forming a disk around the protein with $\gamma = 1.2$ en all five simulations. RIGHT: Colipid staying in a rectangular formation similar to the inital state of the bilayer in three out of five simulations.

membrane showing up (similar to those discussed in Sec. 2.2) in some of the simulations in the case of higher number of proteins, as hinted at by the larger error bars for 3P and 5P and as seen in Fig. A.4.

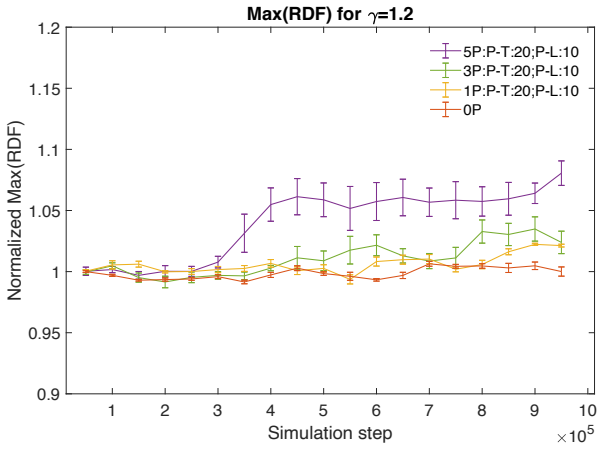

FIGURE 3.8
LEFT: Colipid forming a disk around the protein with $\gamma = 1.2$ en all five simulations. RIGHT: Colipid staying in a rectangular formation similar to the inital state of the bilayer in three out of five simulations.

CHAPTER 4

CONCLUSION

Even though DPD does not describe the details of the molecular interaction between a membrane protein and the surrounding bilayer, it provides a framework to simulate larger scale phenomena that are not accessible to atomistic simulations.

In this work we have explored the effect of clustering proteins in a multi-component bilayer by first defining a dimensionless order parameter to describe the phase transition of a two-component bilayer and then quantifying the effect one or more proteins have on the phase separation of the lipids in the bilayer when the protein prefers one type of lipid over the other. As expected, more proteins in the membrane had a larger effect, with the effect being more pronounced when the fraction of the colipid (that the protein likes) is lower than that of the lipid.

BIBLIOGRAPHY

[1] Vincent A Voelz et al. 'Molecular simulation of ab initio protein folding for a millisecond folder NTL9 (1- 39)'. In: *Journal of the American Chemical Society* 132.5 (2010), pp. 1526–1528.

[2] Kresten Lindorff-Larsen et al. 'How fast-folding proteins fold'. In: *Science* 334.6055 (2011), pp. 517–520.

[3] Maxwell I Zimmerman et al. 'SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome'. In: *Nature Chemistry* (2021), pp. 1–9.

[4] Beili Wu et al. 'Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists'. In: *Science* 330.6007 (2010), pp. 1066–1071.

[5] Laura Martinez-Munoz et al. 'Separating actin-dependent chemokine receptor nanoclustering from dimerization indicates a role for clustering in CXCR4 signaling and function'. In: *Molecular cell* 70.1 (2018), pp. 106–119.

[6] Kristen A Marino et al. 'Impact of lipid composition and receptor conformation on the spatio-temporal organization of μ-opioid receptors in a multi-component plasma membrane model'. In: *PLoS computational biology* 12.12 (2016), e1005240.

[7] Marco Scarselli, Paolo Annibale and Aleksandra Radenovic. 'Cell type-specific β2-adrenergic receptor clusters identified using photoactivated localization microscopy are not lipid raft related, but depend on actin cytoskeleton integrity'. In: *Journal of Biological Chemistry* 287.20 (2012), pp. 16768–16780.

[8] Ioanna Bethani et al. 'Spatial organization of transmembrane receptor signalling'. In: *The EMBO journal* 29.16 (2010), pp. 2677–2688.

APPENDIX A

SUPPLEMENTARY INFORMATION

A.1 ADDITIONAL FIGURES

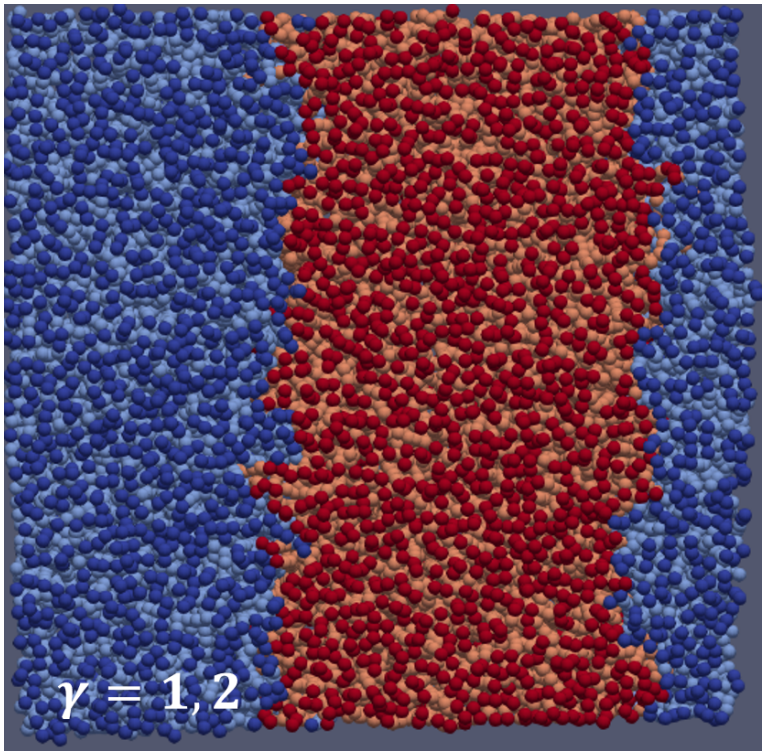

FIGURE A.1
Simulation snapshots of the bilayer at $\gamma = 1.2$ showing clear phase separation with a jagged interface between the two phases

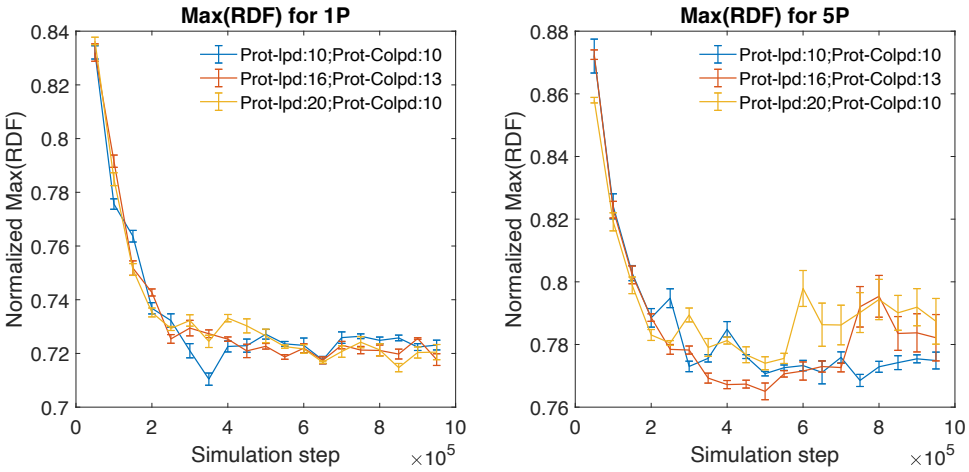

FIGURE A.2

LEFT: RDFs for co-lipid head group G with one protein in the bilayer versus time at three different values of conservative interactions, with 10-10 (blue) being the control case when the protein cannot tell the difference between the two lipids. Compare this to RIGHT: where we see that having fice proteins in the bilayer shows a tangible effect in the RDF

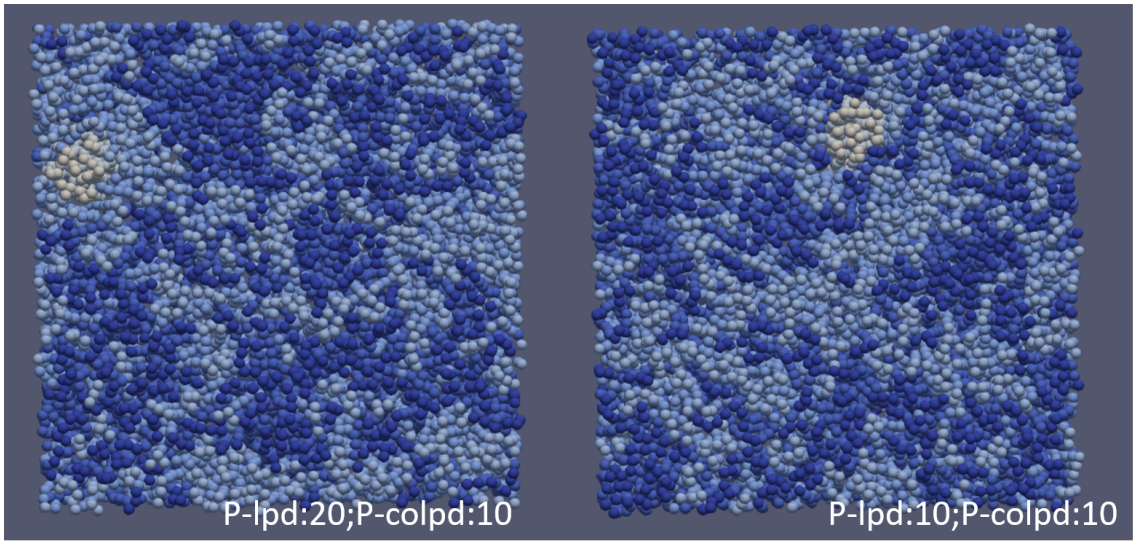

FIGURE A.3

Simulation snapshots of one protein in a 2-component bilayer with the left showing a cluster of colipid (light blue) around the protein (white) for the conservative interactions shown on the image versus right showing a random distribution of lipid (dark blue) and colipid (light blue)

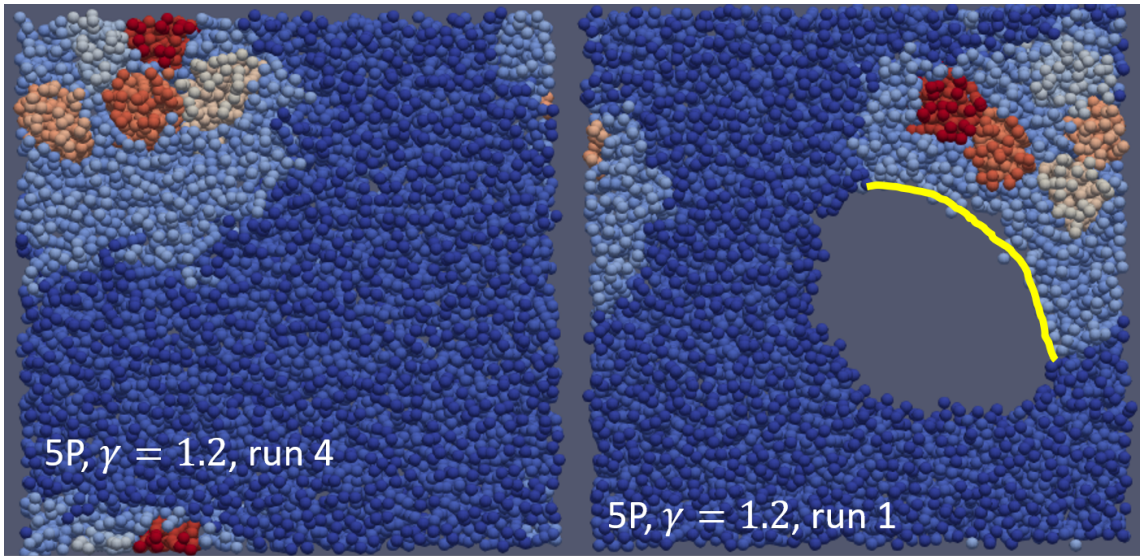

FIGURE A.4
Simulation snapshots of five proteins in a 2-component bilayer with $\gamma = 1.2$ which forces phase separation of lipids. $P - T = 20$ and $P - L = 10$. Left shows a run without a hole showing up while right shows one of the runs with the hole. Notice that the hole creates a new interface for the colipid head group with water (along the yellow marked line)