

# Self-Organization of Alpha Helical Proteins in Bioinspired Membranes and Vesicles

Akash Banerjee,<sup>1</sup> Zachary Finkel<sup>2</sup> and Meenakshi Dutt<sup>1,\*\*</sup>

## ABSTRACT

A fundamental understanding of protein-protein and protein-lipid interactions under various conditions can reveal the energy pathways in photosynthetic bacterial membranes. In this study, we examine the role of key factors such as bilayer curvature, the concentration, helical separation and hydrophobic mismatch of proteins on their self-organization in bilayers. We also develop an understanding of the physical factors underlying the aggregation of proteins. We determine the impact of bilayer curvature by comparing the aggregation of proteins in membranes and vesicles. We identify a threshold helical separation below which small, stable aggregates are observed. Large, unstable protein aggregates are observed above the threshold separation. We examine the effect of the deformations incurred by the proteins via their concentration, and show the aggregation of the proteins to arise from their deformation-induced displacement. We demonstrate the negative hydrophobic mismatch condition to favor a higher degree of protein aggregation. We

Author Information

\*\* corresponding author: [meenakshi.dutt@rutgers.edu](mailto:meenakshi.dutt@rutgers.edu)

<sup>1</sup>Chemical and Biochemical Engineering and <sup>2</sup>Biomedical Engineering,

Rutgers, The State University of New Jersey, Piscataway, NJ 08854

**Supporting Information File.** Building the Lipid Membrane, Building the Lipid vesicle, Protocol to Organize Proteins in Lipid Bilayer, Protocol to Organize Proteins in Lipid Vesicle, Choice of Helical separations, Choice of Cutoff Distance for Cluster Characterization, Solvation of Bilayer and Vesicle, Measurement of Hydrophobic Thickness of Bilayer, Measurement of Hydrophobic Mismatch, Protein Tilt Calculations, Building Protein-Bilayer systems by Varying Helical Spacing and Molecular Dynamics Software Package

adopt the Molecular Dynamics simulation technique along with a coarse-grained force field to capture the behavior spanning extensive spatiotemporal scales. Our results can guide experimental studies of bioinspired materials with structure-function properties mimicking those of photosynthetic bacterial membranes, or assist in understanding the organization of inclusions in polymeric matrices.

Keywords: Photosynthetic Bacterial Membrane, Transmembrane Proteins, Lipid Bilayer and Vesicle, Molecular Dynamics, Inclusions in Soft Materials

## INTRODUCTION

Photosynthetic bacterial membranes <sup>1</sup> can inspire the design of bioinspired materials. Of specific interest are purple bacterial membranes <sup>2</sup> which have energy pathways via the membrane <sup>3</sup> facilitated by transmembrane proteins <sup>4,5</sup> functioning as proton pumps. <sup>6</sup> These transmembrane proteins encompass two key components, namely a light harvesting complex and a reaction center <sup>7</sup> which form a core complex with the ability to oligomerize. Yet, very little is known of the impact of protein-lipid interactions on protein-protein interactions and the self-organization of proteins on the energy pathways in these photosynthetic bacterial membranes. A fundamental understanding of the self-organization of the proteins can aide in mapping out the energy transfer mechanisms in similar photosynthetic membranes. Our study is focused on developing this fundamental understanding through the use of model systems which examine the self-organization of proteins in bioinspired lipid membranes and vesicles.

Prior experimental investigations have identified key factors which control protein-lipid and protein-protein interactions. Membrane curvature has been identified to determine the degree

of protein aggregation.<sup>8,9</sup> Studies have found the concentration of proteins to affect their self-organization into aggregates,<sup>9</sup> with the activity of the aggregates having a sigmoidal dependence on the concentration.<sup>10</sup> Experimental investigations have suggested the dimerization of proteins to be dependent upon their spatial configuration and conformations.<sup>11</sup> However, these studies have been unable to resolve the role of helical separation on the mechanisms and processes underlying the aggregation of proteins.<sup>12</sup> Another factor known as hydrophobic mismatch<sup>13,14</sup> has been known to cause protein-induced deformation of lipids and lipid-induced deformation of proteins,<sup>15</sup> thereby impacting the aggregation of proteins. Yet, limitations in the spatiotemporal resolution of the experimental approaches hinders the identification of the impact of each of these factors on the self-organization of proteins in bilayers. These limitations can be addressed by computational approaches.

Computational approaches have been adopted to characterize the energetics associated with protein-protein interactions as function of the distance of separation,<sup>16,17</sup> with the purpose of identifying the thermodynamics of protein aggregation.<sup>18</sup> Other computational approaches<sup>17,19,20</sup> have been adopted to examine the effect of mismatch between the hydrophobic sections of proteins and lipids, and reported hydrophobic mismatch to result in a higher degree of protein aggregation. Earlier studies<sup>19</sup> have also investigated the effect of curvature on the self-organization of proteins through comparison of the aggregation of proteins in membranes and vesicles. The mean residence times<sup>19</sup> for the average cluster size in membranes and vesicles have been observed to depend upon the orientation of the proteins. In addition, investigations on the effects of protein-induced lipid deformations have been reported to vary with the degree of mismatch and size of the protein.<sup>15</sup> These deformations have been surmised to result in protein aggregation, in order to reduce membrane stress.<sup>21</sup>

In this study, we examine the effect of various factors on the self-organization of proteins in bilayers. Furthermore, our study develops an understanding of the key physical factors underlying the self-organization of proteins into clusters. We investigate the role of bilayer curvature, protein concentration, their initial helical separation and hydrophobic mismatch on their aggregation in bilayers. We would like to note that this is one of the first investigations to examine the isolated impact of bilayer curvature, protein concentration, helical separation and hydrophobic mismatch on the self-organization of proteins in bilayers and vesicles. Due the extended spatiotemporal scales relevant to the problem at hand, the investigation employs the Molecular Dynamics simulation technique in conjunction with coarse grained representations of the molecules.

Our investigations on protein-protein interactions in a lipid bilayer can guide the study of bioinspired materials where the chemistry of the molecular components can yield materials with structure-function properties mimicking those of photosynthetic bacterial membranes. Furthermore, the results from our investigations can be used to provide insight into the predominant physical factors in any experimental setup involving organization of inclusions or macromolecules in polymeric matrices

## **METHODS**

The dynamics underlying the self-organization of proteins<sup>22</sup> in lipid membranes (Figure SI.1) and vesicles (Figure SI.2) typically occurs over extended spatial and temporal scales. Hence, the dynamics of the system is resolved via the Molecular Dynamics (MD) simulation technique used in conjunction with coarse-grained (CG) representations of the molecular components. The bonded and non-bonded interactions<sup>23</sup> between the CG beads is modeled via the explicit solvent

Martini force field.<sup>24</sup> The Lennard-Jones potential is shifted to zero from 0.9 nm to 1.2 nm.<sup>25</sup> The Coulomb electrostatic potential is shifted to zero from 0 and 1.2 nm. The cut-off distance ( $r_c$ ) for the van der Waals and Coulombic interactions is 1.2 nm.

The steepest descent algorithm is used for the energy minimization process. Initially, all the simulations are equilibrated using a time step of 10 fs. Following equilibration, the time step is increased to 20 fs. The simulations performed in the canonical ensemble use a v-rescaling thermostat with a stochastic term,<sup>26</sup> with the coupling constant maintained at  $\tau = 10$  ps. The simulations performed in the NPT ensemble employ a Parrinello-Rahman barostat,<sup>27,28</sup> with the pressure set at 1 bar. The particle trajectories of the systems are computed using the GROMACS<sup>29,30</sup> and LAMMPS<sup>31</sup> simulation packages.

### **Coarse-grained Models for Lipid Membranes and Vesicles**

A pre-assembled lipid membrane encompassing 1348 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) molecules is placed in a simulation box with three dimensional periodic boundary conditions. The simulation box dimensions are given by 30 nm  $\times$  30 nm  $\times$  10 nm. Each CG representation of a DPPC molecule (Figure 1.A) includes two charged head beads (namely, Choline and Phosphate), two non-polar beads (i.e., Glycerol) and 2 saturated acyl tails consisting of four apolar beads each. The CG bonded and non-bonded interaction parameters are obtained from a custom membrane generator program.<sup>32</sup> Following the energy minimization process for 3000 steps, the lipid membrane is equilibrated for 400 ns in a semi-isotropic NPT ensemble with a target pressure of 1 bar. To ensure that the DPPC bilayers are in the fluid phase for this study, we perform the simulations at 323 K. The x- and y- dimensions of the membrane are coupled to each other to maintain the membrane in a tensionless state.<sup>33</sup> This equilibrated lipid bilayer is used

to host protein inclusions. Details of the construction of the pre-assembled membrane are provided in the Supporting Information (SI).

A pre-assembled DPPC vesicle is also used for the study. The characteristics of the vesicle are based upon those associated with a vesicle obtained using the protocol detailed in the SI. The simulations performed in the NPT ensemble employ isotropic coupling of the barostat.<sup>34</sup>

### **Coarse-grained Model for Proteins**

One of the key factors underlying the self-organization of transmembrane helices in bilayers is the ability of the core complex to dimerize. The protein complex is modeled as an alpha helix that will oligomerize with neighboring proteins. The CG model for KALP 23 (that is, lysine (K), alanine (A), leucine (L) and 23 is the number of amino acid residues in the alpha helix), a transmembrane protein with an Alanine-Leucine helical sequence, is based upon earlier computational studies.<sup>35</sup> The alpha helical conformation of the protein is maintained via dihedrals with a force constant of 400 kJ/mol rad<sup>2</sup>.<sup>13</sup> The model for the transmembrane protein enables favorable interactions between the helix side chain and the phospholipid acyl tails. Additionally, a central pore is introduced for the diffusion of ions and solvent across the protein.<sup>19</sup> Whereas both the ends of the protein are identical, the charge on the N terminus is higher as compared to the C terminus. This difference is responsible for the protrusion of the N terminus from the hydrophobic core of the bilayer. In addition, the C terminus is placed at the interface between the solvent and the hydrophobic bilayer core, in concurrence with prior studies.<sup>18</sup> This charge difference helps reproduce the desired tilt of the protein in the bilayer, in agreement with spectroscopic data.<sup>15</sup> Each KALP protein has four units of positive charge corresponding to four positively charged Lysine

residues. In order to neutralize the charges in the system, four negatively charged chloride ions are added for each protein.

The impact of hydrophobic mismatch<sup>36</sup> on the self-organization of the proteins in lipid membranes and vesicles is studied via the introduction of two other proteins with hydrophobic sections of different lengths. We measure the hydrophobic mismatch by subtracting the hydrophobic thickness of the bilayer from the hydrophobic length of the protein (alpha-helical section). The hydrophobic section of the bilayer is estimated by finding the average distance between the glycerol beads on each monolayer. The approach adopted to eliminate inaccuracies in the estimation of hydrophobic thickness due to bilayer undulations is provided in the SI. The first protein has a shorter hydrophobic section (namely, KALP 19, Figure 1.B) as compared to the hydrophobic core of the lipid membrane so as to result in negative mismatch. The second protein has a longer hydrophobic section (namely, KALP 27, Figure 1.D) compared to the hydrophobic section of the bilayer, resulting in positive mismatch. KALP 23 (Figure 1.C) has zero hydrophobic mismatch with the lipid bilayer. The models for the KALP 27 and KALP 19 proteins are obtained through the addition/subtraction of alanine-leucine residues in the hydrophobic section of the model for the KALP 23 protein.<sup>36</sup> The anchoring groups (that is, lysine) at the extremities of the three different protein models are the same to ensure physical consistency.<sup>37</sup>

The models for the three proteins are validated through measurements of the tilt of each protein in a membrane along with its impact on the structural characteristics of the bilayer in its vicinity. These measurements require the substitution of four phospholipid molecules in a bilayer (see SI for details) with a single protein. KALP 19 is observed to tilt (relative to the bilayer normal) and immerse itself between neighboring phospholipid molecules,<sup>15</sup> thereby minimizing the exposure of the protein hydrophobic section to unfavorable interactions with the solvent. Whereas

KALP 27 is observed to tilt significantly and protrude out from the bilayer. Table SI.2 shows the results from the characterization of a single protein in a bilayer (for all three protein models). The measurements for hydrophobic mismatch are observed to be in qualitative agreement with corresponding results obtained using all atom simulations.<sup>20</sup> A quantitative comparison is limited by the resolution of the CG representation of the molecules.<sup>38</sup> Due to this discrepancy, the KALP 23 protein is measured to have a negative mismatch with the DPPC membrane whereas it is expected to have zero hydrophobic mismatch. However, measurements of hydrophobic mismatch of the proteins in vesicles demonstrate KALP 19, KALP 23 and KALP 27 to respectively have negative, zero and positive mismatch with the bilayer (see Table SI.3). A reduction in bilayer thickness can potentially arise from the stress induced by curvature effects of a vesicle,<sup>10,39</sup> which would result in zero hydrophobic mismatch for the KALP 23.

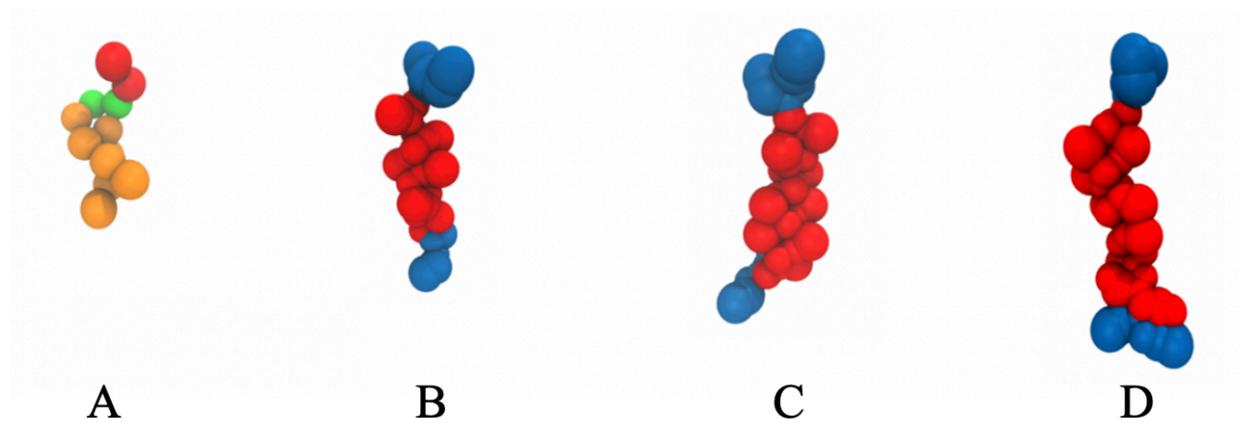


Figure 1: Images of (A) a DPPC phospholipid (yellow, green and red colors represent the lipid acyl tail, hydrophobic head and hydrophilic head beads), and KALP proteins with (B) 19, (C) 23 and (D) 27–amino acid residues forming the alpha helical section. Alanine and leucine are represented using red beads, and lysine is represented using blue beads.

The orientation of each protein in the bilayer is characterized by its tilt angle (with respect to the bilayer normal). The tilt angle is evaluated by representing the helical section as a set of data points spiraling around a cylindrical axis. This axis is assumed to define the axis of rotation of the alpha helix. The unit vector along this axis is computed using a non-linear regression tool called NLREG.<sup>40</sup> The tilt angle of the protein is given by the angle between the unit vector and the bilayer normal. The measurements for the tilt angle (see Table SI.4) are in good agreement with corresponding results from earlier studies<sup>20,41</sup> on the tilt values for proteins with negative hydrophobic mismatch.

The initial arrangement of the proteins in the bilayer is set up by controlling the helical separation between the proteins. This approach is adopted to understand the physical factors driving the equilibrium self-organization of the proteins in a bilayer. The proteins are arranged such that each protein is oriented antiparallel to the nearest neighboring protein.<sup>7</sup> Due to favorable interactions between N- and C-termini, this arrangement is assumed to favor dipole-dipole interactions<sup>42</sup> and result in the formation of stable clusters. The approach used in this study departs from that adopted in earlier studies<sup>43</sup> where proteins were arranged in the bilayer at random separations from each other.

The membrane-protein system is set up by replacing four phospholipid molecules with one protein molecule. The proteins are organized within the bilayer in an array (see Figure SI.3). The backbone of each protein is restrained via a harmonic constraint of 3000 kJ/ mol nm<sup>2</sup> until the acyl tails of the phospholipids in the vicinity of the proteins are sufficiently relaxed. The membrane-protein system is allowed to run for 500 ns without restraints followed by an additional 100 ns to obtain trajectories for analysis. Four independent particle trajectories are used to ensure reproducibility of the results. These trajectories have been generated by random seeds that provide

a unique initial velocity distribution for every new simulation. One must note that the initial position of proteins and lipids are not altered across the four seeds.

The details of the protocols used to set up the membrane-vesicle systems are provided in the SI. The vesicle-protein systems are simulated for a total of 4 microseconds, using four independent particle trajectories to ensure reproducibility of the results.

The particle trajectories from the MD simulations are used to study the impact of bilayer curvature along with protein concentration, helical separation and hydrophobic mismatch on their self-organization within a bilayer.<sup>44</sup> The values of the protein concentration are 15 KALP proteins (low concentration), 35 KALP proteins (medium concentration) and 150 KALP proteins (high concentration). The helical separations are varied from 1.2 nm to 2.0 nm. The hydrophobic mismatch between the protein and the bilayer are varied from positive, zero to negative.

### **Cluster Size Quantification**

The self-organization of the proteins into aggregates is characterized via the protein cluster counts and sizes. The cluster count measurements are performed using the center of mass of each protein as a reference. Hence, protein aggregates are classified as clusters if each protein occupant has its center of mass within 1.8 nm ( $\sim 1.5$  times the interaction cutoff distance) of the nearest neighboring protein. The cluster count is used to compute a characteristic called P value,<sup>9</sup> which is given by the ratio of the number of proteins to the number of clusters. The P value ranges from 1 (corresponding to no clustering) to the total number of proteins (corresponding to the case of a single cluster encompassing all the proteins). Hence, the P value provides an estimate of the average number of proteins per cluster, or the average cluster size. The time evolution of the P value is determined for each system to quantify their propensity to form aggregates. This

measurement is performed over a period of 100 ns using four independent particle trajectories for each system.

## RESULTS AND DISCUSSIONS

### (A) Effect of Bilayer Curvature

Earlier studies<sup>9</sup> suggest that protein inclusions in vesicles have a higher tendency to aggregate than those in membranes. Curved membranes (i.e., vesicles) induce curvature-mediated interactions between proteins, and we are interested in quantifying the effect of these interactions on protein aggregation. On the other hand, depletion-induced interactions<sup>45</sup> between proteins drive protein aggregation in a membrane. Both interactions are a function of the distance of separation between the proteins. We build protein-membrane and protein-vesicle assemblies by varying the helical separation between the protein inclusions. In order to quantify the effect of bilayer curvature, we compare the results of protein aggregation in a vesicle to that in a membrane. A protocol described in another work<sup>19</sup> makes a comparison between protein aggregation in vesicles and membranes. Our models have been built along the same lines, setting the membrane as reference to test for higher degree of aggregation in vesicles.

Our results suggest that the curvature of vesicles is responsible for a higher degree of protein aggregation under a specific set of conditions. To isolate the effect of curvature on the self-organization of proteins, all measurements are performed at the same concentration (namely, medium concentration), helical separations ranging from 1.2 nm to 2 nm and a zero hydrophobic mismatch of the proteins. Protein to lipid ratio in membranes and vesicles are equal. We ensure that the surface density of proteins in membranes (figure SI.3) and vesicles (figure 2B) are approximately similar.

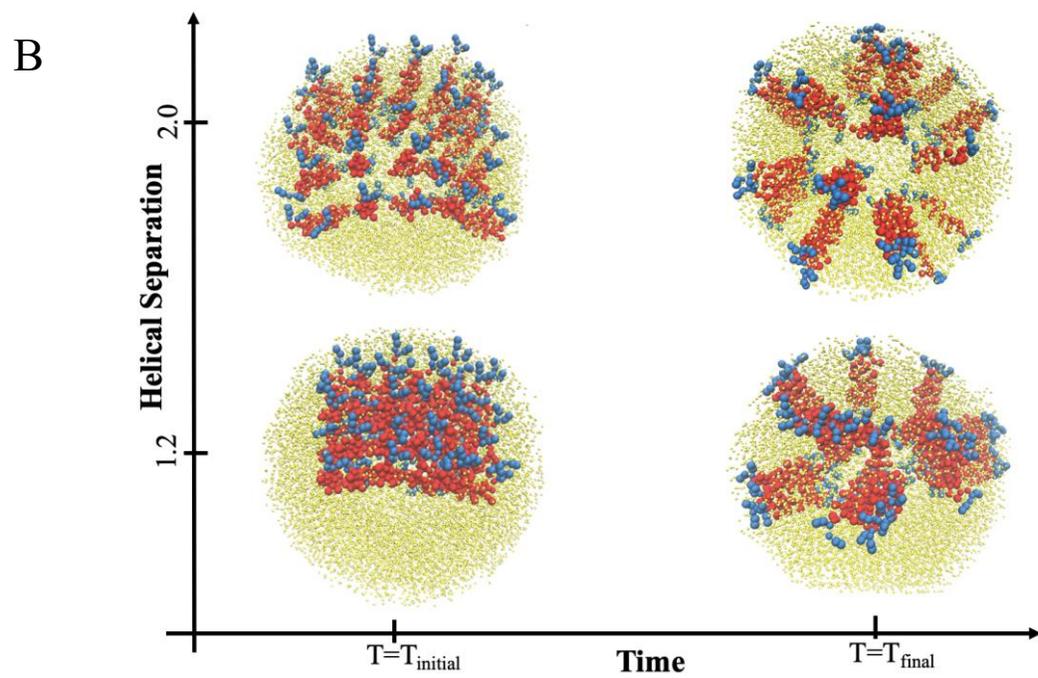
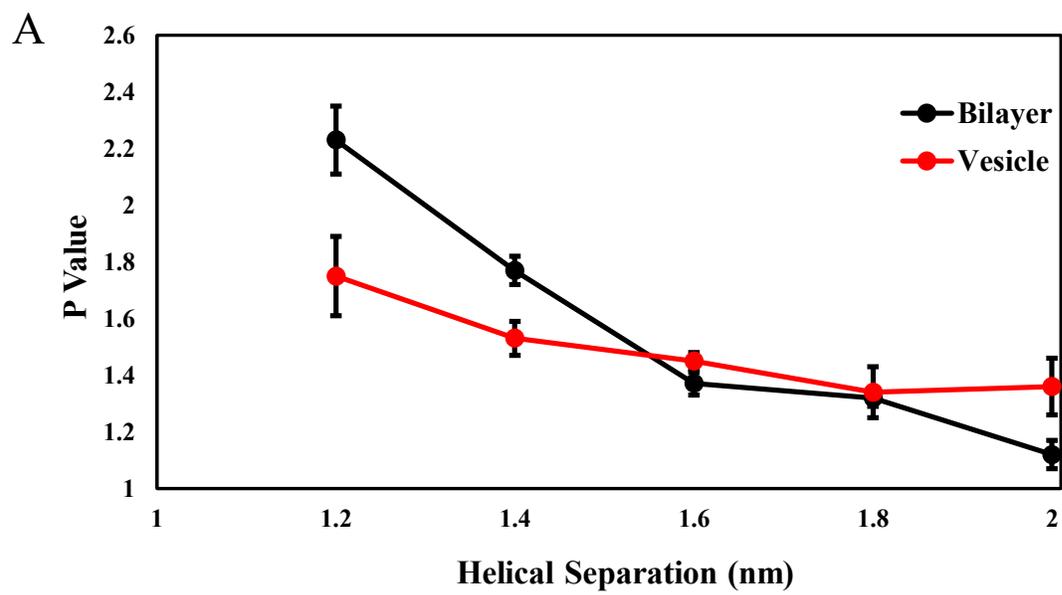


Figure 2: (A) P value calculations as a function of the helical separation of protein inclusions. The red and black lines represent P value measurements for a protein-vesicle and protein-bilayer system, respectively. (B) Protein-vesicle systems for the smallest and largest helical separations (i.e., 1.2 nm and 2.0 nm respectively). The initial unaggregated configurations at 0 ns ( $T_{\text{initial}}$ ) and the final aggregated configuration ( $T_{\text{final}}$ ) are illustrated individually on the vertical axis.

The P value measurements for membranes and vesicles is summarized in Figure 2.A. KALP proteins arranged with a helical separation of 1.2 nm in a membrane are observed to form large clusters ( $P = 2.3$ ) with the proteins organized into elongated linear chains. However, the average aggregate size decreases with helical separations up to a distance of 1.6 nm. As per theory, each protein is considered to have a certain lipid deformation zone <sup>46,47</sup> (namely, regions of the bilayer that are perturbed due to the presence of a protein). Aggregation would be observed when the lipid deformation zones of two or more proteins overlap. We surmise that two neighboring proteins are unable to sense each other at helical separations beyond 1.6 nm as the corresponding deformation zones of the proteins may not significantly overlap. Experimental studies <sup>48</sup> have shown that a high concentration of proteins in bilayers which are placed at distances beyond the calculated deformation length from each other do not aggregate. The decrease in the aggregation of the proteins could be potentially consistent with the reduction in the magnitude of the depletion-induced forces between the proteins. <sup>45</sup>

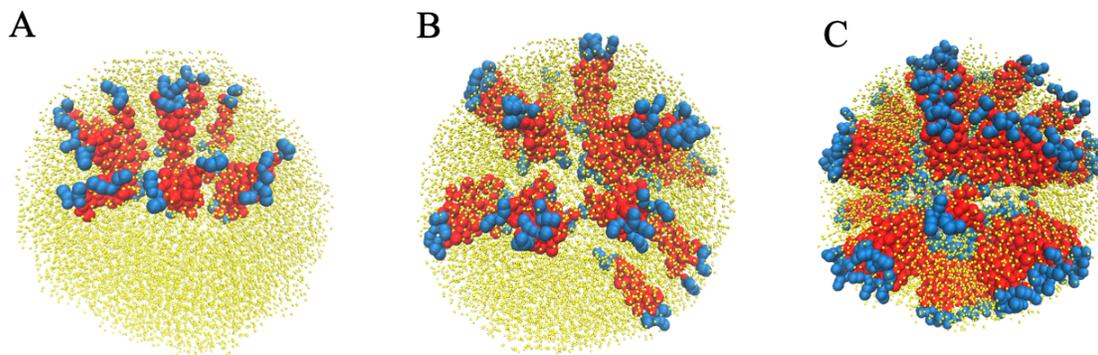
The sizes of aggregates in vesicles are observed to decrease with helical separation. In comparison to membranes, aggregates in vesicles are found to be smaller for shorter helical separations and larger for longer helical distances. A vesicle maintains its spherical morphology via minimization of its interfacial energy by increasing its bending energy. This process gives rise

to an effective repulsive barrier to the aggregation of KALP proteins. However, in the case of proteins placed beyond the interaction cutoff distance, curvature-induced forces<sup>8,9,49</sup> play a critical role in the aggregation of KALPs. The curvature of a vesicle induces a force which drives the displacement of KALPs, bringing the KALPs within interaction distance and causing them to form clusters.<sup>8</sup> Experimental studies<sup>50</sup> on membrane curvature-sensing proteins demonstrated that curvature was the sole determinant of the aggregation of proteins, even if the proteins had a repulsive barrier towards aggregation. From these observations, we infer that curvature-induced displacements are sufficiently high to displace proteins from their original locations and drive them into the deformation zone of the neighboring protein. Therefore, beyond a helical separation of 1.6 nm, the curvature-driven displacements of proteins results in higher P values for vesicles as compared to those for membranes (see Figure 2.B). A consequence of this observation is that vesicles can host moderate-sized clusters of proteins which were initially separated from each other by higher helical separations. An additional observation is that curvature-induced forces do not permit vesicles to host very large clusters, similar to the ones observed in membranes with protein inclusions at initial helical separations of 1.2 nm. This limitation arises as vesicles minimize their interfacial energy by bending to preserve their spherical morphology. Hence, the clustering behavior of proteins in bilayers is coupled to their initial helical separation and the curvature of the bilayer. The aggregation of the proteins in a bilayer are driven by depletion-induced forces below 1.6 nm and curvature-induced forces above 1.6 nm. The stability of these clusters will be discussed in a later section.

The impact of concentration, helical separation and hydrophobic mismatch of the proteins on their self-organization has been investigated using a vesicle to incorporate the effect of

membrane curvature. In most of the investigations, we have used a helical separation of 1.6 nm to reduce the competition between depletion-induced and curvature-mediated forces.

### (B) Impact of Protein Concentration



**D**

Model	P Value
15 KALP Vesicle (P/L = 0.012)	1.44 ± 0.05
35 KALP Vesicle (P/L = 0.028)	1.51 ± 0.02
150 KALP Vesicle (P/L = 0.176)	2.28 ± 0.04

Figure 3: Images of (A) 15 KALPs, (B) 35 KALPs and (C) 150 KALPs in a DPPC vesicle (namely, low, medium and high concentration respectively). The transmembrane proteins are represented by red and blue beads, whereas the vesicle matrix is represented by yellow beads of a smaller size. All protein-lipid systems have been simulated for 1 $\mu$ s and the aggregated states have been shown. (D) Protein/Lipid ratios and P values for low, medium and high concentrations.

We examine the impact of the concentration of proteins on their self-organization in a vesicle bilayer. To isolate the effect of concentration, a helical separation of 1.6 nm is selected as

it is shown to be impervious to the interplay between depletion-induced and curvature-mediated forces. In addition, KALP 23 is chosen for the study due to its zero hydrophobic mismatch with the bilayer. The following concentrations are used for the study: 15 KALPs (low concentration), 35 KALPs (medium concentration) and 150 KALPs (high concentration).

Using the protocol detailed in the SI, three systems are generated with KALP 23 proteins organized in vesicles at different concentrations. Measurements of the P value for each system (see Figure 3) show the low and medium concentration systems to form aggregates with similar size distributions. However, the high concentration system forms aggregates with sizes 30% larger than those corresponding to the low and medium concentrations. Our results suggest that concentrations of protein inclusions beyond a certain threshold results in a higher degree of protein aggregation.

Proteins incur an elastic strain in the bilayer which generates a deformation zone of a specific radius around each cluster. This deformation is stored as elastic bending energy in the bilayer<sup>15</sup> which induces a long range force.<sup>49</sup> Earlier studies have used elastic models<sup>10</sup> to demonstrate that membrane-mediated forces play an important role in the degree of protein aggregation. We hypothesize that a larger total elastic deformation of the bilayer results in a higher contribution to lipid-mediated protein clustering,<sup>17</sup> which becomes pronounced with increasing concentrations of proteins. Hence, the proteins minimize the elastic deformations in the bilayer by forming aggregates which cause the sharing or overlap of individual protein deformation zones.<sup>46,51</sup> However, the size of the aggregates will be constrained due to the preservation of a spherical morphology by the vesicle. Determination of the threshold concentration at which the aggregation of proteins increases significantly is beyond the scope of the present study.

### **(C) Role of Helical Separation**

To understand the effect of helical separation on the aggregation of proteins, vesicles encompassing a high concentration of KALP 23 proteins (zero hydrophobic mismatch) are considered. The helical separation between the proteins is varied from 1.2 nm to 2 nm, in intervals of 0.2 nm. Helical separations of 1.2 nm and 2 nm would respectively capture the effects of depletion-induced and curvature-induced forces. Therefore, the objective of this measurement is to develop a relationship between the helical separation and the underlying driving force for aggregation. The intermediate helical separations are selected to provide insight into the behavior arising from the interplay of the two forces.

The aggregation of the proteins is characterized by the size or number of proteins in a cluster. These measurements require sampling of four independent particle trajectories over a duration of  $1\mu\text{s}$  at a frequency of 10 ns. The distribution of cluster sizes (see Figure 4.B) is measured by normalizing by the total number of clusters. These calculations exclude rare cases such as clusters with more than 10 proteins and unaggregated proteins. Also, we do not base any of our conclusions on average trends using clusters beyond the size of seven. This is due to the extremely low likelihood of such big clusters as it is energetically unfavorable for the lipid vesicle.

Aggregates encompassing 2 to 7 proteins (figure 4.B) represent a specific trend at a helical separation of 1.6 nm: proteins separated by a distance lower than 1.6 nm show the opposite trend as compared to the ones placed at helical separations greater than 1.6 nm. Thus, proteins separated by 1.6 nm seem to have an equal impact of driving forces that are responsible for aggregation at lower (1.2 nm and 1.4 nm) and greater (1.8 nm and 2.0 nm) helical separations.

Systems with proteins initially placed with respect to each other at distances close to the interaction cutoff distance (i.e., 1.2 nm and 1.4 nm) form small clusters with their immediate neighbors. This behavior could be caused by the depletion-induced attraction<sup>45</sup> between two neighboring proteins which slows down the diffusion of proteins fragmenting from and coalescing with aggregates (small black error bars in figure 4.B are indicative of this behavior). The slower diffusion rates would prevent the clusters from growing as a significant population of the proteins are participating in stable clusters encompassing two or three proteins. The curvature of the vesicles is not expected to play a dominant role on the aggregation in these systems as the helical separations are low.

Systems with proteins initially separated from each other by distances well beyond  $r_c$  (namely, 1.8 nm and 2 nm) form larger clusters encompassing four to seven proteins, with a higher relative probability. In these systems, the proteins had a lower tendency to aggregate with their immediate neighbors. As discussed earlier, the curvature-induced forces drive the proteins to translate to regions favoring aggregation, thereby forming large clusters. The higher relative probability of the large cluster sizes for initial helical separations of 1.8 nm and 2 nm highlights the importance of curvature-induced forces<sup>9</sup> on aggregation of proteins. Also, the fluctuations in the sizes of the clusters demonstrates the proteins to fragment from or coalesce with the clusters (large red error bars in figure 4.B are indicative of this behavior).

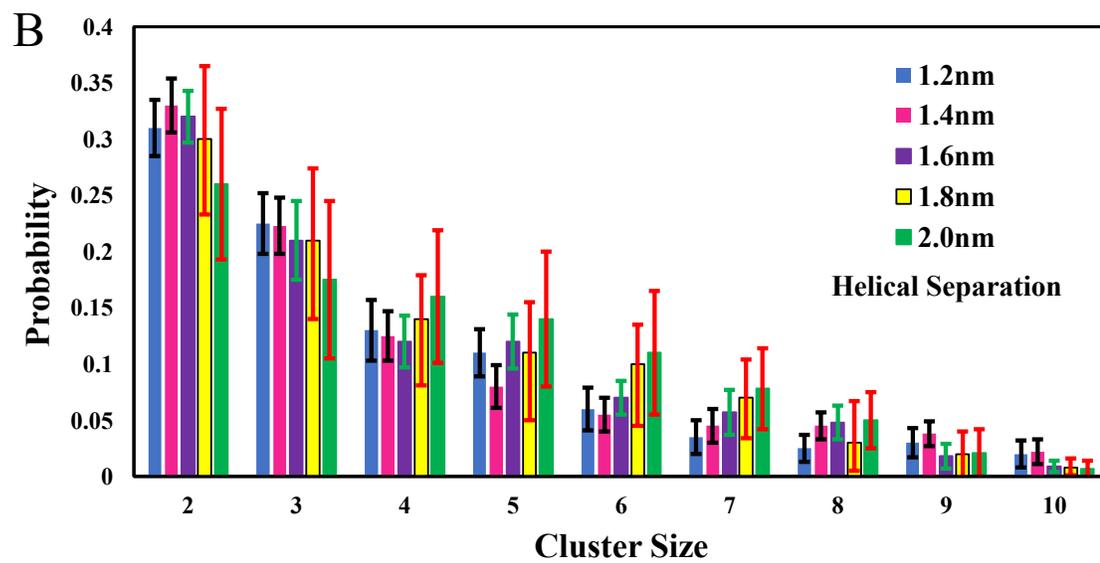
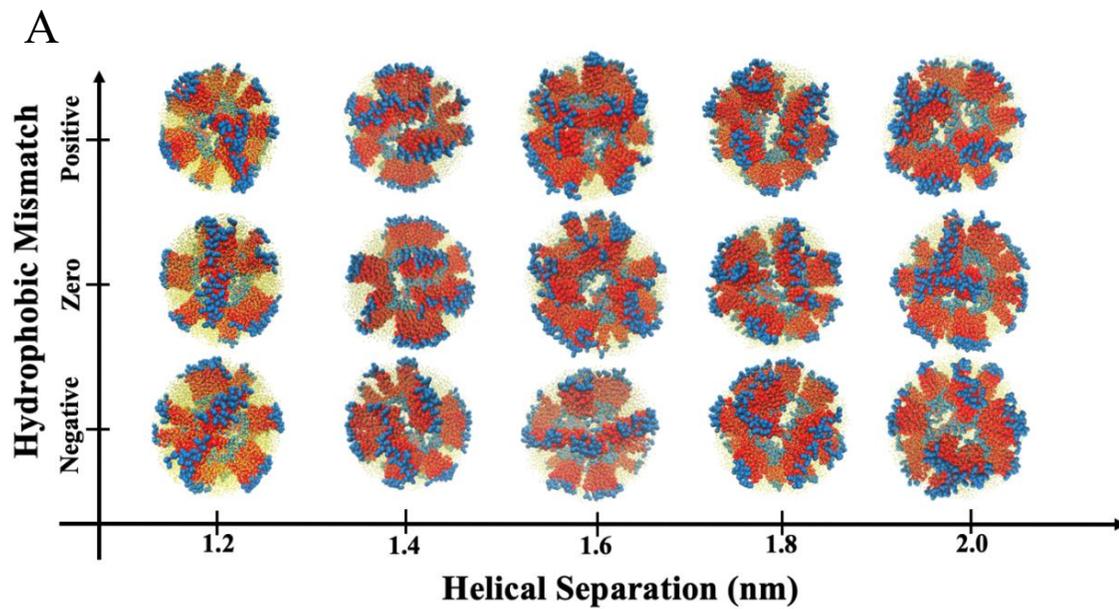


Figure 4: (A) 150 KALP proteins with negative, zero and positive mismatch in DPPC vesicles forming clusters that encompass 2 to 14 proteins. (B) Probability of finding clusters with varying initial helical separations. The black, green and red standard deviation bars correspond to systems with helical separation of 1.2 nm and 1.4 nm, 1.6 nm and 1.8 and 2.0 nm, respectively.

#### **(D) Role of Hydrophobic Mismatch**

The hydrophobic mismatch of a protein with the bilayer imposes a strain on the neighboring membrane lipids which gives rise to an elastic deformation of the bilayer in the vicinity of the protein. Studies have proposed the existence of an annular zone surrounding the protein which encompasses the deformed region.<sup>52</sup> The deformation of the annular zone is noted to be a function of the protein size.<sup>15</sup> The elastic deformation of the bilayer along with curvature-mediated forces will impact the aggregation of proteins for different hydrophobic mismatch conditions. Hence, the effect of hydrophobic mismatch on the organization of proteins in vesicles can be determined by the lifetime of clusters.

These measurements are performed under conditions intended to isolate the effect of hydrophobic mismatch. We record trajectories spanning 100 ns after a 4  $\mu$ s simulation run for all systems. The initial helical separation between the homogeneously distributed proteins in the vesicle is set at 1.6 nm as this distance was not affected by depletion-induced or curvature-induced forces. A high concentration of proteins was used for these measurements. The initial vesicle curvature and the number of lipids is kept the same for all mismatch conditions.

Using measurements of the mean residence time (MRT),<sup>19</sup> the lifetime of clusters for different hydrophobic mismatch conditions is determined (see Figure 5). Four independent particle trajectories, each spanning 1  $\mu$ s and sampled at a frequency of 10 ns, are used for each hydrophobic

mismatch condition. At each time step, the clusters in the system are characterized for their individual sizes and spatial location. If a match between the cluster size at the same location is found for two consecutive timesteps, the cluster is attributed to have a mean residence time of 10 ns. The cluster sizes and positions are tracked over 100 timesteps to evaluate the total MRT. Figure 5 summarizes the MRT distribution for clusters with different sizes. Due to similarities in clusters encompassing two and three proteins for all hydrophobic mismatch conditions, results for clusters constituting four or more proteins are reported. We report the weighted mean residence time (WMRT) using the size of the clusters as weighting factors.

The negative mismatch condition yields a uniform distribution of MRTs for aggregates encompassing four to nine KALPs, with the average WMRT of 144.08 ns. The results show the existence of large clusters encompassing nine KALP 19 proteins over a duration of 400 ns, indicating significantly low diffusivity of the proteins. In addition, clusters encompassing six to nine proteins were observed to exchange KALPs via disintegration, diffusion and coalescence. The WMRT for the positive hydrophobic mismatch condition (94.38 ns) is found to be less than corresponding measurements for the negative mismatch condition. This observation indicates the relative instability of the aggregates composed of KALP 27.

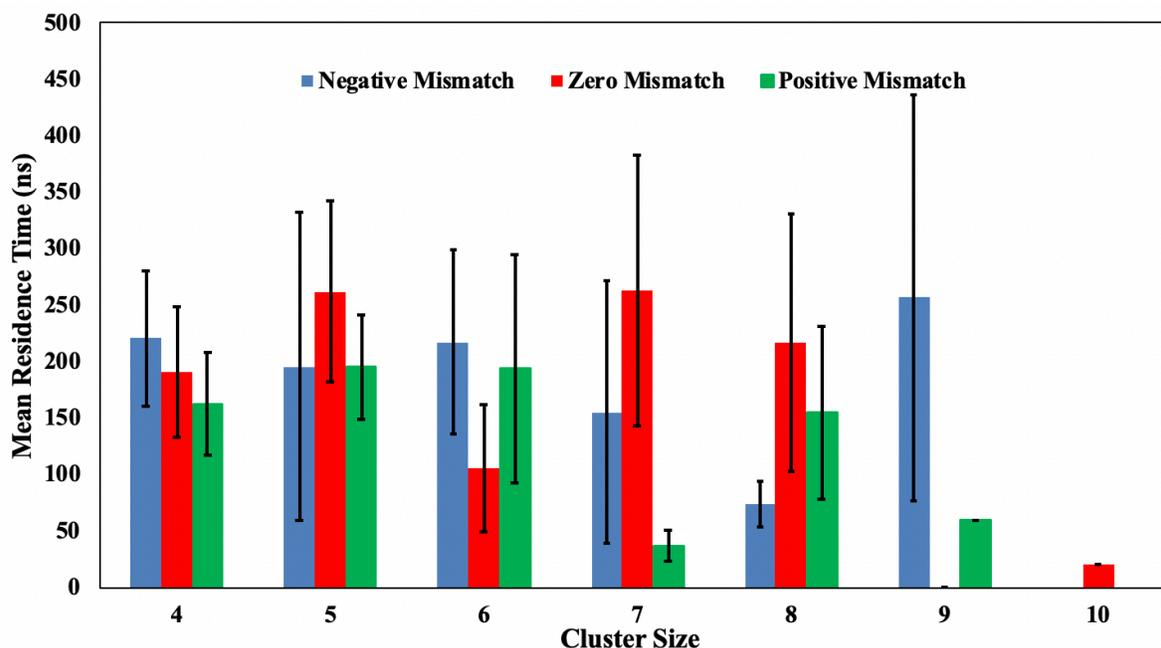


Figure 5: Mean Residence Time (MRT) calculations for each hydrophobic mismatch condition (i.e., negative, zero and positive). Larger clusters have larger standard deviations due to higher number of proteins.

These results demonstrate the negative mismatch condition to be conducive for the formation and stability of larger clusters compared to the positive mismatch condition. This finding could be explained by the following hypothesis; given the small tilt angle of negative mismatch proteins in a bilayer, <sup>15,41</sup> hydrophobic mismatch is compensated by elastic deformation of the bilayer. (The protein shields its hydrophobic components from the solvent by elastically deforming the surrounding bilayer region (namely, the lipid annulus)). The elastic deformations induce the aggregation of the KALP 19 proteins to form large, stable clusters while the vesicle is able to preserve its morphology. In contrast, positive hydrophobic mismatch is compensated by a high degree of protein tilt. <sup>15</sup> This means that proteins with positive hydrophobic mismatch will tilt significantly more to shield their hydrophobic components from the solvent. Hence, the elastic

deformations incurred by these proteins on the surrounding bilayer region will be less than those corresponding to KALP 19. Furthermore, the elastic deformations induced by KALP 27 will induce the aggregation of the proteins to a lesser degree.

## CONCLUSIONS

A fundamental understanding of the self-organization of protein in bilayers could potentially provide insight into the energy pathways in photosynthetic bacterial membranes, thereby guiding the study of bioinspired materials. We examined the aggregation of transmembrane proteins in bilayers as a function of various factors. These factors included the bilayer curvature, protein concentration, helical separation and hydrophobic mismatch. The findings from this study contributes to the fundamental understanding of protein-protein and protein-lipid interactions and highlights the specific factors responsible for the aggregation of proteins.

We find that the impact of membrane curvature on the aggregation of proteins is highly dependent on the initial distance of separation between the proteins. Our results identify a critical distance of 1.6 nm below which the membranes favor the formation of large protein aggregates. Above the critical distance, the curvature-mediated interactions present in a vesicle favors the formation of relatively larger protein clusters.

Investigations on the impact of the concentration of proteins on their aggregation demonstrates a 30 % increase in aggregation upon transitioning from medium to high concentration of KALP proteins in a vesicle. Our results indicate the effects of long range elastic deformations arising from high concentration of proteins induces a homogenous size distribution of clusters encompassing two to ten proteins.

The distinct effects of aggregation driven by depletion-induced forces and curvature-mediated interactions were identified through the variation of the helical spacing using a high concentration of proteins. Our investigation identified 1.6 nm to be the separation distance at which the effect of both the forces were identical. For separation distances below 1.6 nm, we observed the formation of small, stable clusters. The formation of these clusters are driven by the depletion-induced attraction between proteins positioned close to one another. However, proteins separated by a helical separation of 1.8 nm and 2 nm formed relatively large, unstable clusters. The higher distance of separation between the proteins enabled curvature-mediated interactions to displace proteins over a greater distance and form aggregates with a larger number of proteins. The aggregates were observed to exchange proteins frequently via fragmentation and coalescence events.

Our investigations on the effect of hydrophobic mismatch on the self-organization of proteins demonstrated proteins with negative mismatch to favor a higher degree of aggregation. Hence, proteins with negative mismatch were responsible for the formation of larger aggregates which had relatively long life times.

Our results highlight specific conditions that impact the self-organization of proteins in bilayers. Our investigations on protein-protein interactions in a lipid bilayer can guide the design of bioinspired materials where the chemistry of the molecular components can yield materials with structure-function properties mimicking those of photosynthetic bacterial membranes. Furthermore, the results from our investigations can be used by the wider materials chemistry community to understand the organization of inclusions or macromolecules in polymeric matrices. For example, the results reported from this study can be potentially applied to quasi two dimensional materials encompassing inclusions.<sup>53,54</sup> The interaction forces between the inclusions

could be varied via their initial spatial configuration, concentration and mismatch with a quasi-two dimensional material along with the curvature of the material to precisely control the self-organization of the inclusions. This class of materials can be potentially used in a wide range of areas including energy, electronics and biomedicine.

### **Conflicts of Interest**

There are no conflicts to declare

### **ACKNOWLEDGEMENTS**

M.D. would like to acknowledge NSF CAREER award DMR-1654325 and NSF REU award DMR-1659099. Portions of the research presented used computational resources supported by NSF XSEDE (allocation DMR-140125) and Rutgers Discovery Informatics Institute (RDI<sup>2</sup>).

### **REFERENCES**

- 1 K. R. Miller, *Proceedings of the National Academy of Sciences*, 1979, **76**, 6415–6419.
- 2 R. Henderson, *Annual Review of Biophysics and Bioengineering*, 1977, **6**, 87–109.
- 3 J. Baudry, E. Tajkhorshid, F. Molnar, J. Phillips and K. Schulten, *The Journal of Physical Chemistry B*, 2001, **105**, 905–918.
- 4 H. Janovjak, J. Struckmeier, M. Hubain, A. Kedrov, M. Kessler and D. J. Müller, *Structure*, 2004, **12**, 871–879.
- 5 J. Hsin, J. Gumbart, L. G. Trabuco, E. Villa, P. Qian, C. N. Hunter and K. Schulten, *Biophysical Journal*, 2009, **97**, 321–329.

- 6 D. E. Chandler, J. Hsin, C. B. Harrison, J. Gumbart and K. Schulten, *Biophysical Journal*, 2008, **95**, 2822–2836.
- 7 M. D. Winn, G. N. Murshudov and M. Z. Papiz, 2003, pp. 300–321.
- 8 K. S. Kim, J. Neu and G. Oster, *Biophysical Journal*, 1998, **75**, 2274–2291.
- 9 A. H. Bahrami and M. A. Jalali, *The Journal of Chemical Physics*, 2011, **134**, 085106.
- 10 F.-Y. Chen, M.-T. Lee and H. W. Huang, *Biophysical Journal*, 2003, **84**, 3751–3758.
- 11 P. Qian, C. Neil Hunter and P. A. Bullough, *Journal of Molecular Biology*, 2005, **349**, 948–960.
- 12 J. Hsin, J. Strümpfer, M. Şener, P. Qian, C. N. Hunter and K. Schulten, *New Journal of Physics*, 2010, **12**, 085005.
- 13 A. F. Smeijers, K. Pieterse, A. J. Markvoort and P. A. J. Hilbers, *The Journal of Physical Chemistry B*, 2006, **110**, 13614–13623.
- 14 T. K. M. Nyholm, S. Özdirekcan and J. A. Killian, *Biochemistry*, 2007, **46**, 1457–1465.
- 15 M. Venturoli, B. Smit and M. M. Sperotto, *Biophysical Journal*, 2005, **88**, 1778–1798.
- 16 A. J. de Jesus and T. W. Allen, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2013, **1828**, 864–876.
- 17 N. Castillo, L. Monticelli, J. Barnoud and D. P. Tieleman, *Chemistry and Physics of Lipids*, 2013, **169**, 95–105.
- 18 E. Strandberg, S. Özdirekcan, D. T. S. Rijkers, P. C. A. van der Wel, R. E. Koeppe, R. M. J. Liskamp and J. Antoinette Killian, *Biophysical Journal*, 2004, **86**, 3709–3721.
- 19 D. L. Parton, J. W. Klingelhoefer and M. S. P. Sansom, *Biophysical Journal*, 2011, **101**, 691–699.
- 20 S. K. Kandasamy and R. G. Larson, *Biophysical Journal*, 2006, **90**, 2326–2343.

- 21 D. Marsh, *Biophysical Journal*, 2008, **94**, 3996–4013.
- 22 H. I. Ingólfsson, C. A. Lopez, J. J. Uusitalo, D. H. de Jong, S. M. Gopal, X. Periole and S. J. Marrink, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2014, **4**, 225–248.
- 23 Daan Frenkel and Berend Smit, *Understanding Molecular Simulation*, 2002.
- 24 S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *The Journal of Physical Chemistry B*, 2007, **111**, 7812–7824.
- 25 S. Toxvaerd and J. C. Dyre, *The Journal of Chemical Physics*, 2011, **134**, 081102.
- 26 G. Bussi, D. Donadio and M. Parrinello, *The Journal of Chemical Physics*, 2007, **126**, 14101.
- 27 M. Parrinello and A. Rahman, *Journal of Applied Physics*, 1981, **52**, 7182–7190.
- 28 M. Parrinello and A. Rahman, *Physical Review Letters*, 1980, **45**, 1196–1199.
- 29 H. J. C. Berendsen, D. van der Spoel and R. van Drunen, *Computer Physics Communications*, 1995, **91**, 43–56.
- 30 E. Lindahl, B. Hess and D. van der Spoel, *Journal of Molecular Modeling*, 2001, **7**, 306–317.
- 31 S. Plimpton, *Journal of Computational Physics*, 1995, **117**, 1–19.
- 32 T. A. Wassenaar, H. I. Ingólfsson, R. A. Böckmann, D. P. Tieleman and S. J. Marrink, *Journal of Chemical Theory and Computation*, 2015, **11**, 2144–2155.
- 33 C. Hofsaß, E. Lindahl and O. Edholm, *Biophysical Journal*, 2003, **84**, 2192–2206.
- 34 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *The Journal of Chemical Physics*, 1984, **81**, 3684–3690.

- 35 The KALP protein reference structure, <http://cgmartini.nl/index.php/example-applications2/35-downloads/exampleapplications/121-kalpdppc> , (accessed December 2016)
- 36 T. M. Weiss, P. C. A. van der Wel, J. A. Killian, R. E. Koeppe and H. W. Huang, *Biophysical Journal*, 2003, **84**, 379–385.
- 37 T. Baştuğ and S. Kuyucak, *Biophysical Reviews*, 2012, **4**, 271–282.
- 38 P. Buslaev and I. Gushchin, *Scientific Reports*, 2017, **7**, 11476.
- 39 S. J. Marrink and A. E. Mark, *Journal of the American Chemical Society*, 2003, **125**, 15233–15242.
- 40 P.H.Sherrod, Nonlinear Regression Analysis Program (NLREG Version 5.0),Nashville,TN, 2000
- 41 T. Kim and W. Im, *Biophysical Journal*, 2010, **99**, 175–183.
- 42 Y. Yano and K. Matsuzaki, *Biochemistry*, 2006, **45**, 3370–3378.
- 43 S. Sharma, B. N. Kim, P. J. Stansfeld, M. S. P. Sansom and M. Lindau, *PLOS ONE*, 2015, **10**, e0144814.
- 44 E. S. Salnikov and B. Bechinger, *Biophysical Journal*, 2011, **100**, 1473–1480.
- 45 T. Sintès and A. Baumgärtner, *Biophysical Journal*, 1997, **73**, 2251–2259.
- 46 A. G. Lee, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2003, **1612**, 1–40.
- 47 M. v. LeVine, G. Khelashvili, L. Shi, M. Quick, J. A. Javitch and H. Weinstein, *Biochemistry*, 2016, **55**, 850–859.
- 48 M. Schümann, M. Dathe, T. Wieprecht, M. Beyermann and M. Bienert, *Biochemistry*, 1997, **36**, 4345–4351.

- 49 A. V. Botelho, T. Huber, T. P. Sakmar and M. F. Brown, *Biophysical Journal*, 2006, **91**, 4464–4477.
- 50 V. Vogel and M. Sheetz, *Nature Reviews Molecular Cell Biology*, 2006, **7**, 265–275.
- 51 S. Mall, R. Broadbridge, Ram. P. Sharma, J. M. East and A. G. Lee, *Biochemistry*, 2001, **40**, 12379–12386.
- 52 C. Aponte-Santamaria, R. Briones, A. D. Schenk, T. Walz and B. L. de Groot, *Proceedings of the National Academy of Sciences*, 2012, **109**, 9887–9892.
- 53 M. Dutt, M. J. Nayhouse, O. Kuksenok, S. R. Little and A. C. Balazs, *Current Nanoscience*, 2011, **7**, 699–715.
- 54 M. Dutt, O. Kuksenok, M. J. Nayhouse, S. R. Little and A. C. Balazs, *ACS Nano*, 2011, **5**, 4769–4782.