NMR-Based Computational Studies of Membrane Proteins in Explicit Membranes

By

Xi Cheng

Submitted to the graduate degree program in the Center for Computational Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Date Defended: 04/07/2015

The Dissertation Committee for Xi Cheng

certifies that this is the approved version of the following dissertation:

NMR-Based Computational Studies of Membrane Proteins in Explicit Membranes

_____

Wonpil Im, Chairperson

Date approved: 04/07/2015

Abstract

Since nuclear magnetic resonance (NMR) spectroscopy data, including solution NMR from micelles and solid-state NMR from bilayers, provide valuable structural and dynamics information of membrane proteins, they are commonly used as restraints in structural determination methods for membrane proteins. However, most of these methods determine the protein structures by fitting the single-confer model into all available NMR restraints regardless of the explicit environmental effects that are determinant in the structures of membrane proteins. To develop a reliable protocol for obtaining optimal structures of membrane proteins in their native-like environments, various NMR properties were applied in the refinement approaches using explicit molecular dynamics (MD) simulations in this research.

First, solution NMR NOE based-distance measurements were used as restraints in MD simulations to refine an activating immunoreceptor complex in explicit environments. Compared to the structure determined in vacuum, the resulting structures from the explicit restrained simulations yields a more favorable and realistic side-chain arrangement of a key Asp residue, which is highly consistent with mutagenesis studies on such residue.

Incorporating solid-state NMR and solution NMR, MD simulations were performed in the explicit bilayers to refine the structure of membrane-bound Pf1 coat protein. Since solid-state NMR is sparse in its N-terminal periplasmic helix, the protein structure was determined by combining solid-state NMR and solution NMR. Benefiting from the sophisticated energy function and the explicit environments in MD, the orientation of Pf1's periplasmic helix can be identified in simulations restrained by solid-state NMR alone. In the simulations restrained with

both solid-state NMR and solution NMR, physically irrelevant structures were frequently observed, suggesting there are conflicts between the restraints from different sample types (e.g., bilayers and micelles).

As NMR data are ensemble-averaged measures, the solid-state NMR restrained explicit ensemble dynamics (ED) simulations of fd coat protein were performed in different ensemble sizes and compared to the unrestrained MD simulations. As the ensemble size increases, the violations of resulting structures from experimental NMR data decrease, while the structural variations increase to be comparable to the unrestrained MD simulations, indicating the efficacy of restrained ED in refining structures and extracting dynamics.

To investigate the influence of different environments on the structures of membrane proteins, in this research, MD simulations were performed in bilayers and micelles, respectively. Since building a preassembled protein/micelle complex for MD simulation is challenging and requires considerable experience with simulation software, a web-based graphical interface *Micelle Builder* in CHARMM-GUI (http://www.charmm-gui.org/input/micelle) was developed to support users to build micelle systems in a automatic and simplified process. Using this interface, Pf1 coat protein was preassembled in a protein/micelle model and simulated in explicit environment. Compared to previous simulations of Pf1 coat protein in bilayers, different protein conformations were observed in these simulations due to the distinct behavior and geometry of micelles.

Acknowledgements

I would never have been able to finish my dissertation without the guidance of my committee members, help from friends, and support from my family.

First, I would like to express my great gratitude to my advisor, Professor Wonpil Im, for his excellent guidance, motivation, immense knowledge and providing me with an excellent atmosphere for doing research. He is a productive and energetic expert in the field of computational biophysics and always able to give insightful discussions and suggestions about research. I deeply appreciate the opportunities and support that he gave me to learn and apply new skills over the years that I have spent in his lab.

I would like to thank Dr. Ilya Vakser, Dr. John Karanicolas, Dr. Eric Deeds, Dr. Joanna Slusky and Dr. Mario Rivera for reading my thesis and providing insightful comments. During my graduate years, I had chances to discuss various research topics with them and learned a lot.

This work presented here would not be possible without the help of the past and current Im lab members and the collaborators from other institutions. I am very grateful to Dr. Huan Rui, Dr. Sunhwan Jo, Dr. Emilia L. Wu, Dr. Yifei Qi and Dr. Jong Cheol Jeong. They have helped me countless times on the technical issues I encountered in my research. It would be remiss not to mention Dr. Francesca Marassi, Dr. Jeffrey Klauda and Dr. Matthew Call for sharing important experimental data, giving valuable comments and engaging in discussions for the work presented here. During my graduate years in Lawrence, I have met many people who have helped my Ph.D. study directly or indirectly. My thanks go to Dr. Kyuil Lee, Dr. Soohyung Park, Dr. Huisun Lee, Dr. Jaejin Ka, Dr. Dhilon S. Patel, Dr. Zhaowen Duan, Andrew H. Beaven, Jumin

Lee, Hongjing Ma, Seonghoon Kim, Kevin Song, Nathan R. Kern and Eder M. Davila-Contreras.

At last, I want to say thank you to my family for the support they provided me through my entire life. Thanks to my beloved parents, Baoqing Cheng and Mingxia Pan, for their unconditional love and care. And, special thanks go to my “Dr. Right”--Zhuxi Chen.

Table of Contents

Abstract iii

Acknowledgements v

List of Tables x

List of Figures xi

1. Introduction 1

1.1 Biological roles of membrane proteins 1

1.2 Nuclear magnetic resonance (NMR) spectroscopy in protein structure determination 2

1.3 Computational studies of membrane proteins 6

1.4 Outline of dissertation 10

2. Solution NMR based structural refinement of an activating immunoreceptor complex in explicit membranes 11

2.1 Introduction 11

2.2 Materials and methods 13

2.3 Results and discussion 15

2.4 Conclusions 27

3. Solid-state NMR and solution NMR restrained molecular simulations of Pf1 coat protein in explicit bilayers 28

3.1 Introduction 28

3.2 Materials and methods 32

3.2.1 Restraint potentials for CSA, DC, and RDC 32

3.2.2 Restrained MD simulations in explicit bilayers 34

3.3 Results and discussion 37
3.3.1 Validation of restrained simulations 37
3.3.2 Depth of membrane insertion 39
3.3.3 Transmembrane helix orientation, dynamics, and lipid interactions 42
3.3.4 Periplasmic helix orientation, dynamics, and interactions 44
3.4 Conclusions 49
4. Solid-state NMR based ensemble dynamics simulations of fd coat protein in explicit membranes 51
4.1 Introduction 51
4.2 Materials and methods 54
4.2.1 Solid-state NMR ensemble restraint potentials 54
4.2.2 Simulations in explicit lipid bilayers 55
4.3 Results and discussion 59
4.3.1 Validation of solid-state NMR restrained ensemble dynamics 59
4.3.2 Influence of solid-state NMR restraints on helix orientations 63
4.3.3 Flexibility of transmembrane helix conformations in simulations 65
4.3.4 Protein-lipid interactions in explicit membranes 67
4.4 Conclusions 72
5. CHARMM-GUI micelle builder for pure/mixed micelle and protein/micelle complex systems 75
5.1 Introduction 75
5.2 Materials and methods 78
5.2.1 Protein/micelle complex building process in CHARMM-GUI Micelle Builder 78

5.2.2 Detergent-only homogenous micelle systems 86
5.2.3 Pf1 protein/micelle system 88
5.3 Results and discussion 90
5.3.1 Micelle size and shape 90
5.3.2 Micelle structure 94
5.3.3 Distribution and mobility of detergents in protein/micelle systems 96
5.3.4 Protein-detergents interactions and its effect on protein structure 101
5.4 Conclusions 104
Reference 105
Appendix 120
A list of publications 120

List of Tables

Table 2.1 System information for DAP12-NKG2C 14

Table 2.2 The orientation of DAP12-NKG2C TM helix complex in terms of helix-helix distance (D) and crossing angle (Ω). 16

Table 3.1 Restrained MD refinement statistics 33

Table 3.2 Transmembrane helix orientation calculated from last 10 ns of restrained MD 43

Table 3.3 Frequency of interaction between transmembrane helix polar or charged side chains and phospholipid headgroups observed in the last 10 ns of restrained MD 43

Table 4.1 System information for ED and MD simulations. 57

Table 4.2 Structural statistics 57

Table 5.1 Force constants for positional harmonic restraints on each equilibration step. 85

Table 5.2 System information for micelle simulations. 87

Table 5.3 Geometrical parameters of detergent micelles. 88

List of Figures

Figure 1.1 Structures of membrane proteins determined by NMR spectropy. 2

Figure 1.2 Schematics representations of internuclear vectors used to calculate instantaneous CSA and DC. 4

Figure 2.1 Simulation systems of DAP12-NKG2C complex in micelles and bilayers. 14

Figure 2.2 The average number of violated NOE distance restraints in the different systems with a cutoff value of 0.5 Å.. 15

Figure 2.3 Top and side views of the DAP12-NKG2C complex showing the interactions between the key interfacial residues. 17

Figure 2.4 Relative rotation angles of Asp16 in DAP12. 18

Figure 2.5 The distance between the polar groups of DAP12 Asp16s and NKG2C Lys52 in the different systems. 19

Figure 2.6 Water accessible surface area of DAP12 Asp16 in the different systems. 19

Figure 2.7 The interaction network at Asp16-Thr20-Lys52. 21

Figure 2.8 Top and side views of the DAP12-NKG2C complex showing the water molecules near the key interfacial residues. 22

Figure 2.9 The relationship between the number of water molecules near NKG2C Lys52 and DAP12-1 Asp16 and the salt bridge formation between them. 23

Figure 2.10 Density profiles of system components along *Z*-axis. 25

Figure 2.11 Cylinder radial distribution of system components around the DAP12-NKG2C complex principal axis. 26

Figure 3.1 Initial structural models, M1 and M2, used for MD simulations of Pf1 coat protein in explicit lipid bilayers. 36

Figure 3.2 Correlation between experimental and back-calculated CSA, DC, and RDC set 1 and 2 for the Pf1 coat protein. 38

Figure 3.3 Depth of transmembrane helix membrane insertion. 40

Figure 3.4 Interactions between residues and solvent. 41

Figure 3.5 The orientation and depth of membrane insertion of the periplasmic helix. 47

Figure 3.6 Snapshots showing the interactions between the periplasmic helix of Pf1 coat protein and lipid molecules. 48

Figure 4.1 Ensemble dynamics simulation system of fd coat protein. 55

Figure 4.2 Comparison between the experimental and back-calculated CSA and DC for the fd coat protein. 61

Figure 4.3 Validation of fd coat protein structure ensemble. 62

Figure 4.4 Distributions of the tilt angles of the periplasmic helix and the transmembrane helix with respect to the membrane normal. 64

Figure 4.5 Distributions of the angles between the principal axes of two transmembrane helical segments Y21-A35 and G38-F45. 66

Figure 4.6 Interactions between each residue and solvent components. 68

Figure 4.7 The average *Z* coordinates of lipid phosphorus atoms in the lower leaflet in ED32 and one snapshot from ED32 showing the interactions between lipids and K40. 69

Figure 4.8 The distributions of the (hinge) angles between the principal axes of the periplasmic and transmembrane helices and the (crossing) angles between their principal axes projected on the *XY* plane. 71

Figure 5.1 Chemical structures of DHPC, DPC, TPC and SDS. 77

Figure 5.2 Overview of building process of protein/micelle complex simulation systems in CHARMM-GUI *Micelle Builder*. 81

Figure 5.3 Cross-sectional area profile of Pf1 coat protein (PDB:2KSJ) along the Z-axis, generated by *Micelle Builder*. 82

Figure 5.4 Schematic representation of a Pf1 coat protein/micelle system. 83

Figure 5.5 Representative structures of DHPC, DPC, SDS and TPC micelles. 87

Figure 5.6 Initial and equilibrated structures of Pf1 coat protein in the DHPC micelles. 89

Figure 5.7 Schematic representation of the micelle radius. 91

Figure 5.8 Micelle radius of gyration of all replicas as a function of time. 92

Figure 5.9 Density profiles for the specific components from the micelle center of mass. 95

Figure 5.10 Protein/micelle complex structures. 96

Figure 5.11 The number of Pf1-contact detergents as a function of time. 98

Figure 5.12 Two snapshots in Pf1-DHPC100 taken at 87 ns and 90 ns. 99

Figure 5.13 Interactions between Pf1 residues and various components. 100

Figure 5.14 Snapshots showing the interactions between Pf1 coat protein periplasmic helix and detergent molecules. 101

Figure 5.15 Distribution of angles between the transmembrane and periplasmic helices of Pf1 protein in the DHPC micelle and DOPC/DOPG bilayer. 103

Figure 5.16 Snapshots of Pf1 protein in bilayer and micelle systems. 103

1. Introduction

1.1 Biological roles of membrane proteins

Membrane proteins are abundant and biologically important. 20-30% of all genes in most genomes are estimated to encode membrane proteins (1). In various cellular processes, membrane proteins play distinct and crucial roles (2). Membrane receptor proteins transduce signals to trigger changes in the function of the cell (3), membrane transport proteins move molecules and ions across the membranes (4), and membrane adhesion proteins promote cell recognitions (5). Due to their critical status in cell functions, membrane proteins make extremely interesting targets for drugs. In fact, membrane receptors and transporters are targets of over 50% of all modern medicinal drugs (6).

Determination of structures and dynamics of membrane proteins is essential for understanding of their functions. The interactions between membrane proteins and surrounding lipids are important determinants for protein structures (7). However, these lipid bilayers hampered the expression and purification of the transmembrane domains of membrane proteins in the classical experimental techniques, such as X-ray diffraction (8). Nowadays, advanced experimental and computational methods have been developed to overcome limitations in the determination of membrane protein structure.

1.2 Nuclear magnetic resonance (NMR) spectroscopy in protein structure determination

NMR spectroscopy is a powerful approach providing information about the structure and dynamics of membrane proteins dissolved in detergent micelles or reconstituted into lipid bilayers (Figure 1.1) (9). Using solution NMR methods, it is feasible to determine the structures of small membrane proteins in micelles. However, solid-state NMR can determine the structures of larger membrane proteins in their definitive environments of lipid bilayers (10).

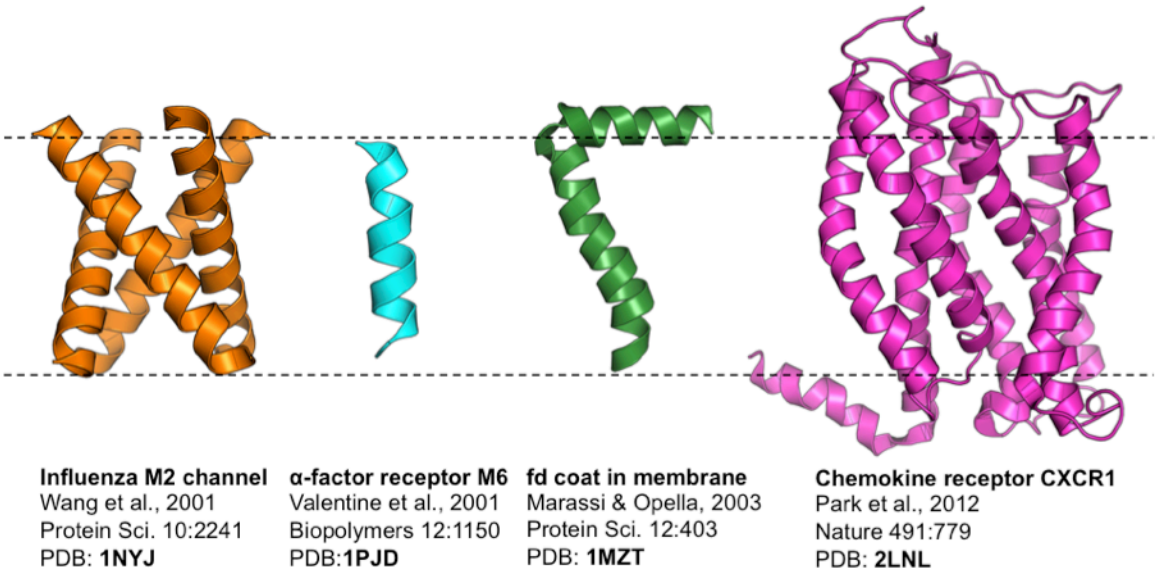

Figure 1.1 Structures of membrane proteins determined by NMR spectropy.

In solution NMR spectroscopy, detergent-solubilized membrane proteins tumble freely on a nanosecond timescale, allowing the collection and assignment of NMR spectra. In this process, the nuclear Overhauser effect (NOE) is commonly observed showing certain atoms are in close proximity to each other. The inter-atomic distances derived from the observed NOE can help confirm the three dimensional protein structures (11). In addition, in solution, the partial

alignment of proteins is induced either by their own magnetic susceptibility anisotropy (12, 13) or by external alignment media (14, 15). This partial molecular alignment leads to an incomplete averaging of anisotropic magnetic interactions such as the dipolar coupling and the chemical shift anisotropy (16). In this scenario, the measurable residual dipolar coupling (RDC) can provide information about the orientation of each internuclear vector with respect to the static magnetic field and therefore determine protein structures (17). The following equation shows the RDC (v_{RDC}) between a pair of nuclei P and Q with a spin of 1/2.

$$v_{RDC} = \frac{2}{3}\frac{K_{RDC}}{r_{PQ}^3}\langle\frac{3}{2}cos^2\varphi - \frac{1}{2}\rangle \tag{1-1}$$

where $r_{PQ} = \left|\mathbf{r}_P - \mathbf{r}_Q\right|$ and φ is the angle between the internuclear vector and the magnetic field. $K_{RDC} = -S\mu_0\gamma_P\gamma_Q h/8\pi^3$, where μ_0 is the magnetic permeability of vacuum, γ_P and γ_Q are gyromagnetic ratios of P and Q, h is Plank's constant, and S is the generalized order parameter describing the internal motion of the internuclear vector.

Solid-state NMR spectroscopy does not rely on rapid molecular reorientation, so it is suited for membrane proteins immobilized in phospholipid bilayers (9). When an immobile sample is oriented to the direction of the applied magnetic field, the resulting NMR spectra such as chemical shift anisotropy (CSA) and dipolar coupling (DC) are characterized. These observed resonances are dependent on the orientation of their molecular sites relative to the static magnetic field (18). As shown in Equation 1-2 and Equation 1-3, CSA (σ_{CSA}) relies on the peptide-plane orientation with respect to the direction of the magnetic field ($\mathbf{B_0}$) which is aligned to the *Z*-axis; DC (v_{DC}) relies on the angle (θ) between the nuclei pair vector (r_{NH}) and the direction of the magnetic field ($\mathbf{B_0}$) (19).

$$\sigma_{CSA} = \sigma_{11}e_{1,z}^2 + \sigma_{22}e_{2,z}^2 + \sigma_{33}e_{3,z}^2 \quad (1\text{-}2)$$

$$v_{DC} = \frac{K_{DC}}{2}(3cos^2\theta - 1) \quad (1\text{-}3)$$

where σ_{nn} and $e_{n,z}$ (n = 1, 2, 3) are the instantaneous magnitude and unit vector Z-component of chemical shift tensors (see Figure 1.2 for e_n orientations). $K_{DC} = (\gamma_N\gamma_H h\mu_0)/(8\pi^3 r_{NH}^3)$, μ_0 is the magnetic permeability of vacuum, γ_N and γ_H are gyromagnetic ratios, h is Plank's constant.

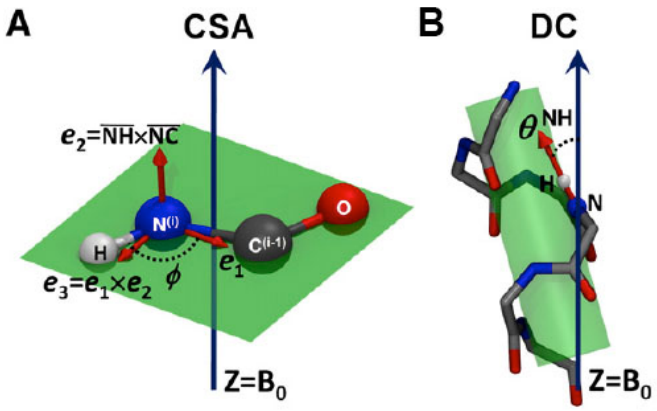

Figure 1.2 Schematics representations of internuclear vectors used to calculate instantaneous (A) CSA and (B) DC. For CSA, e_2 is generally defined by the cross production of r_{NC} and r_{NH}. e_1 is defined by a rotation angle ϕ from r_{NH} on the peptide plane defined by N, C and H atoms. $e_3 = e_1 \times e_2$.

These NMR data depending on the orientations of molecular sites can be applied in structural determination methods based on orientational restraints (20). In these approaches, a protein orientational search is carried out to minimize the RMSD between the experimental data (χ^{exp}) and the calculated values (χ^{calc});

$$RMSD_{\chi} = \sqrt{\frac{1}{N_{\chi}} \sum_{i=1}^{N_{\chi}} \left(\chi_i^{exp} - \chi_i^{calc}\right)^2} \tag{1-4}$$

One conventional and popular determination method is the semi-static fitting approach. In this approach, the experimentally derived orientations are determined by searching a particular set of orientations of an ideal, rigid secondary structure, i.e., a helix, to minimize the deviations between the experimental and calculated NMR data. Previous studies showed that the accessible orientations determined by the semi-static fitting approach are very narrow or delta function-like. The most probable reason is this method does not consider any motional averaging and therefore neglects the ensemble average (21).

1.3 Computational studies of membrane proteins

Molecular dynamics (MD) simulation is a powerful tool for complementing experimental studies and providing explanation for experimental observations at an atomic level (22). By setting up simulation systems of membrane proteins under the same conditions as experiment, the MD simulations can mimic the biological situations having such membrane proteins and capture their motions at time scales from picoseconds to microseconds. At every simulation time step, the future positions of all the atoms in the system are predicted following Newton's classical equation of motion:

$$\boldsymbol{F}_\alpha = m_\alpha \bar{\boldsymbol{r}}_\alpha = -\frac{\partial U(\{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_N\})}{\partial \boldsymbol{r}_\alpha} \tag{1-5}$$

where $\boldsymbol{F}_\alpha$ is the force exerted on atom α with mass m_α and its position r_α. The potential energy function of the system $U(\{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_N\})$ involves the positions of all the atoms in the system. In a typical biological simulating system, the potential energy function is expressed in the form of a summation over bonded energies of bond, angle and torsion, and non-bonded energies of van der Waals and electrostatics interactions (23):

$$U(\{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_N\}) = \sum_{bonds,i} \frac{1}{2} k_{b,i}(b_i - b_i^0)^2 + \sum_{angles,i} \frac{1}{2} k_{\theta,i}(\theta_i - \theta_i^0)^2$$

$$+ \sum_{torsion,i} k_{\emptyset,i}[1 + cos(n_i \emptyset_i - \delta_i)] \tag{1-6}$$

$$+ \sum_i \sum_{j,j \neq i} \left\{ \varepsilon_{ij} \left[\left(\frac{r_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^{min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon r_{ij}} \right\}$$

Compared to the semi-static fitting approach, which does not have motional information, MD simulations can provide not only structural but also dynamic information of membrane proteins. As a result, simulated annealing with restrained MD is increasingly used in the NMR-based structure calculations. However, in most of them, all non-bonded interactions are represented by a simple repulsive term with no contributions from van der Waals, electrostatic or solvation energy. A more sophisticated NMR based refinement approach was developed to include all energy terms in Equation 1-6 (19, 24). As shown in Equation 1-7 and Equation 1-8, NMR data is converted into harmonic restraint potentials ($U_{\chi^i}(\{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_N\})$) added to the total potential energy applied in the MD simulations of the membrane proteins:

$$U(\{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_N\})_{total} = U(\{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_N\}) + \sum_i U_{\chi^i}(\{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_N\}) \tag{1-7}$$

$$U_{\chi^i}(\{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_N\}) = \sum_j^N k_{\chi^i} \left(\chi_j^{i,exp} - \chi_j^{i,calc} \right)^2 \tag{1-8}$$

where χ^i is a NMR measure, *N* is the number of χ^i, and k_{χ^i} is the force constant for χ^i.

Using solid-state NMR data as restraints in implicit MD simulations, this structural calculation method can produce plenty of structures with few violations from the experimental NMR data (19). However, since NMR data are time- and ensemble- averaged measures, a strong restraint potential fitting one particular structural conformer to the averaged NMR data might corrupt the atomic configuration of the protein in an undesirable way (25).

One alternative strategy is the NMR-restrained ensemble dynamics (ED) simulation. Its scheme is performing parallel MD simulations of *N* replicas (N_{rep}) in the presence of NMR restraint potentials that enforces the ensemble averaged NMR properties ($\langle\chi_i^{\mathrm{calc}}\rangle_{ens}$) toward known experimental values (χ_i^{exp}) at each simulation time step:

$$\langle\chi_i^{\mathrm{calc}}\rangle_{\mathrm{ens}} = \sum_{j=1}^{N_{\mathrm{REP}}} \chi_i^{\mathrm{calc}(j)}/N_{\mathrm{REP}} \tag{1-9}$$

$$U_\chi(\{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_N\}) = N_{\mathrm{REP}} \sum_{i=1}^{N_\chi} k_\chi \left(\langle\chi_i^{\mathrm{calc}}\rangle_{\mathrm{ens}} - \chi_i^{\mathrm{exp}}\right)^2 \tag{1-10}$$

where χ is a NMR measure, N_χ is the number of χ, and k_χ is the force constant for χ. With respect to the refinement of transmembrane helix structure, the implicit restrained ED showed good agreement with experimental observables including motional averaging and a great orientational variability (26-28).

Most of these structural refinement approaches use implicit membrane models to reduce computational expenses (26-29). In these implicit models, a low-dielectric slab is defined to mimic the hydrophobic core of the membrane bilayer (30, 31). This static, low-dielectric slab can capture the general effect of biological membrane, such as hydrophobic mismatch (between the length of a protein's hydrophobic transmembrane domain and the thickness of the bilayer hydrophobic core) (32-35). Nonetheless, the undeformable nature of the hydrophobic slab is one of its most significant drawbacks. The biological membrane bilayer is fluid and the membrane protein function is in part regulated by the changes in lipid bilayer thickness and intrinsic lipid curvature (36). Particularly, in the case of a hydrophobic mismatch, the bilayer adaptation involving local changes in lipid bilayer thickness (also known as local membrane thickening or thinning) may change the orientation of the membrane protein (37-43). Therefore, it is of particular interests to perform the structural refinement of membrane proteins in an explicit environment presenting protein-lipid interactions at an atomics level.

1.4 Outline of dissertation

This dissertation mainly focuses on the NMR-based refinements and structural analyses of membrane proteins using MD simulations in explicit membranes. In Chapter 2, solution NMR data were applied to the explicit MD simulations of an activating immunoreceptor complex to address the ambiguities of side-chain conformations and produce optimal structures consistent with the experimental observations. Then, both solution NMR from micelles and solid-state NMR from bilayers were used in the structure refinement of the Pf1 coat protein in Chapter 3. In Chapter 4, the solid-state NMR restrained explicit ED simulations of the fd coat protein were compared to the unrestrained molecular dynamic simulations. All protein/micelle complex models described in these NMR computational studies can be built with *Micelle Builder* in CHARMM-GUI (http://www.charmm-gui.org/input/micelle) in an automatic and simple process. In Chapter 5, this web-based graphical interface was introduced and tested. The different structures of membrane proteins in the micelles and the bilayers are also discussed.

2. Solution NMR based structural refinement of an activating immunoreceptor complex in explicit membranes

2.1 Introduction

NMR observables, such as NOE-based distance, CSA, and various DC measurements, are increasingly being used to characterize membrane protein structures (44, 45). However, membrane proteins are challenging subjects for NMR and result in a relatively small number of such measurements, which can create ambiguities in determining critical side chain-side chain interactions. Additionally, most membrane protein structure calculations do not consider several unique features of the membrane environment that may affect the determined structures, such as the low degree of hydration and associated electrostatic interactions, and the spatial constraints enforced by bilayer geometry. Therefore, the resulting structures, even with few violations of NOE-based distances and/or other observables, can present side-chain conformations that may not reflect the most energetically favorable arrangements.

Recently, Call et al. (46) determined the NOE-based solution NMR structure (PDB:2L35) of a micelle-embedded transmembrane (TM) hetero-trimeric complex DAP12-NKG2C, representing the membrane-embedded portions of the natural killer cell-activating receptor complex DAP12-NKG2C/CD94. DNAX-activation protein 12 (DAP12) is a homodimer containing an immunoreceptor tyrosine-based activation motif in its cytoplasmic domain and non-covalently associated with natural killer group 2C (NKG2C). NKG2C forms a heterodimer with the C-type lectin CD94 and recognizes the human nonclassical MHC class I molecule HLA-E, delivering activating signals via the DAP12 immunoreceptor tyrosine-based activation

motifs (47). Mutagenesis studies demonstrated that five polar residues, including one Asp and one Thr in each DAP12 TM helix and one Lys in the NKG2C TM helix, mediate the key TM contacts between DAP12 and NKG2C (46, 48). The NMR structure provided the first structural insight into the TM contacts within an assembled immunoreceptor complex. Nonetheless, this complex structure shows a puzzling aspect in that one of Asp residues faces the hydrophobic core, which may not be energetically favorable. To explore whether these side-chain conformations are optimal in membrane environments, we have performed a refinement of the DAP12-NKG2C structure using NOE-based restrained MD simulations in both explicit micelles and bilayers.

2.2 Materials and methods

The representative DAP12-NKG2C-micelle and DAP12-NKG2C-bilayer systems are shown in Figure 2.1 (see Table 2.1 for detailed system information). The average NMR structure of PDB:2L35 was used as a starting structure. For simplicity, the DAP12 with a short linker to NKG2C is named DAP12-1 in this work; the linker was introduced in the NMR study to produce a covalently stabilized three-TM complex. For the micelle simulations, 13 sodium dodecyl sulfate (SDS) and 130 FOS-Choline 14 (TPC) molecules (1:10 ratio) were radially distributed around the protein surface to mimic the NMR experimental conditions (46). For the bilayer simulations, DAP12-NKG2C was inserted into a bilayer of 129 dimyristoylphosphatidylcholine (DMPC) molecules using CHARMM-GUI Membrane Builder (49, 50). Each system was replicated and assigned with different initial velocities to generate three independent simulation systems. All calculations were performed in NPT (constant particle number, pressure, and temperature) ensembles (51) at 303.15 K using CHARMM (52) with the CHARMM all-atom protein force field (53) including the dihedral cross-term correction (54), CHARMM36 lipid force field (55), and a modified TIP3P water model (56). A time step of 2 fs was used with the SHAKE algorithm (57). Each initial system was equilibrated for 300 ps with the nonbonded and dynamics options in the Membrane Builder input; the van der Waals interactions were smoothly switched off at 10 – 12 Å by a force-switching function (58), and the electrostatic interactions were calculated using the particle mesh Ewald method (59) with a mesh size of ~1 Å for fast Fourier transformation, κ = 0.34 $Å^{-1}$, and a sixth-order B-spline interpolation. After equilibration, a total of 10-ns production run was performed for each micelle system and 40 ns for each bilayer system under available NOE-based distance restraint potentials.

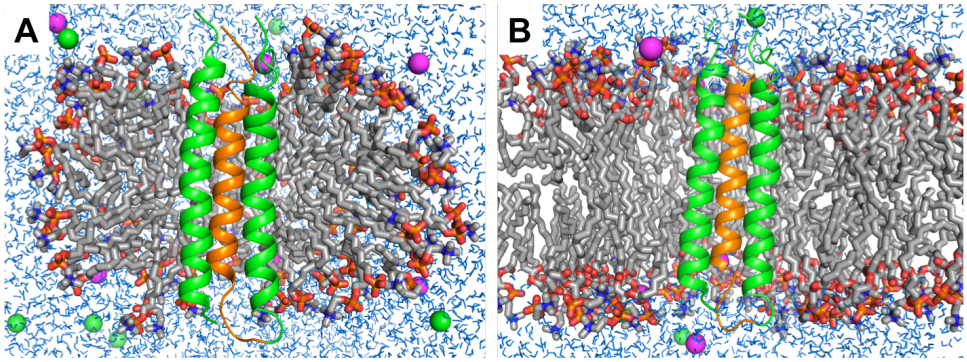

Figure 2.1 Simulation systems of DAP12-NKG2C complex in (A) micelles and (B) bilayers. Protein is shown in cartoon. DAP12 is green, NKG2C is orange. Detergent and lipid molecules are shown in sticks, ions in spheres, and water molecules in lines.

Table 2.1 System information for DAP12-NKG2C

System	# of water molecules	# of detergents/lipids	# of ions	# of atoms	Box size (Å^3)
Micelle	16,476	13 SDS, 130 TPC	55 K^+, 44 Cl^-	60,230	83.8×83.8×83.8
Bilayer	7,411	129 DMPC	19 K^+, 18 Cl^-	38,937	65.6×65.6×87.3

2.3 Results and discussion

The average distance violation is a direct measurement to check whether the MD-refined structures satisfy the NMR observables. The total number of restraints is 238. Given a cutoff value of 0.5 Å, 15 structures with the least violations were selected for each system, as in the conventional NMR structure determination. The average numbers of violated restraints are <1 for all these structures in the different systems, which are comparable to the 15 structures in PDB:2L35 (Figure 2.2). The root mean-squared deviations (RMSD) of the TM helix backbone atoms from the average NMR structure are 1.0 ± 0.1 Å in both micelle and bilayer systems. There are no significant differences between PDB:2L35 and MD-refined structures in terms of helix-helix distance and crossing angles (see Table 2.2). This result indicates that the MD-refined complex structures well satisfy the NMR observables and their overall structures are similar to the PDB:2L25 structures.

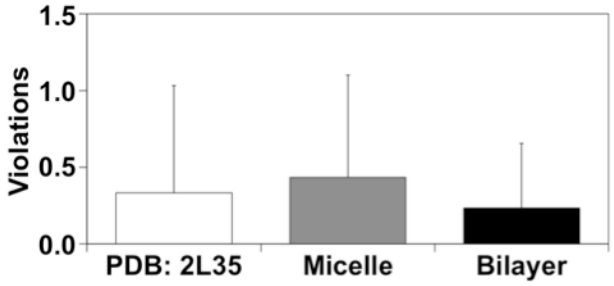

Figure 2.2 The average number of violated NOE distance restraints in the different systems with a cutoff value of 0.5 Å. 15 structures with the least violations were selected for the restrained MD simulations.

Table 2.2 The orientation of DAP12-NKG2C TM helix complex in terms of helix-helix distance (D) and crossing angle (Ω).

Structural element	PDB:2L35		Micelle		Bilayer	
	D (Å)	Ω (°)	D (Å)	Ω (°)	D (Å)	Ω (°)
DAP12-1 and NKG2C	9.6±0.4	3.6±1.8	9.9±0.5	3.9±1.8	9.9±0.3	4.0±2.2
DAP12-2 and NKG2C	11.0±0.4	10.9±2.8	10.6±0.6	11.8±2.5	10.4±0.7	10.7±2.2
DAP12-1 and DAP12-2	9.2±0.4	12.5±1.7	9.4±0.6	10.8±2.8	9.7±0.4	11.3±0.3

Compared to PDB:2L35, however, a different side-chain conformation of DAP12-1 Asp16 (numbering based on PDB:2L35) is observed in the MD-refined structures (Figure 2.3). This Asp16 orients toward the complex interface in the MD-refined structures, but faces the membrane hydrophobic core in PDB:2L35 (Figure 2.3 and Figure 2.4). Because there is no distance restraint for the Asp16 side chain, its orientation can be varied, depending on how interactions and environments are treated. Interestingly, the side chain of DAP12-2 Leu19 is also reoriented to shield this Asp16 from the hydrophobic core (Figure 2.3 *B2* and *B3*).

The Asp16 rotation is related to Asp16s' electrostatic interactions with NKG2C Lys52 (Figure 2.3 *A2* and *A3*). The distances between DAP12 Asp16s and NKG2C Lys52 indicate that both Asp16s can form stable salt bridges with Lys52 in the MD-refined structures (Figure 2.5), which provides a plausible explanation for the strict requirement for two DAP12 and only one NKG2C TM domains in the complex (46). The opposite face of DAP12 in the refined structures is essentially devoid of an appropriate NKG2C TM association site due to the rotation (Figure

2.6), which attractively explains the experimental observation that a second NKG2C TM domain cannot join in the assembly (46).

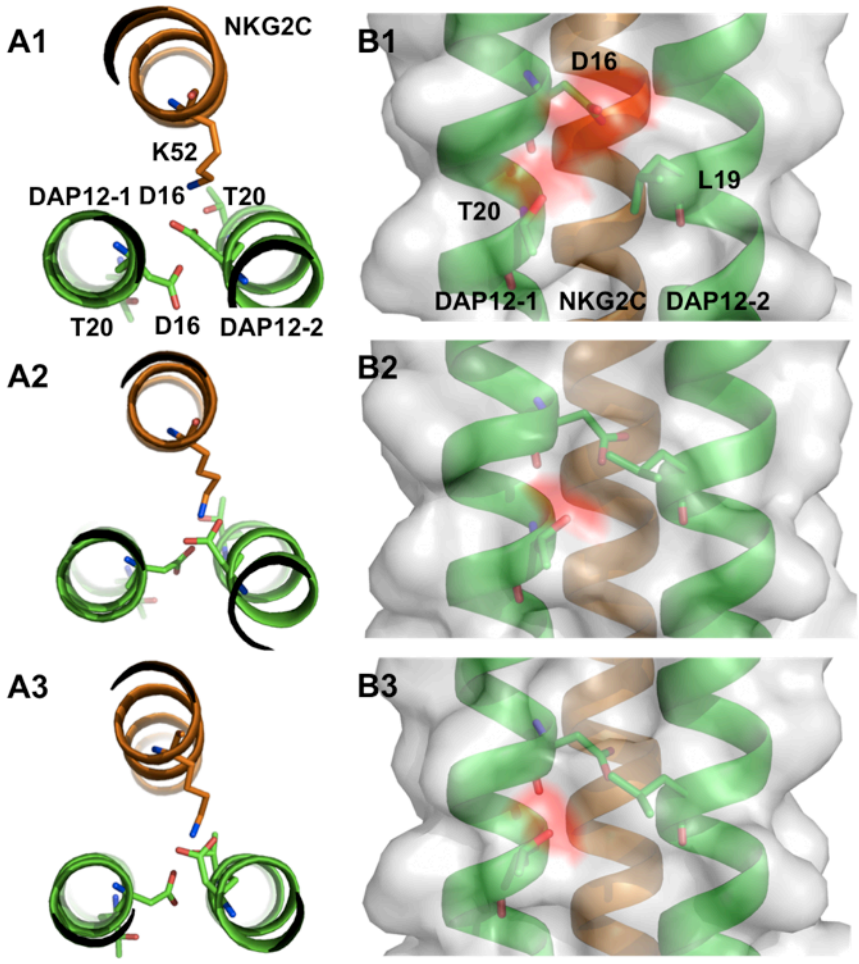

Figure 2.3 Top and side views of the DAP12-NKG2C complex showing the interactions between the key interfacial residues in (A1, B1) PDB:2L35, (A2, B2) micelle systems, and (A3, B3) bilayer systems, respectively. All other side chains are omitted for clarity. (Red and gray) Negatively charged and neutral protein surfaces, respectively.

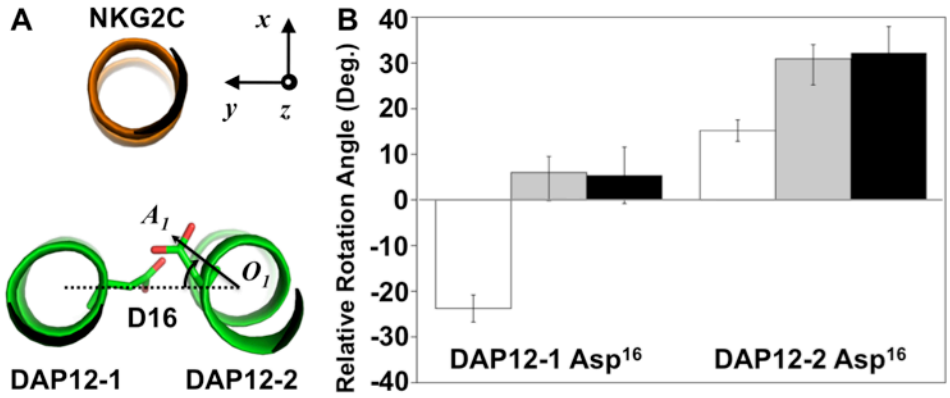

Figure 2.4 Relative rotation angles of Asp16 in DAP12. (A) The top view of one representative structure from the micelle systems. The rotation angles were calculated by aligning DAP12's principal plane to the *YZ* plane, keeping the NKG2C above the principal plane, and measuring the angles of the geometrical center of DAP12-2 Asp16 side-chain oxygen atoms (*A1*) from the DAP12-2 (*O1*) relative to the principal plane (similarly for DAP12-1 Asp16). If the angle is negative, this Asp16 is below the DAP12 principal plane and outside the interface. (B) Average DAP12 Asp16's relative rotation angle in different systems (PDB:2L35: white, micelle: gray, bilayer: black).

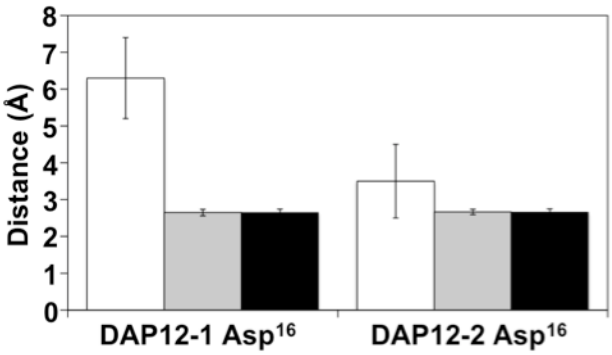

Figure 2.5 The distance between the polar groups of DAP12 Asp16s and NKG2C Lys52 in the different systems (PDB:2L35: white, micelle: gray, bilayer: black). The distance is between the geometrical center of one DAP12 Asp16 side-chain oxygens and the NKG2C Lys52 side-chain nitrogen.

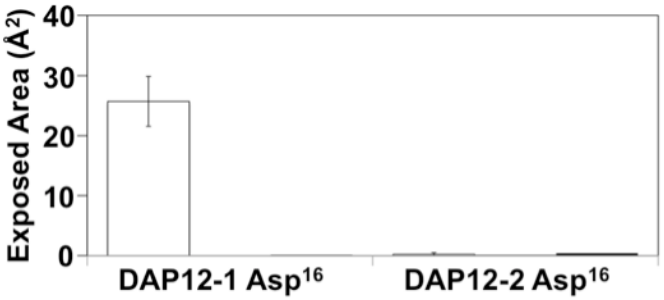

Figure 2.6 Water accessible surface area of DAP12 Asp16 in the different systems (PDB:2L35: white, micelle: gray, bilayer: black).

In addition, DAP12-2 Thr20 in the MD-refined structures shows higher probabilities of hydrogen bond (H-bond) formations with DAP12-1 Asp16 and with NKG2C Lys52 than in PDB:2L35 (Figure 2.7). DAP12-1 Thr20 also has a chance to form intra-helical H-bond with DAP12-1 Asp16. These Thr20-associated intra- and inter-helical H-bonds can keep the DAP12 dimer compact and further stabilizes the TM physical contact (Figure 2.3). Therefore, Thr20s in the refined structures show direct contributions to the complex stability, which is supported by the mutagenesis study in which the substitution of DAP12 Thr20 to Ala leads to serious defects in the complex assembly (46).

The Asp ionization states and the presence of water near the polar residues influence their electrostatic interactions and can yield different side-chain conformations. In the refinement, both Asp16s were not protonated. However, two additional normal MD simulations with protonated DAP12-1 Asp16 (atom types OD1 or OD2) showed that the Asp rotated from the hydrophobic core and formed stable interactions with NKG2C Lys52 (data not shown). Although this result does not provide an answer to the Asp16 ionization state in the DAP12-NKG2C complex, it indicates that the critical polar interactions in this refinement do not depend on the Asp16 ionization state. Although a small number (<5) of water molecules were observed near the key polar residues in some micelle or bilayer systems (Figure 2.8), the interaction pattern of the interfacial residues remained identical in those systems, illustrating that the presence of such water does not interfere with the key residue interactions (Figure 2.9).

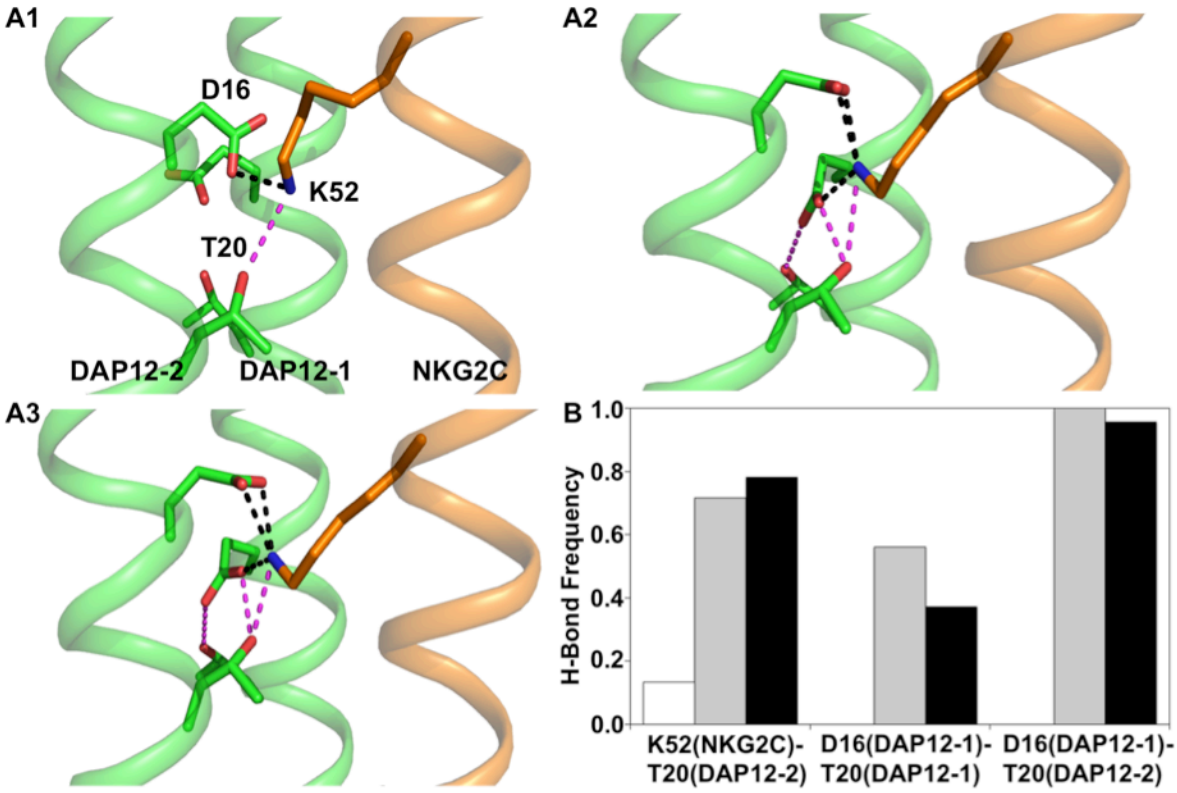

Figure 2.7 The interaction network at Asp16-Thr20-Lys52 in (A1) PDB:2L35, (A2) the micelle system, and (A3) the bilayer system, respectively. The H-bond is defined by $d \leq 2.8$ Å and $120° \leq \theta \leq 180°$, where d is the distance between donor and acceptor atoms, and θ is the H-bond angle. The donors are HZ1/HZ2/HZ3 of NKG2C Lys52 side chain, HG1 of DAP12-1/DAP12-2 Thr20 side chain, and the acceptors are OG1 of DAP12-2 Thr20 side chain, OD1/OD2 of DAP12-1 Asp16 side chain (CHARMM atom types). The black dotted line represents the salt bridges. The magnate dotted line shows the all-possible putative H-bonds. In the refined structures, the NKG2C Lys52 can form two H-bonds maximally. The frequency of forming those H-bonds in the different systems (PDB:2L35: white, micelle: gray, bilayer: black) are shown in (B).

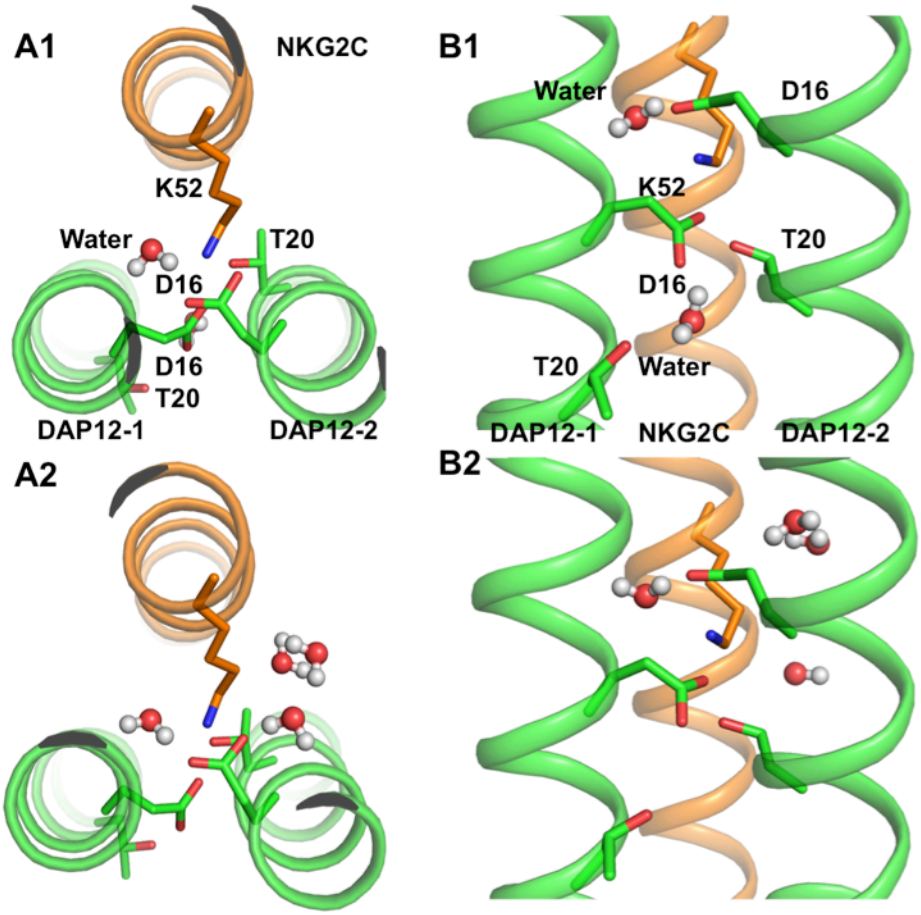

Figure 2.8 Top and side views of the DAP12-NKG2C complex showing the water molecules (spheres) near the key interfacial residues (sticks) in (A1, B1) micelle or (A2, B2) bilayer systems. All other bulky water molecules and the protein side chains are omitted for clarity.

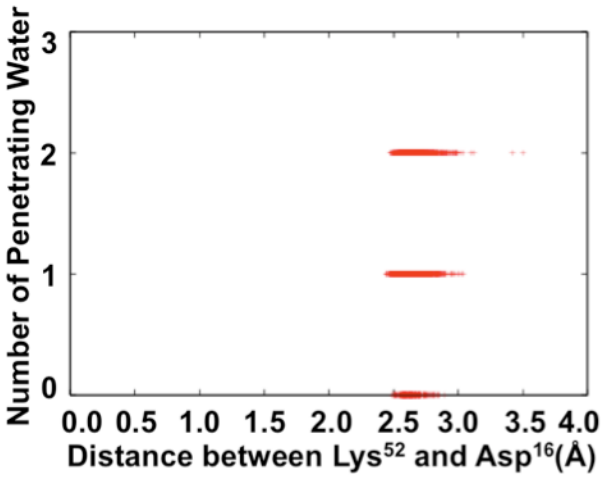

Figure 2.9 The relationship between the number of water molecules near NKG2C Lys52 and DAP12-1 Asp16 and the salt bridge formation between NKG2C Lys52 and DAP12-1 Asp16. The salt bridge formation is indicated in the *X*-axis by the distance between the NKG2C Lys52 side-chain nitrogen and DAP12-1 Asp16 side-chain oxygen atoms from a MD-refined simulation system that had penetrating water molecules during the simulation. This plot illustrates that the salt bridge is maintained regardless of the water molecules near the key polar groups.

Possibly, non-additive effects, not captured in this study, could play a role in such electrostatic interactions (including H-bonds) in membranes, and our results need to be further confirmed by simulations when a polarizable force field becomes available in the future. Nonetheless, it is the advantage of a structure refinement in the explicit membranes (with currently available additive molecular force field) to obtain optimal side-chain conformations (through side chain-side chain and side chain-lipid/detergent interactions) in a more realistic environment. Obviously, the overall distribution of detergent/lipid molecules is different in the micelle and bilayer systems (Figure 2.1). Nevertheless, the local chemical environments surrounding the DAP12-NKG2C complex in both systems are surprisingly similar (Figure 2.10 and Figure 2.11). This result indicates that the protein side chains show similar conformations in the similar local environments. An additional advantage of the bilayer simulation is to refine the TM helix orientation with respect to the bilayer normal. This orientation information, which is not available from micelle simulations, could be important structural information for membrane protein function. In the case of DAP12-NKG2C, the tilt angle of the complex principal axis with respect to the bilayer normal is 9.0 ± 4.8° and the tilt angle of each helix is 10.8 ± 4.6° (DAP12-1), 10.3 ± 5.2° (DAP12-2), and 9.1 ± 4.2° (NKG2C).

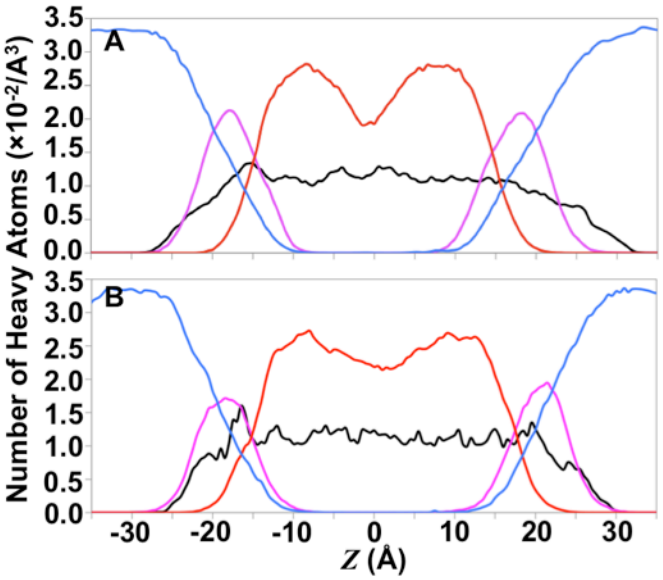

Figure 2.10 Density profiles of system components along *Z*-axis in (A) micelle systems and (B) bilayer systems. In the micelle systems, the principal axis of DAP12-NKG2C was aligned to the *Z*-axis. In the bilayer systems, the *Z*-axis corresponds to the membrane normal. The profiles involve the heavy atoms of the protein (black), detergent/lipid headgroup (magenta), carbon tail (red), and water (blue) within a radius of 20 A° around the *Z*-axis based on Figure 2-11.

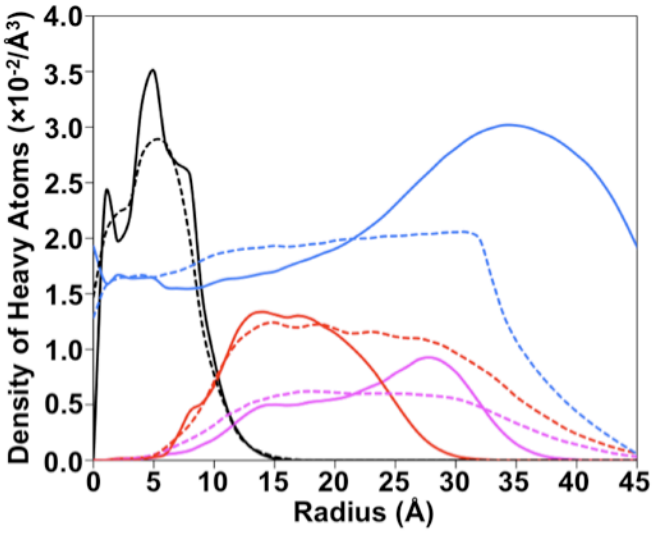

Figure 2.11 Cylinder radial distribution of system components (protein: black, detergent/lipid head group: magnate, detergent/lipid carbon tail: red, water: blue) around the DAP12-NKG2C complex principal axis in the micelle (solid lines) and in the bilayer (dotted lines) systems. In the micelle systems, the principal axis of DAP12-NKG2C was aligned to the *Z*-axis. In the bilayer systems, the *Z*-axis corresponds to the membrane normal. The component distributions are similar in the micelle and in the bilayer within the radius of 20 Å.

2.4 Conclusions

In conclusion, we refined the DAP12-NKG2C structure (PDB:2L35) in explicit micelle and bilayer membranes using the NOE-based distance restraints. The refined structures are globally similar to PDB:2L35, but show different side-chain orientations/conformations of the five functionally required interfacial residues in the middle of the TM domains. Instead of being exposed to the hydrophobic core, DAP12-1 Asp16 stays in the complex interface and forms a stable salt bridge with NKG2C Lys52 in the refined structures. In addition, the refined DAP12 Thr20s form H-bonds with Asp16 and Lys52, which also enhances the complex's structural stability. These features of side-chain interactions are also consistent with the available mutagenesis data. The refined structures provide novel structural information to understand the key TM contact. In the case of the DAP12-NKG2C TM complex, the detergent molecules provide effective local environments similar to lipid bilayers. Given the considerable challenges in collecting sufficient NMR observables (NOE-based distance, chemical shift, and various dipolar coupling) to define all critical side chain-side chain interactions in membrane protein NMR studies, our study illustrates the efficacy of a structure refinement using restrained MD simulations in explicit micelles and bilayers to provide side-chain orientations in more realistic environments and the protein's orientation relative to bilayers.

3. Solid-state NMR and solution NMR restrained molecular simulations of Pf1 coat protein in explicit bilayers

3.1 Introduction

NMR spectroscopy can provide high-resolution three-dimensional (3D) structures of membrane proteins embedded in phospholipid bilayers or bicelles, or dissolved in detergent micelles (60-62). A significant advantage of this technique is that NMR observables represent time- and ensemble-averaged structural restraints, and therefore provide information about protein dynamics that can be used to study collective motions relevant to protein function (63-66). NMR studies performed in lipid bilayers or bicelles have the most biological relevance because they capture the structures and dynamics of membrane proteins in their native-like membrane environment. Membrane proteins can be reconstituted in phospholipid proteoliposomes, planar bilayers, or bicelles, for solid-state NMR structure determination, as described recently for the influenza virus M2 proton channel (67), the human chemokine receptor CXCR1 (68), and the membrane-anchored electron-carrier protein Cytochrome b5 (69). Furthermore, significant progress has been made in incorporating membrane proteins in phospholipid nanodiscs for solution NMR structural studies, as described recently for the bacterial β-barrel outer membrane protein OmpX (70).

Computational methods for NMR protein structure determination are designed to effectively translate NMR observables into 3D structures with minimal deviations between the calculated and the experimental observables (71). However, with few exceptions (e.g., influenza M2 (67)), most NMR structure calculations do not consider explicit protein-solvent or protein-lipid

interactions, although the latter has profound effects on the structure and function of membrane protein (72, 73). Including the proper environment in NMR structure determination significantly influences calculations of protein structure and dynamics. For example, it has been shown that a solution NMR-restrained molecular dynamics (MD) simulation protocol, using either a generalized Born implicit solvent model or an explicit solvent water model, yields soluble protein structures with more favorable backbone dihedral angles (64). Furthermore, for the solution NMR structure of an integral membrane signaling complex (DAP12-NKG2C) determined in micelles (46), NMR-restrained MD refinement in explicit micelles and bilayers yields a more favorable and realistic arrangement of a key Asp residue, compared to refinement in vacuum (74). Notably, the results of refinement in these more realistic environments are consistent with mutagenesis studies while still satisfying the NMR observables. These improvements suggest that there could be multiple structures that are degenerate to the NMR observables and the proper environment must be taken into account to extract the conformations likely adopted in the native-like environment.

In the DAP12-NKG2C example, the NMR observables consisted solely of the NOE distance restraints. However, recent methods for NMR structure determination of membrane proteins rely significantly on orientation restraints derived from measurements of CSA, DC and RDC signals, as well as on dihedral angle restraints derived from measurements of isotropic chemical shift (CS) frequencies. Such restraints are used increasingly for membrane protein structure determination, where NOE distance measurements are more challenging to obtain. Orientation restraints are particularly valuable in solid-state NMR studies performed in lipid bilayers or bicelles where they provide information about a protein's 3D structure as well as its orientation

with respect to the membrane. Orientation restraints could be equally valuable if measured for samples of proteins in nanodiscs by solution NMR.

In this study, we describe a restrained MD simulation protocol for membrane protein structure calculation and refinement in explicit phospholipid bilayer membranes, which can incorporate a wide range of NMR observables, including CSA, DC, and RDC orientation restraints measured by solid-state NMR and solution NMR. This approach overcomes the limitations associated with NMR structure calculations in the absence of solvent. NMR-restrained MD simulations performed in explicit membranes consider protein-lipid interactions explicitly and thus maximize the information content of experimental NMR restraints measured for membrane-embedded proteins. Furthermore, they can facilitate structure determination and remove ambiguities in cases where only sparse restraints can be measured for side-chain sites. As a first application, we use the membrane-bound form of the major coat protein of bacteriophage Pf1, whose structure was determined recently by combining solid-state NMR and solution NMR orientation and dihedral angle restraints (75). The structure is composed of a N-terminal, amphipathic periplasmic helix (residues 5-15) that aligns parallel to the lipid bilayer surface, a short linker, and a C-terminal transmembrane helix (residues 23-45) that adopts a tilt of ~30° relative to the lipid bilayer normal.

Here, we refine the Pf1 coat protein structure, characterize detailed protein-lipid interactions, and explore its dynamics by performing restrained MD simulations in explicit lipid bilayer membranes using the NMR observables: ^{15}N CSA and ^{1}H-^{15}N DC obtained by solid-state NMR, and ^{1}H-^{15}N RDC obtained by solution NMR. In addition, performing the simulations in explicit membranes allows us to determine the protein's depth of insertion into the lipid bilayer, which

cannot be directly characterized from NMR observables without consideration of explicit protein-lipid interactions. Finally, we consider the validity of pooling structural restraints extracted from different sample types (e.g., bilayers and micelles) for NMR structure determination. The restraints for Pf1 coat protein were measured in two types of samples: lipid bilayers for solid-state NMR and detergent micelles for solution NMR. Highly converged, consistent results obtained by using only the solid-state NMR restraints and excluding the solution NMR RDC restraints from the MD simulations suggest that a hybrid approach, combining restraints from different sample types, is not valid, at least for Pf1 coat protein.

3.2 Materials and methods

3.2.1 Restraint potentials for CSA, DC, and RDC

In this work, we used the solid-state NMR (19) and RDC restraint potentials (17) implemented in CHARMM (52), in which a simple harmonic potential was applied to restrain each NMR property calculated from protein structure at each time step (χ^{calc}) to the experimental value (χ^{exp})

$$U_{\chi} = \sum_{i=1}^{N_{\chi}} k_{\chi} \left(\chi_i^{calc} - \chi_i^{exp}\right)^2 \tag{3-1}$$

where χ is either ^{15}N CSA, ^{1}H-^{15}N DC, or ^{1}H-^{15}N RDC, N_{χ} is the number of target experimental observables, and k_{χ} is the force constant; see (17, 19) for detailed calculation methods for each NMR property. Note that in the case of ^{1}H-^{15}N RDC, the alignment tensor is a variable and the forces from the alignment tensor are explicitly considered in the RDC restraint potential (17, 24).

The total number and types of restraints used in the calculations are listed in Table 3.1. The solution NMR ^{1}H-^{15}N RDC restraints were derived from two types of samples: Pf1 coat protein in micelles weakly aligned with fd bacteriophage; and Pf1 coat protein in micelles weakly aligned with stressed polyacrylamide gels (75). The optimal force constants in Equation 3-1 were empirically determined by testing different sets of values and examining their effects on restraint violations and embedded dynamics. To avoid an over-fitting of the protein structure to the experimental data, cross-validation of restraint potentials with different force constants is commonly used (26, 76, 77). We used a similar approach and determined an optimal set of force

constants for k_{CSA} = 0.01 kcal/(mol • ppm^2), k_{DC} = 0.5 kcal/(mol • kHz^2), and k_{RDC} = 5.0 kcal/(mol • Hz^2).

Table 3.1 Restrained MD refinement statistics

	S1/M1	S1/M2	S2/M1	S2/M2
Deviation from NMR structure (Å)[c]				
Residues 5–43	4.6 ± 1.2	4.4 ± 1.0	3.7 ± 0.5	3.8 ± 0.5
Residues 5–13 (periplasmic helix, PP)	0.9 ± 0.3	0.8 ± 0.1	1.1 ± 0.1	0.9 ± 0.2
Residues 21–43 (transmembrane helix, TM)	0.9 ± 0.3	1.0 ± 0.1	0.9 ± 0.1	0.7 ± 0.1
Number of experimental NMR restraints[a]				
^{15}N CSA (23 for TM; 3 for PP)	26	26	26	26
^{1}H-^{15}N DC (for TM)	23	23	23	23
^{1}H-^{15}N RDC phage (25 for TM; 18 for PP)			43	43
^{1}H-^{15}N RDC gel (23 for TM; 17 for PP)			41	41
Deviation from NMR restraints[b]				
^{15}N CSA (ppm)	7.3 ± 1.5	6.6 ± 1.3	8.2 ± 1.1	8.8 ± 1.1
^{1}H-^{15}N DC (kHz)	1.0 ± 0.2	1.0 ± 0.2	1.3 ± 0.2	1.3 ± 0.2
^{1}H-^{15}N RDC phage (Hz)			1.2 ± 0.1	1.2 ± 0.1
^{1}H-^{15}N RDC gel (Hz)			1.3 ± 0.1	1.3 ± 0.1

[a] CSA and DC restraints were derived from solid-state NMR experiments with Pf1 in lipid bilayers oriented with n||B_o. RDC restraints were derived from solution NMR experiments with Pf1 in micelles weakly aligned with phage or stressed gels.

[b] Evaluated as RMSD.

[c] Evaluated as RMSD for backbone atoms.

3.2.2 Restrained MD simulations in explicit bilayers

To investigate the influence of the mixed restraints derived from solid-state NMR and solution NMR experiments, we setup two simulations: simulation S1 using only solid-state NMR (DC, CSA) restraints, and simulation S2 using both solid-state NMR (DC, CSA) and solution NMR (RDC) restraints. In addition, we performed a third simulation $S2_{-RDC}$, in which we extended the simulation S2 after removing the RDC restraints at 25 ns of S2.

The Membrane Builder module (49, 50) in CHARMM-GUI (www.charmm-gui.org) (78) was used to insert the Pf1 coat protein structure (average NMR structure in PDB:2KSJ) (75) into a bilayer composed of a 9:1 mixture of 108 DOPC (di-oleoyl-phosphocholine) molecules and 12 DOPG (di-oleoyl-phosphoglycerol) molecules, consistent with the experimental conditions (75). Furthermore, 0.15 M KCl was used for all simulations. Because NMR observables themselves do not provide direct information about insertion depth in a membrane bilayer, we setup two initial models (M1 and M2) of the protein embedded in lipid bilayers (Figure 3.1). In initial model M1, the TM helix was embedded in the bilayer and the periplasmic helix fully solvated by water. In initial model M2, the C-terminus of the TM helix protrudes out of the lipid bilayer (lower leaflet) and the periplasmic helix is embedded in the lipid headgroup region. Initial models M1 and M2 were each used to setup simulations with systems S1 and S2.

For each of these four simulations (S1/M1, S1/M2, S2/M1, S2/M2), the system was replicated and assigned with different initial velocities to generate three independent simulation systems, resulting in a total of 12 simulations. All calculations were performed in NPT (constant particle number, pressure, and temperature) ensembles (51) at 303.15 K using CHARMM (52) with the CHARMM all-atom protein force field (53) including the dihedral cross-term correction

(54), CHARMM36 lipid force field (55), and a modified TIP3P water model (56). To consider potential difference in the number of lipid molecules at the top leaflet depending on the location of the periplasmic helix (i.e., M1 or M2), we have used the P21 periodic boundary condition (79) to allow the lipid molecules to move between the top and bottom leaflets of the bilayer during the simulations. A time step of 1 fs was used with the SHAKE algorithm (57). Each initial system was equilibrated for 300 ps with the nonbonded and dynamics options in the Membrane Builder input; the van der Waals interactions were smoothly switched off at 10–12 Å by a force-switching function (58), and the electrostatic interactions were calculated using the particle mesh Ewald method (59) with a mesh size of ~1 Å for fast Fourier transformation, $\kappa = 0.34$ Å^{-1}, and a sixth-order B-spline interpolation. After equilibration, a 40-ns production run was performed for each system. All the results are presented as an average of three independent simulations and the variations are the standard deviations from the average.

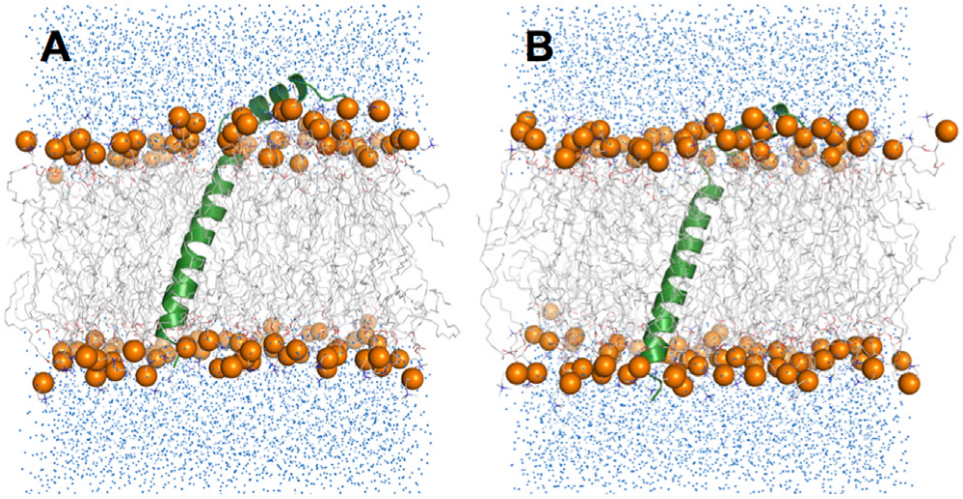

Figure 3.1 Initial structural models, M1 and M2, used for MD simulations of Pf1 coat protein in explicit lipid bilayers. (A) Model M1 has the transmembrane helix well embedded in the lipid bilayer and the periplasmic helix fully solvated by water. (B) Model M2 has the C-terminus of the transmembrane helix protruding out of the lipid bilayer and the periplasmic helix well embedded in the lipid headgroup region. Water molecules are shown as blue dots, the protein is shown in green, lipid hydrocarbon chains are shown as gray sticks, and lipid headgroup phosphorus atoms as orange spheres. Ions are omitted for clarity.

3.3 Results and discussion

3.3.1 Validation of restrained simulations

The average root mean-squared deviation (RMSD) between calculated and experimental NMR observables provides a direct measure of restraint violations in the MD-refined structures. The data in Table 3.1 and Figure 3.2 show that the MD-refined structures in all four simulations satisfy the NMR observables and have a precision that is comparable to those reported for the structure of membrane-bound Pf1 coat protein determined by NMR and refined with XPLOR-NIH (75). The typical experimental errors for these NMR data are <0.5 Hz for RDCs, 0.3 kHz for DCs, and 5–10 ppm for CSAs.

For the individual transmembrane and periplasmic helices, the average backbone RMSDs from the average NMR structure are around 1 Å in all simulations (Table 3.1), indicating that these segments of the protein are stable. However, when the entire structure is considered, the RMSDs are >3.5 Å in all simulations, suggesting that the relative orientations of the transmembrane and periplasmic helices are less well defined, potentially due to a flexible linker and/or difficulties in determining a helix-helix orientation using only orientation-dependent NMR observables. Nonetheless, the MD-refined structures in lipid bilayers can provide more information about protein-lipid interactions and protein dynamics (see below).

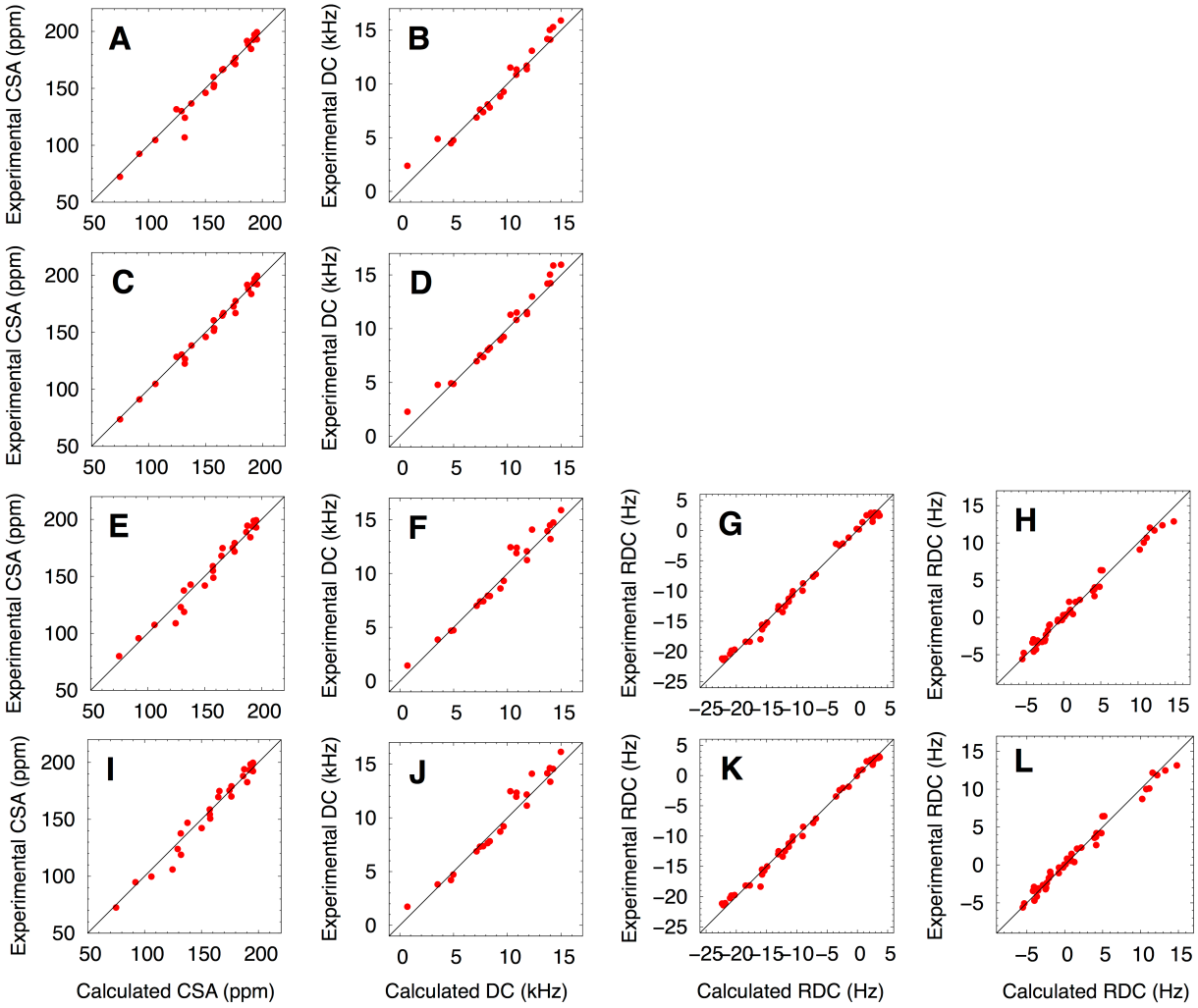

Figure 3.2 Correlation between experimental and back-calculated CSA, DC, and RDC set 1 and 2 for the Pf1 coat protein in (A, B) S1/M1, (C, D) S1/M2, (E, F, G, H) S2/M1 and (I, J, K, L) S2/M2.

3.3.2 Depth of membrane insertion

To quantify the protein's depth of insertion in the lipid bilayer, the membrane insertion depth is defined as the Z coordinate of transmembrane center of mass ($Z_{\text{TM-COM}}$), with the phospholipid bilayer normal aligned parallel to the Z-axis and centered at $Z = 0$. Figure 3.3 shows the change in $Z_{\text{TM-COM}}$ with time during the MD simulations. At the start of the simulations, the values of $Z_{\text{TM-COM}}$ reflect the depth of membrane insertion of the initial models M1 and M2 (Figure 3.1); the transmembrane helix of M1 is well embedded in the bilayer hydrophobic core region and has a larger $Z_{\text{TM-COM}}$ value, whereas the transmembrane helix of M2 is not well embedded and has a smaller $Z_{\text{TM-COM}}$ value (Figure 3.3 *A* and *E*). However, within 6 ns of MD, the transmembrane helix of M2 moves toward that of M1 (Figure 3.3 *B–D* and *F–H*), showing a rapid convergence of transmembrane helix solvation in the membrane environment.

Interestingly, a relatively large range (~8 Å) of transmembrane helix thermal motion along the membrane normal is observed in the last 34-ns MD, reflected as a greater distribution of $Z_{\text{TM-COM}}$ ranging from ~0 to −8 Å (Figure 3.3, *D* and *H*). Even at the limits of this large distribution (−8 Å and 0 Å), only a few transmembrane hydrophobic residues near the membrane interface interact with water molecules occasionally (Figure 3.4) because both the protein and the contact lipid molecules can dynamically adjust to each other. This observation emphasizes the fluidity of the membrane and the dynamics of the protein-environment interactions. Instead of statically staying in the membrane, the protein constantly moves together with several proximal lipids within thermally allowed extents, at least for a single-pass transmembrane helix such as Pf1 coat protein.

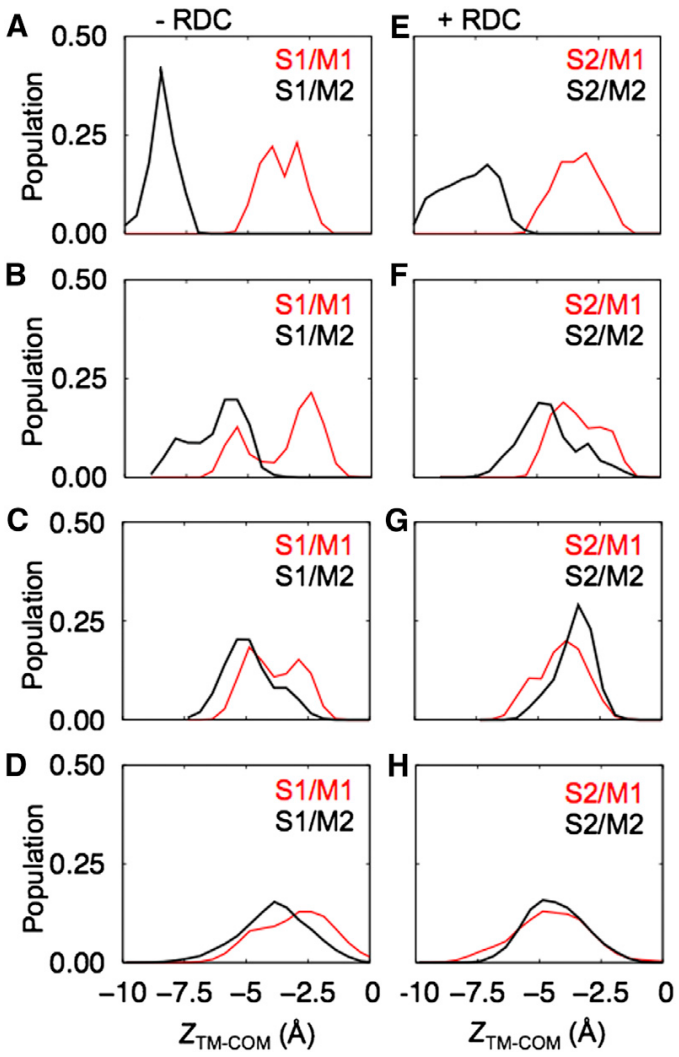

Figure 3.3 Depth of transmembrane (TM) helix membrane insertion. Distribution of the position of the transmembrane helix's heavy atom center of mass along the membrane normal (i.e., the *Z*-axis) obtained at different times of MD simulation for system S1 (A–D) and S2 (E–H) starting from models M1 (red) and M2 (black). Distributions were calculated for MD trajectories of (A, E) 0–0.2 ns, (B, F) 2–3 ns, (C, G) 5–6 ns, and (D, H) 6–40 ns.

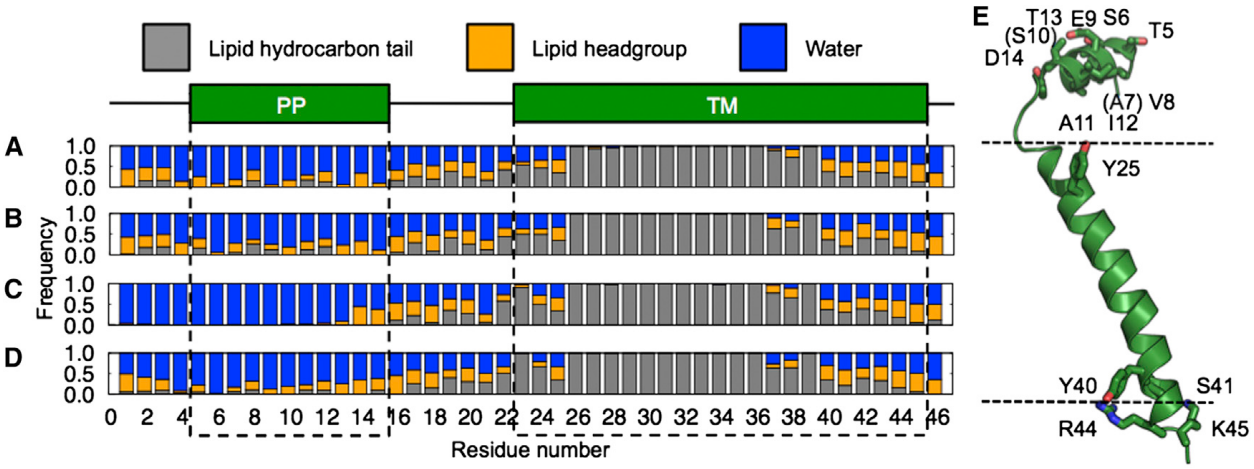

Figure 3.4 Interactions between residues and solvent. The graph shows the frequency with which each one or more atoms of a specific residue is found within 4A° of a lipid hydrocarbon site (gray), lipid headgroup site (orange), or water molecule (blue), after MD simulations with system (A) S1/M1, (B) S1/M2, (C) S2/M1, and (D) S2/M2. Data are from the last 10 ns of MD trajectory. The periplasmic (PP) and transmembrane (TM) helices are designated by green boxes and dashed lines. (E) Residues important for stabilizing transmembrane and periplasmic helix orientations relative to the membrane. Some residues, i.e., A7 and S10, cannot be seen in this side view.

3.3.3 Transmembrane helix orientation, dynamics, and lipid interactions

In all MD-refined systems, the transmembrane helix orientations are similar with small fluctuations in terms of the tilt angles with respect to the membrane normal and the rotation angles along its principal axis (Table 3.2). These orientations are also very similar to those measured from the NMR (PDB:2KSJ) structure.

Solid-state NMR CSA and DC observables provide information about the protein's orientation in the lipid bilayer. In a recent study with solid-state NMR ensemble dynamics simulations of VpuTM, the transmembrane helix of the HIV viral protein Vpu, the fluctuations in helix orientation are larger in the ensemble dynamics simulations than those determined from single conformer simulations restrained by solid-state NMR data (21, 26). With less strict restraints in the former, the variations in helix tilt and rotation increase to 9° and 50° in terms of standard deviations, respectively, similar to the values observed for restraint-free MD simulations. Therefore, the smaller variations observed for Pf1 coat protein in this study (Table 3.2) are possibly due to the strong influence of the applied solid-state NMR restraints.

All four MD-refined systems display a similar pattern of interactions between the transmembrane helix and the lipid molecules (Figure 3.4). In general, the hydrophobic match between the hydrophobic lipid bilayer thickness and the length of the transmembrane domain contributes to the transmembrane helix orientation (37, 39, 42, 80, 81). Specifically, there are five polar or charged residues (Y25, Y40, S41, R44, and K45) located at the water-membrane interface of the transmembrane helix, (Figure 3.4 *E*), which interact extensively with the lipids (Table 3.3) and help stabilize the TM helix orientation and depth of membrane insertion. Indeed,

after 10 ns of MD simulation, the TM helix orientation, depth of membrane insertion, and pattern of protein-environment interactions are converged to similar values.

Table 3.2 Transmembrane helix orientation calculated from last 10 ns of restrained MD

	S1/M1	S1/M2	S2/M1	S2/M2	PDB: 2KSJ
Transmembrane helix tilt (deg)[a]	28.2 ± 1.7	29.2 ± 1.4	29.0 ± 1.3	29.4 ± 1.4	27.0 ± 1.2
Transmembrane helix rotation (deg)[b]	188.7 ± 11.9	185.5 ± 10.0	190.8 ± 8.4	190.8 ± 8.6	192.1 ± 13.2

[a] Defined as the angle between the helix principal axis and the lipid bilayer normal.

[b] Defined as the angle between the perpendicular vector (r_s) from the helical axis to G28 Cα atom and the projection vector (z_p) of the z axis onto the plane made by the second and third principal axes. The sign of the rotational angle becomes positive if $z_p \times r_s$ is in the opposite direction to the helical axis, or negative otherwise (39).

Table 3.3 Frequency of interaction between transmembrane helix polar or charged side chains and phospholipid headgroups observed in the last 10 ns of restrained MD[a]

	S1/M1	S1/M2	S2/M1	S2/M2
Tyr-25	0.91	0.92	0.89	0.93
Tyr-40	0.81	0.73	0.61	0.95
Ser-41	0.95	0.90	0.62	0.82
Arg-44	1.00	1.00	1.00	0.96
Lys-45	0.77	0.96	0.77	0.93

[a] The interaction occurs when any heavy atom of a particular residue's side chain is within 4 Å of a lipid headgroup atom.

3.3.4 Periplasmic helix orientation, dynamics, and interactions

Although the transmembrane helix orientations are similar in all systems, the periplasmic helix behaves differently with strong dependence on the types of NMR restraints as well as the starting model used in the MD simulations. Compared to the 10 structures in the NMR ensemble of PDB:2KSJ, which show small variations of the periplasmic helix tilt angle (68° ± 3°), the restrained MD simulations produced structures with larger flexibility (Figure 3.5).

In system S1, where simulations were performed using only solid-state NMR restraints, the periplasmic helix tilt angle ranges from 40° to 90° regardless of which initial model, M1 or M2, was used (Figure 3.5 *A*). The average periplasmic helix tilts obtained in these cases are consistent with the 68° tilt obtained for the NMR structures calculated previously. Furthermore, for both M1 and M2 initial models, the periplasmic helix displays a similar pattern of interactions with the membrane lipids (Figure 3.4), even though in model M1, the periplasmic helix started out fully solvated by water with little lipid interactions (Figure 3.1). This indicates good convergence of the solid-state NMR restrained MD simulations.

In system S2, where simulations were performed using both solid-state NMR restraints and solution NMR RDC restraints, the simulations with starting models M1 and M2 do not converge to a single tilt angle distribution. Instead, each simulation yields a separate helix tilt distribution, as evidenced by the presence of two distinct peaks each associated with either model M1 or M2 (Figure 3.5 *B*). Simulations that were initiated with model M1 converge to a periplasmic helix tilt of 55° with a range of 45° to 65°. By contrast, simulations initiated with model M2 converge to a periplasmic helix tilt of 75° with a range of 65° to 80°. To examine the contribution of the RDC restraints to this bimodal distribution of tilt angles, we removed the RDC restraints from the

system S2 at 25 ns, and continued the MD simulations for an additional 10 ns with models M1 and M2 (Figure 3.5 *C*). Within 5 ns of MD simulation with this new system ($S2_{-RDC}$), the periplasmic helix reoriented toward the membrane and its orientation converged toward the 68° tilt angle observed for simulation S1, regardless of the initial model. After removing the RDC restraints, the population of structures with small PP helix tilt angles (>60°) diminished rapidly.

Therefore, this result in system S2 (Figure 3.5 *B*) may be due to the presence of degenerate orientation solutions possible for RDC restraints, and also indicates the conflicting influences of solution NMR RDC and solid-state NMR CSA restraints on the PP helix orientation. It is important to note that solution NMR and solid-state NMR restraints are measured on samples with dramatically different physical properties. Micelles used in solution NMR experiments are highly dynamic and very different from biological membranes. Notably, micelles do not provide a proper membrane-water interface as that found in bilayers (82). By contrast, lipid bilayers used in solid-state NMR experiments provide an environment that closely matches the physical and chemical properties of biological membranes. Indeed, structures determined in detergent micelles can exhibit notable differences from those determined in lipid bilayers (83).

Overall, the results in Figure 3.5 suggest that the simultaneous use of solution NMR and solid-state NMR restraints for membrane structure determination may not be universally valid. Such hybrid approaches are based on the assumption that membrane proteins adopt the same structure in micelles as they do in bilayers (76). However, for many proteins this is not the case, especially for membrane proteins with membrane-water interfacial domains (83). In the case of Pf1 coat protein, the greater degree of fluctuations associated with the RDC restraints measured in detergent micelles appears to be incompatible with the solid-state NMR restraints measured in

phospholipid bilayers, and with the structure of the amphipathic periplasmic helix located at the membrane-water interface.

As shown in Figure 3.4, *A* and *B*, the structures with large periplasmic helix tilt angles resulting from simulations performed exclusively with solid-state NMR restraints display significantly more extensive interactions of the periplasmic helix with the lipid bilayer. These structures satisfy the amphipathic character of the periplasmic helix, enabling hydrophobic interactions to occur between lipid molecules and apolar residues (A7, V8, A11, I12), and hydrophilic interactions to occur between water and polar residues (T5, S6, E9, S10, T13, D14) on the opposite side (Figure 3.4 *E*, Figure 3.6, *A* and *B*). By contrast the structures resulting from MD simulations that include RDC restraints measured from micelle samples show little or no interaction of the periplasmic helix with the membrane surface (Figure 3.4, *C* and *D*, Figure 3.6, *C* and *D*); in these cases the result is highly dependent on the starting model and the periplasmic helix remains highly solvated by water.

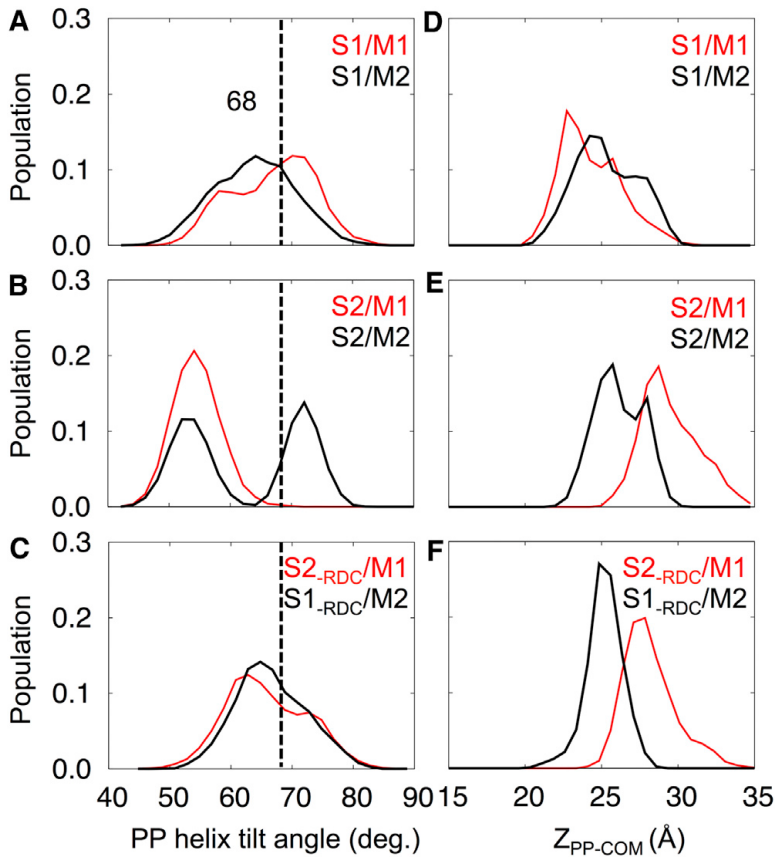

Figure 3.5 The orientation and depth of membrane insertion of the periplasmic (PP) helix. (A–C) The distribution of periplasmic helix tilt angle with respect to the membrane normal (i.e., *Z*-axis). (D–F) the distribution of the center of mass of heavy atoms in periplasmic helix along the *Z*-axis. These distributions were calculated from the last 10 ns of MD trajectory for systems S1, S2, and $S2_{\text{-RDC}}$ starting with models M1 (red) and M2 (black). The dashed lines mark the average helix tilt angle obtained for the NMR structure (PDB:2KSJ).

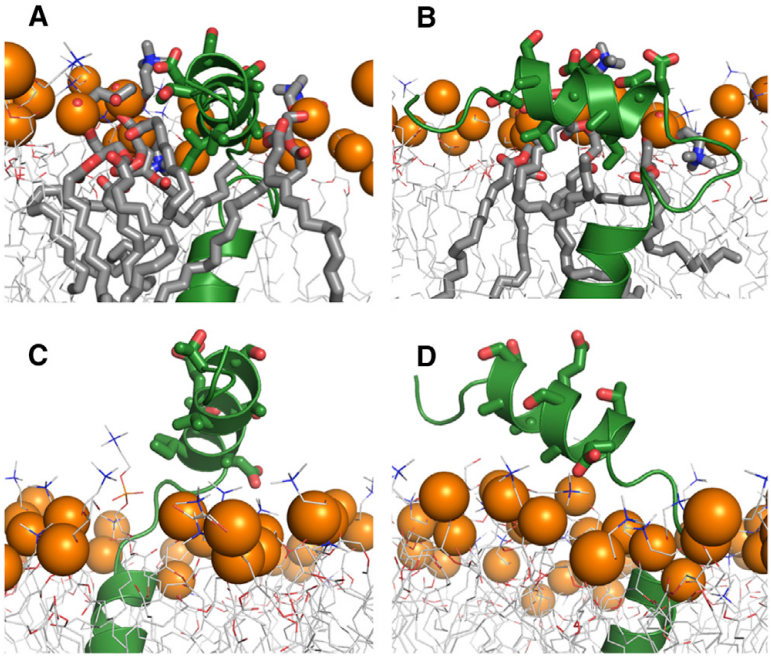

Figure 3.6 Snapshots showing the interactions between the periplasmic helix of Pf1 coat protein and lipid molecules in (A, B) system S1 and (C, D) system S2. The protein is shown in green. Residues in periplasmic helix are shown as sticks. Lipids are shown as gray lines; those contacting with periplasmic helix are drawn as sticks. Lipid headgroup phosphorus atoms are presented as orange spheres. Water molecules and lipids blocking the views are omitted for clarity.

3.4 Conclusions

The results of NMR-restrained MD simulations of Pf1 coat protein in explicit phospholipid bilayer membranes show that it is possible to generate stable MD trajectories using NMR orientation restraints. These restrained simulations provide a powerful refinement tool for membrane protein structure determination by NMR spectroscopy. The NMR restraints are well satisfied and the simulations provide rich structural and dynamics information about the protein embedded in a realistic membrane environment, including: detailed side chain-side chain and side chain-lipid interactions, depth of membrane insertion, and protein dynamics in the context of the experimental restraints. Such detailed information cannot be obtained solely from NMR observables. Thus, the present simulations illustrate the usefulness of the NMR membrane protein structure refinement in explicit membranes.

The results also highlight the importance of performing both the calculations and the experiments in the proper environment. MD simulations performed in explicit lipid bilayers are fully compatible with experimental restraints also measured in lipid bilayers. By contrast, the incorporation of structural restraints measured in a dramatically different setting (such as RDCs measured in detergent micelles) lead to conflicting and ambiguous structural results, which do not converge to a stable structure. This effect is particularly evident for the periplasmic amphipathic helix of Pf1 coat protein. Indeed, the influence of environment is likely to be very important for those structural elements of proteins located at membrane-water interfacial regions, as noted recently for the influenza M2 ion channel (83). Additionally, as reported for paradaxin, even in the bilayers, other membrane components, such as cholesterol, can alter the protein structure (84).

Finally, we have recently shown that solid-state NMR-restrained ensemble dynamics, performed in implicit membrane models, are very effective for rendering the structural and dynamics information that is embedded in solid-state NMR restrains (21, 26). In solid-state NMR ensemble dynamics, an ensemble of structures is simulated in parallel MD calculations, and ensemble-averaged solid-state NMR observables across the ensemble, rather than a single individual structure, are restrained to the experimental values. This approach is valid because the observables measured in solid-state NMR experiments are time- and ensemble-averaged properties. Therefore, it will be of particular interest to perform solid-state NMR ensemble dynamics of Pf1 coat protein in explicit bilayers and compare the result with this study.

4. Solid-state NMR based ensemble dynamics simulations of fd coat protein in explicit membranes

4.1 Introduction

Solid-state nuclear magnetic resonance (NMR) spectroscopy is a powerful method for determining the structures of membrane proteins in native-like phospholipid bilayer membranes. Measurements of dipolar coupling (DC) and chemical shift anisotropy (CSA) solid-state NMR signals provide precise orientation-dependent restraints that can be used to determine the three-dimensional structure and global orientation of a membrane embedded in phospholipid bilayers (85, 86). Solid-state and solution NMR observables represent time- and ensemble-averaged measurements, and therefore contain both structural and dynamics information (87-89). Recently, we developed an ensemble dynamics (ED) technique that uses solid-state NMR observables and enables detailed characterization of the orientational fluctuations of transmembrane helices (26, 27). In these calculations, the membrane was represented implicitly, with the lipid bilayer modeled as a static slab of fixed width, with low dielectric constant in the hydrophobic core, to mimic the properties of biological membranes. In these calculations, the lipid bilayer membrane was represented implicitly to increase computational efficiency (30, 31). However, even though such implicit model calculations are computationally efficient they cannot provide detailed atomic-level information about protein-lipid interactions.

By contrast, calculations performed with explicit representation of the lipid bilayer membrane reflect the atomic-level interactions of proteins with the surrounding lipid molecules,

and the phospholipid bilayer is fluid and deformable. For example, in our recent structural refinement of the membrane-embedded Pf1 coat protein using NMR-restrained molecule dynamics (MD) simulations, a relatively large thermal motion (~8 Å) of the transmembrane helix along the membrane normal was observed (90). Even at the limits of such a large amplitude fluctuation, only a few transmembrane hydrophobic residues near the membrane interface were seen to interact with water molecules, because both the membrane-embedded protein and the surrounding lipid molecules can dynamically adjust to complement each other during the simulations. In another recent MD simulation study of the transmembrane peptide gramicidin A in membranes of various phospholipid compositions, we observed that local changes in lipid packing are induced by protein-lipid interactions, and lead to variations in lipid bilayer thickness as a function of radial distance from the gramicidin A channel (37). Such bilayer adaptations resulting from the interactions of proteins with lipids are difficult to model in simulations performed with implicit membranes, because most implicit membrane models cannot capture the adaptation of lipids to protein and vice-versa.

In this study, we describe a solid-state NMR restrained ED protocol for membrane protein structure refinement in explicit bilayer membranes. By incorporating the effect of the fluid lipid bilayer, it is expected that this ED approach in explicit membranes can readily characterize not only protein dynamics information but also protein-lipid interactions in detail, while maintaining the agreement with the experimental NMR data. As the first application of this approach, the membrane-bound form of the major, pVIII coat protein from filamentous fd bacteriophage (hereinafter called fd coat protein) is used as a model membrane protein. The fd coat protein resides within the membrane of infected bacteria before assembling into new virus particles (91,

92). In its membrane-embedded state, it forms one N-terminal periplasmic helix (residues 8-18) that lies on the membrane surface and one transmembrane helix (residues 21-45) linked by a short loop (residues 19-20) (91). During bacteriophage assembly, the transmembrane helix extrudes from the bacterial membrane while positively charged side chains in its C-terminus interact with the bacteriophage's DNA that is packaged in the new phage particle (92).

The previously reported structure of membrane-inserted fd coat protein (PDB: 1MZT) was determined using ^{15}N chemical shift and ^{1}H-^{15}N dipolar coupling frequencies measured in PISEMA NMR resonances (91). These frequencies depend on the orientations of the corresponding molecular sites, and provide orientational restraints for structure determination. As one of the first examples of membrane protein structure determination by solid-state NMR, the fd coat protein structure was calculated by converting the experimental NMR frequencies into backbone dihedral angles, without including the side chains, without refinement by simulated annealing and without considering environmental effects. Using this previously reported structure and all its experimental ^{15}N CSA and ^{1}H-^{15}N DC solid-state NMR observables (91), we performed restrained ED simulations of fd coat protein in explicit phospholipid bilayers. To determine an optimal ensemble size for extracting dynamics of the protein and characterizing its interactions with the lipid molecules, different numbers of replicas were used to investigate the influence of solid-state NMR restraints on the resulting ensemble structures and dynamics. In addition, a standard MD simulation without solid-state NMR restraints was performed and its results were compared with those from solid-state NMR-restrained ED simulations.

4.2 Materials and methods

4.2.1 Solid-state NMR ensemble restraint potentials

We implemented the solid-state NMR ensemble restraint potentials (U_χ) (93) as simple harmonic restraint potentials in CHARMM (52). U_χ is applied over a certain number of replicas (N_{REP}) to constrain the ensemble-averaged property ($\langle \chi_i^{\mathrm{calc}} \rangle_{\mathrm{ens}} = \sum_{j=1}^{N_{\mathrm{REP}}} \chi_i^{\mathrm{calc}(j)} / N_{\mathrm{REP}}$) to the experimental value (χ_i^{exp}) at each simulation time step:

$$U_\chi = N_{\mathrm{REP}} \sum_{i=1}^{N_\chi} k_\chi \left(\langle \chi_i^{\mathrm{calc}} \rangle_{\mathrm{ens}} - \chi_i^{\mathrm{exp}} \right)^2 \tag{4-1}$$

where χ is either ^{15}N CSA or ^{1}H-^{15}N DC, N_χ is the number of target experimental observables, and k_χ is the force constant (see reference (19) for detailed calculation methods for each solid-state NMR observable). To avoid over-fitting protein structure to the experimental data, we performed cross-validation of the restraint potentials with different force constants (26, 76, 77). Using this approach, we determined an optimal set of force constants for k_{CSA} = 0.01 kcal/(mol·ppm^2) and k_{DC} = 0.5 kcal/(mol·kHz^2). The total potential energy (U_{TOTAL}) of the ensemble system is

$$U_{\mathrm{TOTAL}} = U_{\mathrm{CHARMM}} + U_{\mathrm{CSA}} + U_{\mathrm{DC}} \tag{4-2}$$

where U_{CHARMM} is the standard CHARMM potential energy.

4.2.2 Simulations in explicit lipid bilayers

Membrane Builder (49, 94) in CHARMM-GUI (www.charmm-gui.org/input/membrane) (78) was used to insert the fd coat protein structure (PDB:1MZT) (91), determined with ^{15}N CSA and ^{1}H-^{15}N DC solid-state NMR observables, in a phospholipid bilayer composed of a 4:1 molar mixture of POPC (palmitoyl-oleoyl-phosphatidylcholine) and POPG (palmitoyl-oleoyl-phosphatidylglycerol), consistent with the experimental conditions. In addition, 150 mM KCl was used for all simulations.

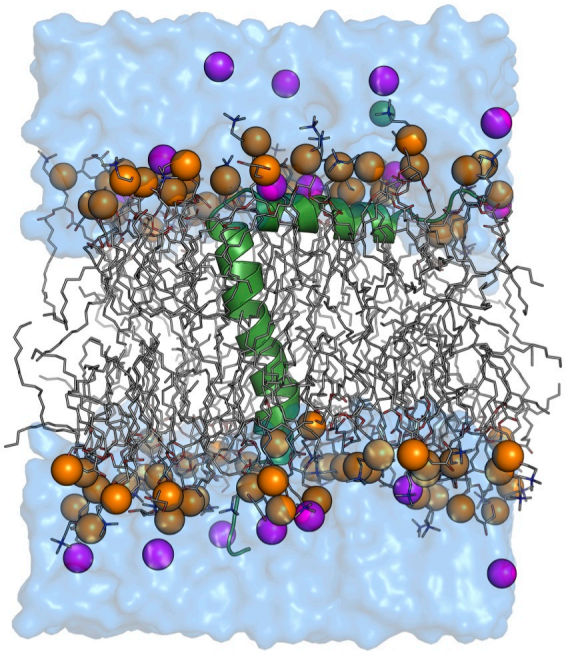

Figure 4.1 Ensemble dynamics simulation system of fd coat protein in a POPC/POPG bilayer. Protein is shown as a cartoon in green; lipids are shown as sticks with phosphorus atoms as spheres (orange); ions are shown as spheres (potassium in magenta and chloride in green); water is shown as surface (blue).

The ED simulation system (Figure 4.1) has an initial size of 56.3 Å × 56.3 Å × 71.9 Å, and contains 68 POPC, 17 POPG, and 3,488 water molecules, yielding a total number of 22,499 atoms. To examine the impact of solid-state NMR restraints on structure and dynamics, as a function of ensemble size, six solid-state NMR-restrained ED simulation systems were built: ED1, ED2, ED4, ED8, ED16, and ED32, each with 1, 2, 4, 8, 16 or 32 protein replicas (Table 4.1). All available experimental solid-state NMR restraints of PDB:1MZT (Table 4.2) were applied in each restrained system. To avoid a bias resulting from an unequal number of replicas, all calculations were performed with an equal total number of 32 replicas in each system (Table 4.1). Each replica was assigned with different initial velocities. All calculations were performed in NPT (constant particle number, pressure, and temperature) ensemble (51) at 295.15 K using CHARMM (52). We used the CHARMM all-atom protein force field (53) including the dihedral cross-term correction (54), the CHARMM36 lipid force field (55), and a modified TIP3P water model (56). To account for potential differences in the number of lipid molecules present in each bilayer leaflet due to the location of the protein's N-terminal helix on the membrane surface, we used the $P2_1$ periodic boundary condition (79) that allows lipid molecules to move between the top and bottom leaflets of the bilayer during the simulations. A time step of 1 fs was used for all ED simulations with the SHAKE algorithm (57), as the ED simulations with 2 fs were not stable due to the solid-state NMR restraint potentials. Each initial system was equilibrated for 15 ns; van der Waals interactions were smoothly switched off at 10-12 Å by a force-switching function (58) and the electrostatic interactions were calculated using the particle mesh Ewald method (59) with a mesh size of ~1 Å for fast Fourier transformation, $\kappa = 0.34$ Å^{-1}, and a sixth-order B-spline interpolation. After equilibration, a 30-ns production was performed for each solid-state NMR ED simulation. All the analyses were done using the last 20-ns trajectories of ED simulations,

and presented as an average of replicas in the ensembles. The variations are the standard deviations from the average.

Table 4.1 System information for ED and MD simulations.

System ID	# Replicas per Ensemble	# Ensembles per Systems
ED1	1	32
ED2	2	16
ED4	4	8
ED8	8	4
ED16	16	2
ED32	32	1
MD1	1	1

Table 4.2 Structural statistics

	MD1	ED1	ED2	ED4	ED8	ED16	ED32
Number of experimental solid-state NMR restraints							
^{15}N CSA[a]	0	38	38	38	38	38	38
^{1}H-^{15}N DC[b]	0	38	38	38	38	38	38
RMSD from solid-state NMR restraints[c]							
^{15}N CSA (ppm)	26.7 ± 5.7	5.6 ± 0.6	4.2 ± 0.6	3.2 ± 0.3	2.9 ± 0.2	2.7 ± 0.2	2.6 ± 0.2
^{1}H-^{15}N DC (kHz)	5.7 ± 1.2	0.8 ± 0.1	0.6 ± 0.1	0.5 ± 0.0	0.5 ± 0.0	0.4 ± 0.0	0.4 ± 0.0
RMSD from PDB structure (Å)[d]							
Residues 1-50	4.5 ± 0.6	4.5 ± 0.6	4.8 ± 0.7	4.7 ± 0.7	4.5 ± 0.7	4.7 ± 0.7	4.8 ± 0.7
Residues 8-18 (PP[e])	0.5 ± 0.1	0.4 ± 0.1	0.5 ± 0.1	0.5 ± 0.1	0.5 ± 0.1	0.5 ± 0.1	0.5 ± 0.2
Residues 21-45 (TM[f])	2.5 ± 0.2	2.3 ± 0.2	2.2 ± 0.2	2.2 ± 0.2	2.2 ± 0.2	2.3 ± 0.2	2.3 ± 0.2
Residue 8-18 and 21-45	3.0 ± 0.4	2.4 ± 0.2	2.6 ± 0.5	2.6 ± 0.5	2.5 ± 0.5	2.7 ± 0.5	2.7 ± 0.5
Residue 21-35	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.2	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.1
Residue 38-45	0.9 ± 0.1	0.5 ± 0.1	0.7 ± 0.3	0.7 ± 0.3	0.7 ± 0.3	0.8 ± 0.4	0.8 ± 0.4

[a]11 CSA for TM, 25 CSA for PP, and 2 for the loop linking TM and PP.

[b]11 DC for TM, 25 DC for PP, and 2 for the loop linking TM and PP.

[c]Evaluated as RMSD. [d]Evaluated as RMSD for backbone atoms (CA, C, N and O).

[e]PP for the periplasmic domain. [f]TM for the transmembrane domain.

For a standard, unrestrained MD simulation, the MD1 system had a greater initial size (81.2 Å × 81.2 Å × 82.3 Å) and thus larger total number of atoms (54,391) to meet the technical requirements for simulations on Anton computer (95). The MD1 system has 148 POPC, 37 POPG, and 9,688 water molecules. We performed a 15-ns equilibration and a 500-ns production on Anton (95) using the CHARMM36 force field (53-56). The simulations were performed at constant temperature (295.15 K) and pressure (1 atm) using the Nosé-Hoover thermostat and the semi-isotropic Martyna-Tuckerman-Tobias-Klein barostat (96, 97). The time-step was 2 fs and trajectories were saved every 240 ps. Short-range non-bonded and long-range electrostatic interactions were evaluated with a cutoff of 9.52 Å every 2 fs and 6 fs, respectively. Long-range electrostatic interactions were calculated using the k-Gaussian Split Ewald method (98) with a 64 × 64 × 64 grid. SHAKE was used to constrain all bonds involving hydrogen atoms. The last 200-ns trajectory of MD1 simulation was used for analysis.

4.3 Results and discussion

4.3.1 Validation of solid-state NMR restrained ensemble dynamics

When the average root-mean squared deviations (RMSD) of back-calculated CSA of a simulation structure (or an ensemble of structures) from the experimental CSA is smaller than 10 ppm, the structure "satisfies" the experimental CSA. In terms of DC, the criterion is 1 kHz. The average CSA and DC RMSD of PDB:1MZT are 5.7 ppm and 0.7 kHz (Figure 4.2 *A* and *B*). Since the solid-state NMR restraints were applied over an ensemble, by definition, the CSA and DC RMSD for each replica in a certain ensemble do not necessarily satisfy the experimental solid-state NMR observables (Figure 4.2 *E* and *F*). However, the ensemble-averaged CSA and DC RMSD are very similar to the experimental measures (Figure 4.2 *C* and *D*). The data in Figure 4.3 and Table 4.2 show that the ensemble structures resulting from all solid-state NMR restrained ED simulations satisfy the experimental solid-state NMR data. As the number of replicas per ensemble increases, the RMSD of the CSA and DC decrease (Figure 4.3), indicating that the resulting structure ensembles obtained with more replicas better represent the experimental data. The same trend was observed in our previous solid-state NMR restrained ED simulations performed in implicit membranes (26, 27). The CSA and DC RMSD observed for MD1, performed without solid-state NMR restraints, are about 5-fold larger than those for ED1.

A number of factors contribute to the observed RMSD and complicate the precise reproduction of the experimental NMR data with the restrained ED simulations. The use of a single order tensor for the entire protein does not accurately represent the actual system, where the order parameter of a membrane-embedded site is likely greater than of a water-exposed site. Furthermore, the use of a single ^{15}N CSA tensor for all amino acids in the protein facilitates the

calculations but also contributes to the ^{15}N CSA RMSD. Similarly, uncertainty about the precise length of the amide NH bond, and use of a single value for all amino acids, contribute to the ^{1}H-^{15}N DC RMSD. Furthermore, the relative values of the restraining force constants for ^{15}N CSA and ^{1}H-^{15}N DC have been shown to impact structure calculations (99). In the context of molecular dynamics calculations, it is difficult to reproduce the experimental solid-state NMR data directly from standard MD simulations (as in MD1), because each residue is unrestrained by experimental data, and therefore has more orientational degrees of freedom. Since a change in the orientation of a molecular site leads to a distinctive solid-state NMR measurement, subtle differences in the orientation and conformation of fd coat protein during MD can lead to large RMSD.

Notwithstanding these factors, the CSA and DC RMSD observed for both ED and MD simulations do show that the calculated structures are representative of the experimental NMR data. Given the complicating factors described above it is also notable that the unrestrained free MD simulation also produces a reasonable level of agreement with the experimental data. We anticipate that future calculations performed with more realistic order parameters for different protein segments, different values of the ^{15}N CSA and ^{1}H-^{15}N DC NMR spin tensors, and possibly different values of the restraining force constants, will yield structures with significantly lower experimental RMSD.

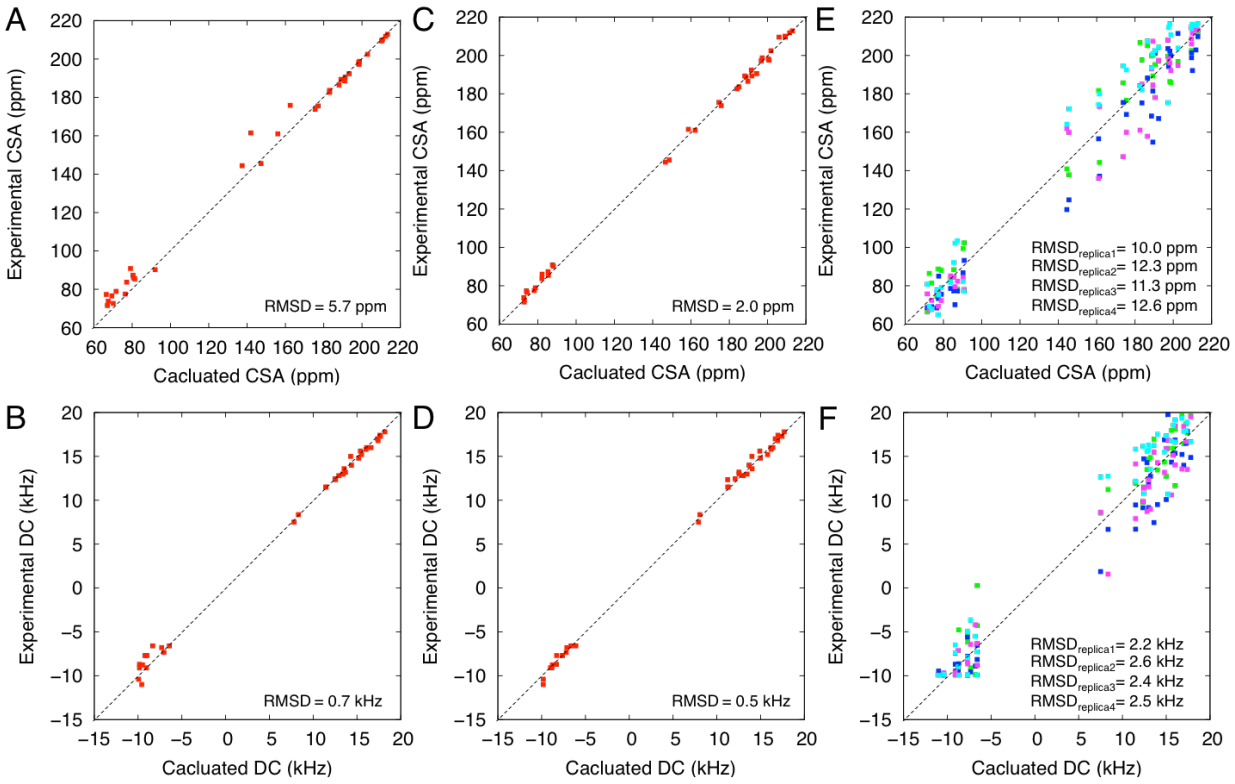

Figure 4.2 Comparison between the experimental and back-calculated CSA and DC for the fd coat protein in (A, B) PDB:1MZT, and (C-F) a representative snapshot of ED4 system at 20 ns. In (C, D), CSA and DC for each residue are averaged over the ensemble structures from the four replicas. In (E, F), CSA and DC for each residue are calculated for each structure in the four replicas (represented in four different colors; replica1 in green, replica2 in blue, replica3 in cyan, replica4 in magenta). Note that by definition each structure in the restrained ensemble simulations (with more than one replica) may not satisfy the experimental restraints, but the ensemble structures from individual replicas collectively satisfy the experimental restraints.

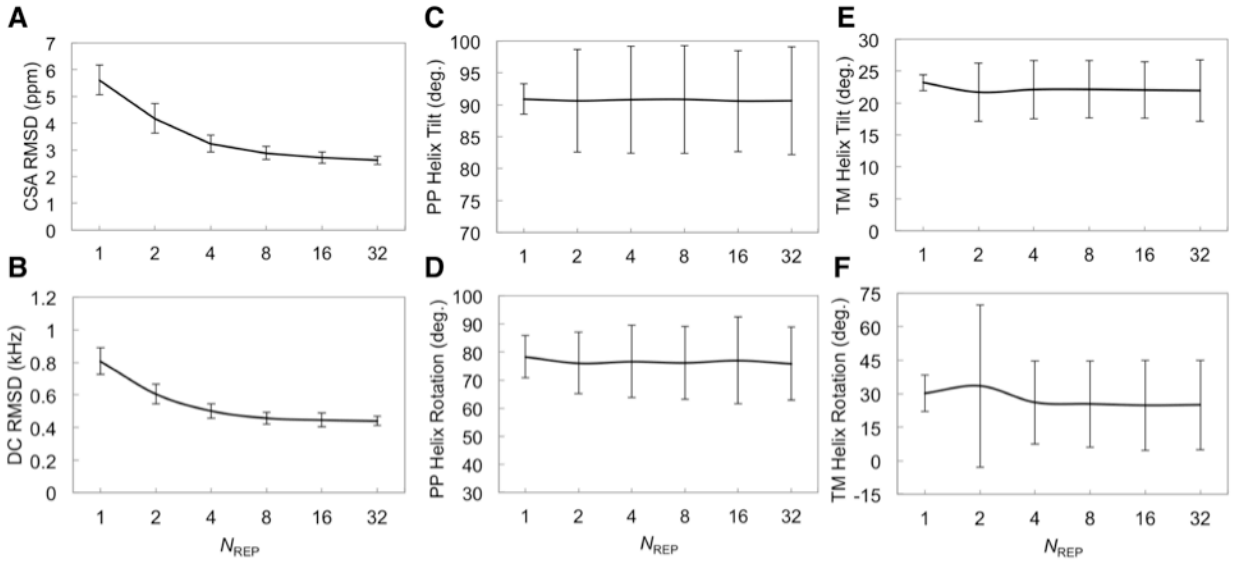

Figure 4.3 Validation of fd coat protein structure ensemble. (A and B) CSA and DC RMSD with respect to the experimental observables as a function of the number of replicas per ensemble simulation. (C and D) The tilt and rotation angle of the periplasmic helix in fd coat protein. (E and F) The tilt and rotation angle of the transmembrane helix in fd coat protein. The tilt angle is defined as the angle between the helix principal axis and the lipid bilayer normal. The rotation angle is defined as the angle between the perpendicular vector ($\mathbf{r}_s$) from the helical axis to a Cα atom (S13 for the periplasmic helix and G34 for the transmembrane helix) and the projection vector ($\mathbf{z}_p$) of the Z axis onto the plane made by the second and third principal axes. The sign of the rotational angle becomes positive if $\mathbf{z}_p \times \mathbf{r}_s$ is in the opposite direction to the helical axis, or negative otherwise. The error bars show the standard deviations from the average.

In all ED and MD simulations, the backbone atom RMSD relative to the PDB structure is between 4.5 and 4.8 Å, when all residues (1-50) are included in the calculation. This is a relatively large structural deviation for a small membrane protein like fd coat. However, the backbone RMSD becomes smaller (between 2.4 and 3.0 Å) when only helical residues (8-18 and 21-45) are included, indicating structural flexibility of the loop and terminal residues. Furthermore, as shown in Table 4.2, the N-terminal periplasmic helix (residues 8-18) has smaller backbone atom RMSD (0.4-0.5 Å) than those (2.2-2.5 Å) of the transmembrane helix (residues 21-45). Although both segments Y21-A35 and G38-F45 have backbone RMSD in the range of 1.2 Å, it is their different relative orientation that contributes to a higher overall RMSD with respect to the PDB structure. As shown in Figure 4.2, the average per-residue RMSD for the solid-state NMR restraints are <10 ppm for CSA and <1.5 kHz for DC for all transmembrane helix residues (Y21-F45) in all six ED systems. This suggests that the structures from all ED simulations generally have different transmembrane conformations from PDB:1MZT, but they also satisfy the experimental observables very well. This fact is further elaborated below.

4.3.2 Influence of solid-state NMR restraints on helix orientations

As shown in Figure 4.3 *C-F*, the ensemble-averaged helix tilt angles (with respect to the membrane normal) and helix rotation angles (along the helix axis) are similar in all ED systems, and also agree with the PDB structure (tilt angles of 91.7° for periplasmic and 20.5° for transmembrane helices, and rotation angles of 76.2° for periplasmic and 25.7° for transmembrane helices). However, the variations for tilt and rotation angles in ED1 is obviously smaller than those in ED2, ED4, ED8, ED16, and ED32 with multiple replicas in each ensemble.

Similarly, as shown in Figure 4.4, ED1 always yields highly homogenous orientational populations in terms of tilt angles, indicating such single conformer simulations are under a large influence of the solid-state NMR restraints. In Figure 4.4 *A*, system ED2 presents two major orientational populations of the periplasmic helix. However, no visible, distinct inter-atomic interactions are observed to cause such difference in ED2. Theoretically, it has been shown that restrained ED simulations provide maximum likelihood distributions that satisfy the imposed restraints (100, 101). These distinct populations in ED2 disappear in the ensembles calculated with more replicas, and are likely an artifact of applying ensemble-averaged restraints with too small number of replicas. In addition, the distributions obtained from ED4, ED8, ED16, and ED32 are very similar and their broadness is comparable to that of MD1, indicating that the minimum, optimal number of replicas is 4 for these ED simulations.

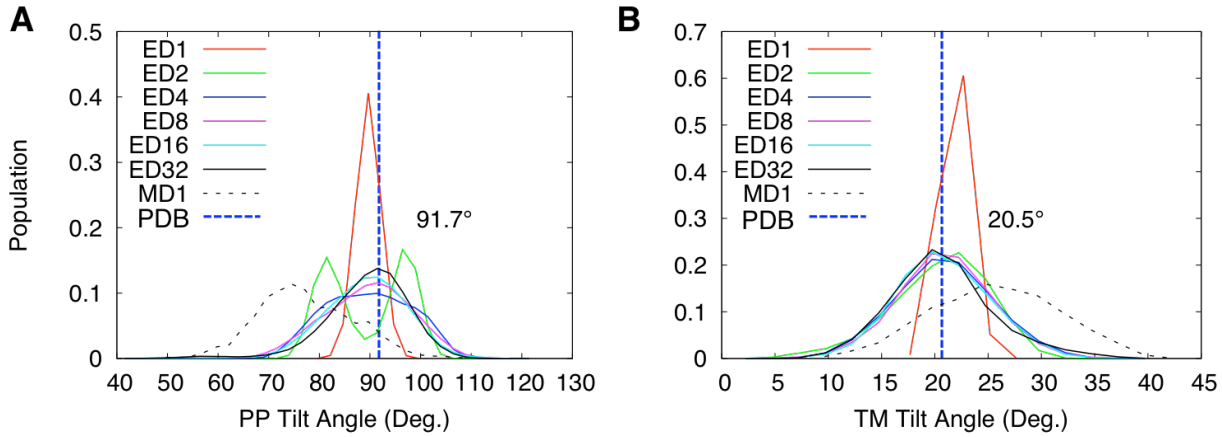

Figure 4.4 Distributions of the tilt angles of (A) the periplasmic (PP) helix and (B) the transmembrane (TM) helix with respect to the membrane normal.

4.3.3 Flexibility of transmembrane helix conformations in simulations

In the PDB structure, a visible kink in the transmembrane helix occurs at I37 with an angle of 17.7°. Such pronounced kinks (>15°) at I37 are rarely observed in the ED and MD simulations and most structural populations appear to have a straight transmembrane helix (Figure 4.5). We note that the PDB structure was calculated directly from the NMR data, only for backbone atoms, without applying any refinement protocol. Improvements in structural quality upon refinement of NMR structures with simple repulsive potentials or in the presence of either implicit or explicit solution are well known (19, 73, 102-105). Therefore, structure calculation using restrained ED in an explicit lipid bilayer membrane, as described here, is expected to yield a significant improvement in quality. To further examine the transmembrane helix structure that best fits the solid-state NMR data, we selected 32 structures with the lowest NMR restraint violations; in this population, the transmembrane helix structures have the RMSD of < 3.5 ppm in CSA and < 0.8 kHz in DC. Within this population, a few structures have a transmembrane helix kink angle near 17° (Figure 4.5 *B*), as observed in PDB:1MZT, indicating that the latter still represents a thermally accessible conformation of the transmembrane helix (Figure 4.5 *C*-*H*) with relatively low population. The kink angle distribution range is ~25° in multiple-replica systems, indicating the transmembrane helix has flexibility near its C-terminal end. Notably, two adjacent glycines (G34/G38) lining on the same side of the transmembrane helix may help confer flexibility in this region of the protein.

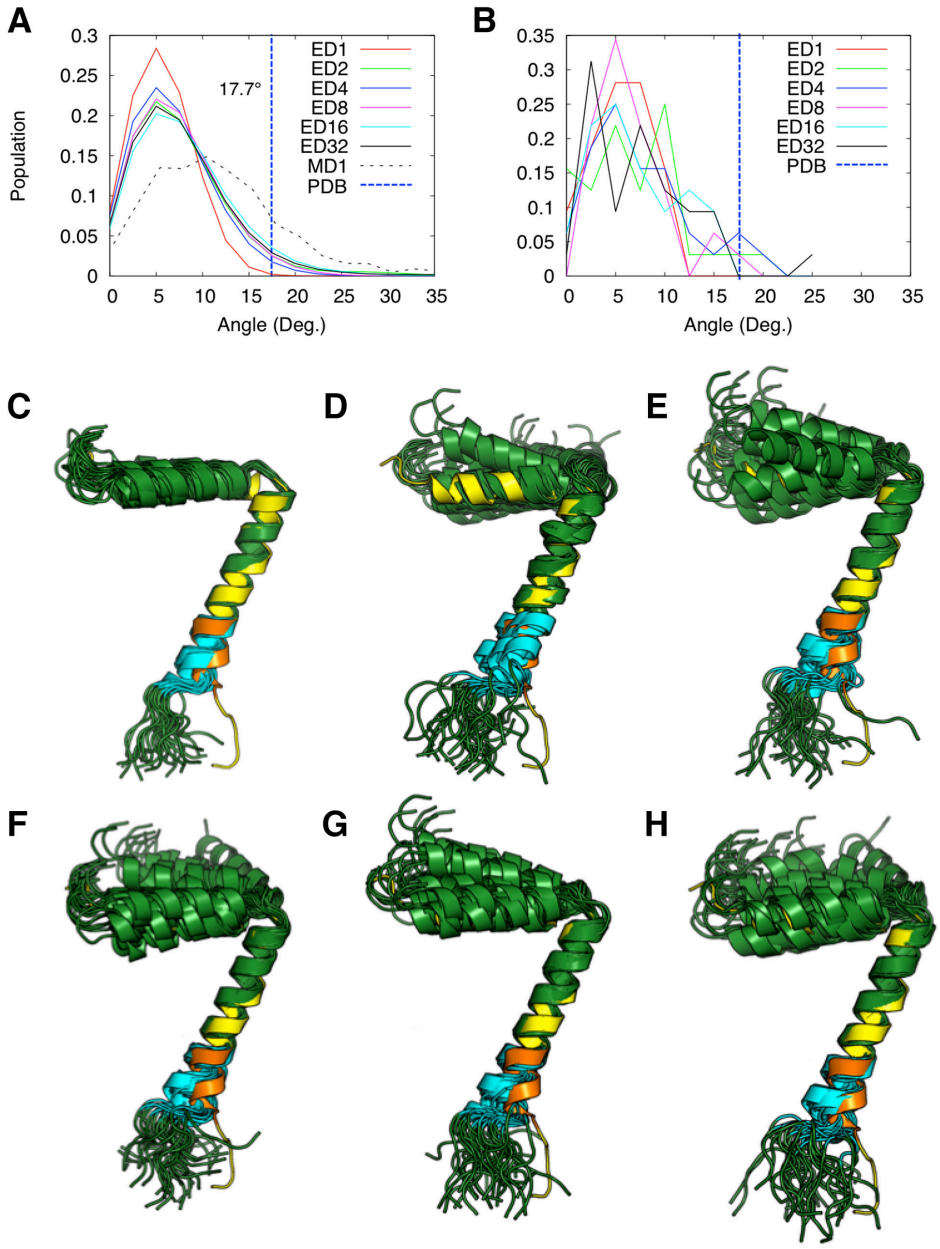

Figure 4.5 Distributions of the angles between the principal axes of two transmembrane helical segments Y21-A35 and G38-F45 in (A) all structures from simulations and (B) top 32 structures with the least NMR violations. Top 32 structures (green) are aligned to PDB:1MZT

(yellow) with respect to Y21-T36: (C) ED1, (D) ED2, (E) ED4, (F) ED8, (G) ED16, and (H) ED32. The G38-F45 region is highlight in orange in PDB:1MZT and in cyan in top 32 structures.

4.3.4 Protein-lipid interactions in explicit membranes

The flexibility of the transmembrane helix at the C-terminus could be functionally important. During bacteriophage extrusion across the bacterial membrane, flexibility in this region could facilitate binding of bacteriophage DNA by the charged amino groups of lysine side chains (K40, K43, and K44), and thus assist bacteriophage packaging. In all ED and MD simulations, these three lysine residues near the C-terminus make frequent interactions with the lipid polar headgroups, lipid hydrocarbon tails, and surrounding water, and thus help anchor the transmembrane helix within the membrane (Figure 4.6). Interestingly, even though K40 is positioned close to the hydrophobic core, lipids in the lower leaflet (i.e., the C-terminal side) can move up and adjust themselves to have the charged side chain of K40 exposed to the solvent. As shown in Figure 4-7 *A*, using ED32 as a representative system, the Z coordinates of lipid phosphorus atoms near K40 (black area) are closer to the bilayer center, i.e., the membrane thinning near K40. In other words, mainly K40 can induce membrane curvature, which is observed in all ED and MD simulations. It is interesting to see lipid deformation induced by fd coat protein, which is involved in the assembly and extrusion of virus particle (106), but linking this observation to the protein function characterization is beyond our scope.

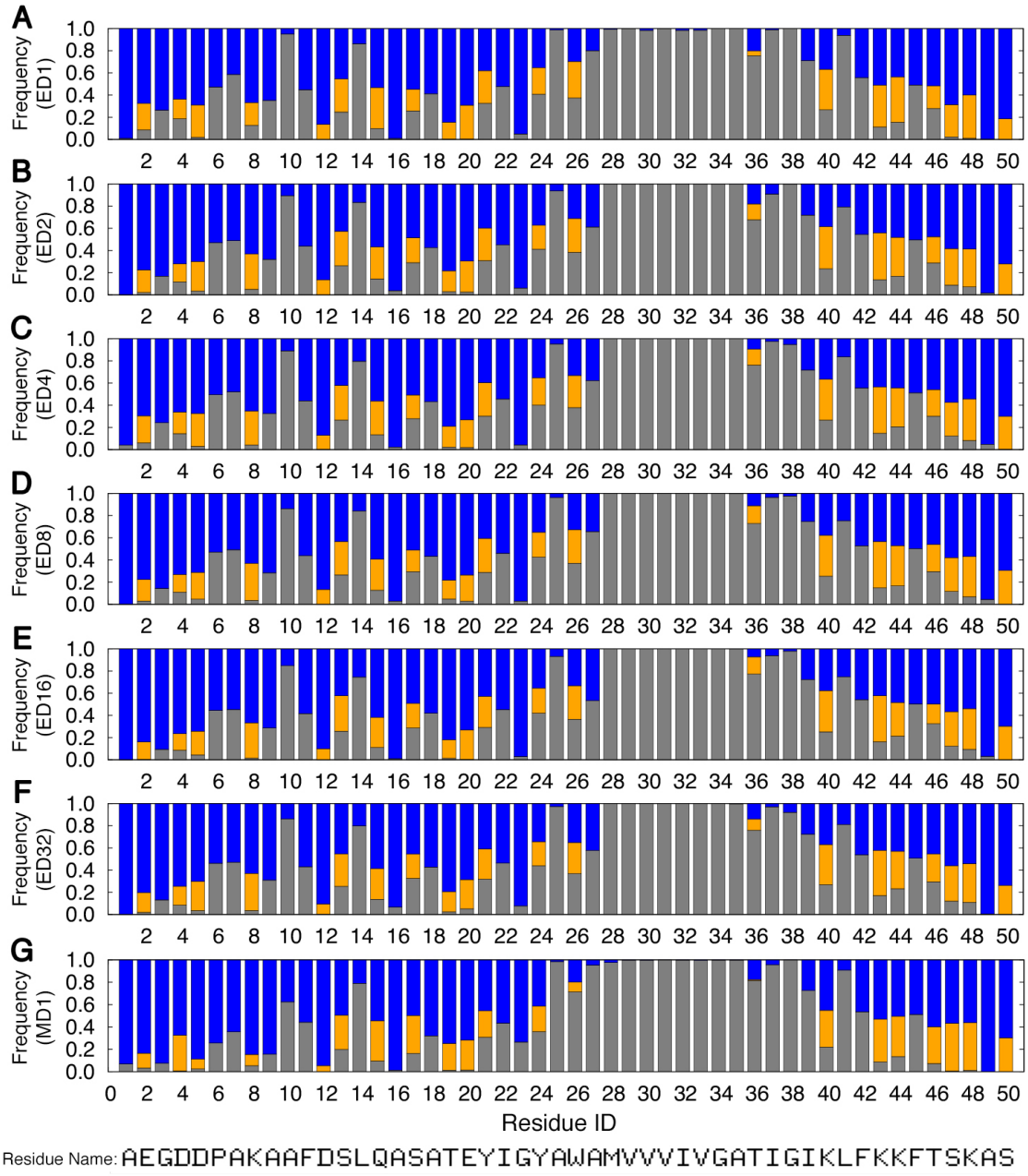

Figure 4.6 Interactions between each residue and solvent components. The graph shows the frequency with which each one or more atoms of a specific residue is found within 4 Å of a lipid hydrocarbon site (gray), lipid headgroup site (orange), or water molecule (blue), in system (A) ED1, (B) ED2, (C) ED4, (D) ED8, (E) ED16, (F) ED32, and (G) MD1.

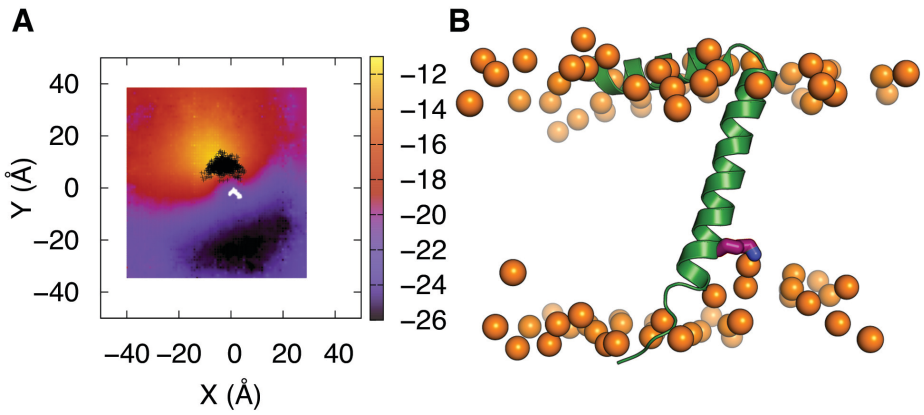

Figure 4.7 (A) The average *Z* coordinates of lipid phosphorus atoms in the lower leaflet in ED32. The bilayer center was aligned to *Z*=0, and then protein from each snapshot was aligned via translation on *XY*-plane. The Z coordinate values of phosphorus atoms in the lower leaflet are presented in gradient color (purple, red, and yellow). The black area shows where K40 charged side chain is positioned during the simulations. (B) One snapshot from ED32 showing the interactions between lipids and K40 (phosphorus atoms in orange sphere, K40 in purple stick, protein in cartoon presentation).

While the protein-lipid interactions change the geometry of the membrane, such interactions also contribute to stabilize the protein structure. For example, the periplasmic helix has hydrophobic residues (A7, A9, A10, F11, L14, and A16) on one side, and polar or charged residues (K8, D12, S13, Q15 and S17) on the other side. The frequent hydrophobic interactions of the former with the lipid carbon tails, as well as the hydrophilic interactions of the latter with the lipid headgroups and/or water, help determine the orientation of the periplasmic helix on the membrane surface, as consistently observed in the MD1 system without any restraints (Figure 4.6). Other than K40, K43, and K44, residues Y21 and Y24 also frequently interact with the

membrane, helping stabilize the transmembrane helix orientation (Figure 4.6). Because the present ED simulations explicitly include solvent molecules, such atomic-level interactions are also essential determinants of the protein structure (and orientation) in addition to the solid-state NMR restraints.

The relative orientation of the periplasmic and transmembrane helices determines the overall structure of the protein and, therefore, is functionally important. As shown in Figure 4-8 *A*, the fd coat protein maintains a well-defined L-shape with a helix-helix hinge angle of ~70° in all systems. Under strong solid-state NMR restraints, ED1 show a narrow distribution with variations of ~10°, while all the other ED systems having multiple replicas show broad distributions with variations of ~18°, which is similar to that of MD1. Figure 4-8 *B* shows the distributions of the crossing angle between the projections of principal helical axes of the periplasmic and transmembrane helices on the membrane surface. Similarly, all ED systems (except ED1) show broad distributions of crossing angles with variations of ~23° as in MD1, highlighting the conformational dynamics of fd coat protein.

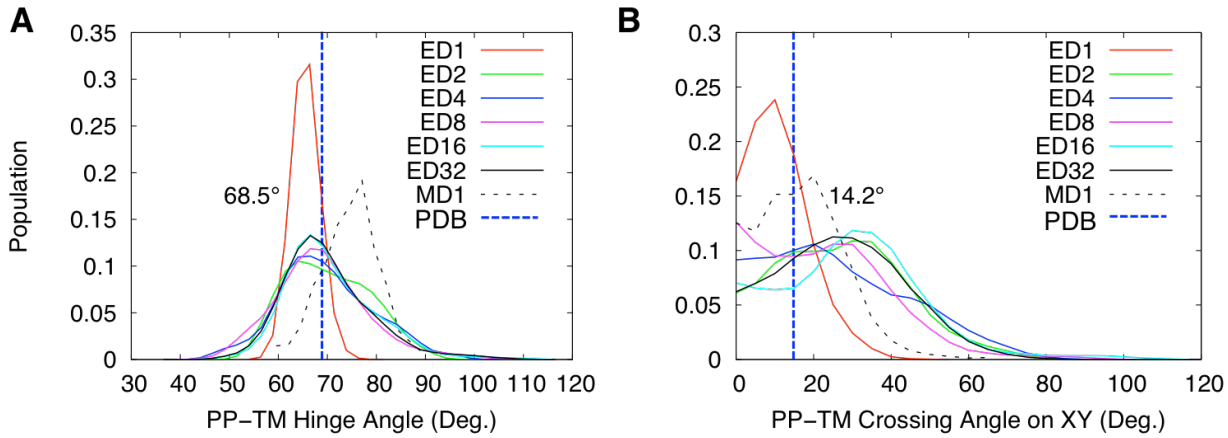

Figure 4.8 The distributions of (A) the (hinge) angles between the principal axes of the periplasmic and transmembrane helices and (B) the (crossing) angles between their principal axes projected on the *XY* plane.

4.4 Conclusions

The experimental solid-state NMR data are well satisfied in all NMR-restrained ED simulations and better represented by structure ensembles obtained with more replicas. By increasing the number of replicas, the strong bias imposed by a set of experimental restraint potentials diminishes, and inherent protein dynamics governed by the physical protein-lipid interactions becomes apparent. Therefore, ED simulations can simultaneously capture protein dynamics as well as an ensemble of conformations that satisfies the experimental observables.

For membrane-integrated fd coat protein, the average helical orientations are consistent in all simulation systems in explicit membranes and agree with the PDB structure that was calculated directly from the solid-state NMR data. Compared to systems with a single-conformer (ED1) or two replicas (ED2), which are under stronger influence of the NMR restraints, the multiple-replica ED systems (ED4, ED8, ED16, and ED32) show similar distribution patterns in helix orientations and compare well with those obtained from the standard, unrestrained MD simulation (MD1). This demonstrates that, with a proper ensemble size (more than two replicas in each ensemble), the NMR-restrained ED method can be very effective for generating stable dynamics trajectories of membrane proteins in explicit membranes. Specifically for fd coat protein, significant flexibility is observed in the C-terminal end of the transmembrane helix, which is likely to have functional importance.

The efficacy of ensemble-restrained ED has been previously demonstrated with the refinement of transmembrane helix structure in implicit membrane environments (26-28). Most implicit membrane models define a low-dielectric slab that effectively mimics the hydrophobic core of the membrane bilayer (30, 31). This static, low-dielectric slab can capture the general

effect of biological membrane, such as hydrophobic mismatch (between the length of a protein's hydrophobic transmembrane domain and the thickness of the bilayer hydrophobic core) (32-35), yet the computational cost of such system is substantially lower than that of an explicit membrane system. Thus, it has been widely used in many computational studies of membrane systems, including our earlier work on structure determination of single-pass transmembrane helices with solid-state NMR-restrained ED simulations (26, 27, 29). Nonetheless, despite many advantages in implicit membrane models, the undeformable nature of the hydrophobic slab is one of its most significant drawbacks. The biological membrane bilayer is fluid and membrane protein function is in part regulated by changes in lipid bilayer thickness and intrinsic lipid curvature (36). In particular, in the case of a hydrophobic mismatch, the bilayer adaptation involves local changes in lipid bilayer thickness (also known as local membrane thickening or thinning), and possibly changes in protein's transmembrane domain orientation (37-43). Most implicit membrane models are not able to provide such bilayer adaptation determined by detailed protein-lipid interactions. To address these drawbacks, several deformable implicit membrane models have been recently proposed (107-109), but their accuracy and general applicability have yet to be validated.

The present work shows that using an explicit membrane environment in solid-state NMR-restrained ED simulations has the important advantages of providing detailed atomic-level information about protein-lipid interactions, as observed for residues K40, K43 and K44, which are involved in DNA binding in the bacteriophage, and are anchored to the membrane surface in the membrane-bound form of the protein. Notably, K40 is observed to induce membrane curvature in the lower bilayer leaflet.

We conclude that solid-state NMR-restrained ED simulations of membrane-bound fd coat protein performed in explicit lipid bilayer membranes are very effective for determining protein structure within the membrane and extracting protein dynamics. This approach could be very useful for improving the accuracy and quality of membrane protein structures determined by solid-state NMR as it is increasingly applied to membrane proteins in nanodiscs.

5. CHARMM-GUI micelle builder for pure/mixed micelle and protein/micelle complex systems

5.1 Introduction

Micelles are frequently used in biochemical studies as a mimetic of cell membranes to solubilize integral membrane proteins (110-112). Understanding the effects of micelles on membrane proteins is significant in determining optimal detergent conditions for extraction, purification, and characterization of proteins (112-114). As a result, studying membrane proteins in micelles with atomic resolution has been of great interest to experimental and computational biophysicists (115-130). Molecular dynamics (MD) simulation is an attractive approach to study such systems because simulation can provide information about the structures, dynamics, interactions, and energetics of micelle systems at the atomic level, which is generally hard to obtain solely from experiments.

It is well known that detergents in water can readily self assemble to form micelles. Nevertheless, self-assembly simulation requires long relaxation time and a large amount of water and detergent molecules, which are computationally expensive. Recently, Pires et al. performed a 1 μs coarse-grained MD simulation to characterize the sodium dodecyl sulfate (SDS) micelle formation (115). 360 SDS and 90,000 water particles were included in the micelle system, whose relaxation occurred after about 250 ns. For the study of protein in micelles, Böckmann and Caflisch simulated spontaneous micelle formation of 1,2-dihexanoyl-*sn*-glycero-phosphocholine (DHPC) around OmpX, which included 125~188 detergents and required ~40 ns for equilibration (116). To avoid long relaxation times and large systems, most computational

studies are conducted on preassembled models (117, 125-127). However, building a preassembled protein/micelle complex is still challenging and requires considerable experience with simulation software. The main difficulty in building such a complex system arises from how to distribute detergents around a protein, especially when the protein is in an irregular shape, although detergents adjust themselves in concert with the embedded protein in the course of MD simulations (116, 117).

To simplify and automate the building process of protein/micelle preassembled complex systems for MD simulations, we have developed *Micelle Builder* (www.charmm-gui.org/input/micelle), a graphical user interface (GUI), available at the CHARMM-GUI website (78). Using *Micelle Builder*, a user can upload a membrane protein structure, or download from a database and choose one or multiple detergent types to generate a preassembled micelle around the protein. The complicated building process of protein/micelle systems can be dramatically simplified by automating the process with the intuitive user interface. In addition to the protein/micelle system, a user can also generate a detergent-only system without protein. All the necessary input files for building the system are provided to users, so that more complex modeling or adaptation of the protocol is possible, if necessary.

In this paper, we describe and illustrate the standardized building process of micelle and protein/micelle complex systems in CHARMM-GUI *Micelle Builder*. For illustration, we simulated four representative homogenous micelle systems composed of DHPC, *n*-dodecylphosphocholine (DPC or FC-12), *n*-tetradecylphosphocholine (TPC or FC-14), and SDS, respectively (Figure 5.1). The resulting models were compared with experimental data and previous simulation studies. In addition, we also used *Micelle Builder* to build and simulate the

major coat protein of Pf1 virus (Pf1 protein hereinafter) in three micelles with 50, 75, and 100 DHPC molecules. Pf1 protein is composed of a C-terminal single-pass transmembrane helix and a N-terminal periplasmic helix. Its structure was recently determined based on both solid-state and solution NMR observables, assuming the protein structure is highly similar in bilayers and in micelles (75). On the basis of these simulations, we explored the differences between protein-detergent interactions in micelles and protein-lipid interactions in bilayers.

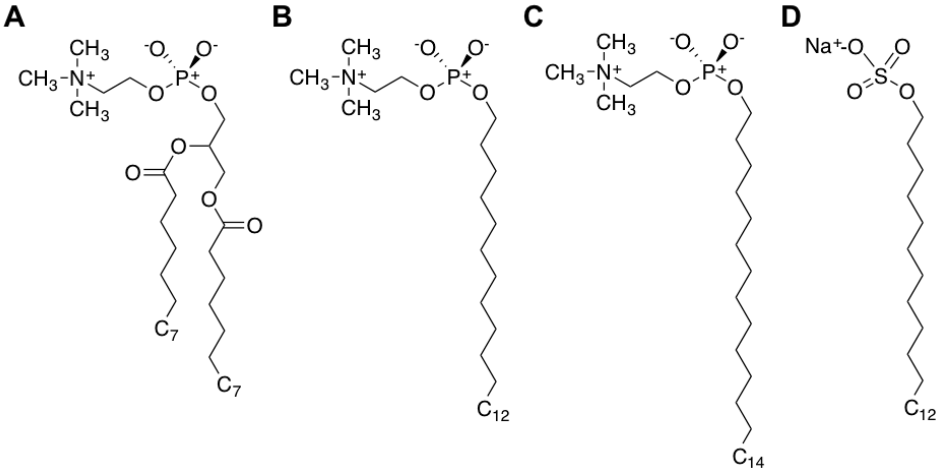

Figure 5.1 Chemical structures of (A) DHPC, (B) DPC, (C) TPC and (D) SDS.

5.2 Materials and methods

5.2.1 Protein/micelle complex building process in CHARMM-GUI Micelle Builder

The overall process to build a protein/micelle complex simulation system has been generalized and automated in six subsequent steps (Figure 5.2), similar to *Membrane (Bilayer) Builder* (49, 50) in the CHARMM-GUI website, www.charmm-gui.org (78). Each step is designed to incorporate user-specified parameters through a web browser and generate/execute CHARMM input files. The user can download and check the generated system in each step so that, if necessary, one can go back to the previous step and modify the options interactively. For a detergent-only micelle, the building procedure starts from the size-determination *STEP 3* (Figure 5.2). Individual input and output files or archives of all files are available for download and edit.

STEP 1: Read protein structure

The building process of a protein/micelle system starts with reading a protein structure into CHARMM. Users can upload their own protein structure, or specify a Protein Data Bank (PDB) (131) entry ID and a database to download the PDB file. Protein structures from the OPM (Orientations of Proteins in Membranes) database (132), http://opm.phar.umich.edu, are preoriented with respect to the membrane normal (the *Z*-axis by definition). Like *Membrane (Bilayer) Builder*, the usage of OPM database is recommended over the PDB (133) in *Micelle Builder* (see below).

STEP 2: Orient protein

Micelle Builder assumes that the initial protein structure is oriented along the *Z*-axis and the hydrophobic region of the protein is placed around $Z = 0$, which is the same as in *Membrane (Bilayer) Builder*. In the case that the protein is not properly oriented along the *Z*-axis, *Micelle Builder* provides a few options so that the user can place the protein appropriately in a micelle by reorienting protein in this step. These options are: aligning the protein principal axis along the *Z*-axis; aligning a vector between two user-specified Cα atoms along the *Z*-axis; translating the protein along the *Z*-axis; and rotating the protein around the *X*-axis. After alignment, *Micelle Builder* generates pore water (24), if specified, and calculates the protein cross-sectional area along the *Z*-axis, which provides information on the protein position with respect to the *Z*-axis (Figure 5.3).

STEP 3: Determine system size

This is a critical step to determine the system size, based on a number of user-specified parameters such as detergent type, number of detergents, water thickness around the micelle, and the size of the protein's hydrophobic region (Figure 5.4). In the current *Micelle Builder* setup, a cubic box is assumed for a micelle system, and four detergent molecules (DHPC, DPC, SDS, and TPC; Figure 5.1) are available. A user needs to select one or more detergent types and specify the number of selected detergent type(s) in the table under "Number of Detergent Molecules". For homogenous micelles, an initial micelle radius (r_m) is set to 21.4 Å (DHPC), 21.4 Å (DPC), 26.9 Å (TPC), and 22.6 Å (SDS) based on the pure micelle simulations in this study. As shown in Figure 5.4, the protein hydrophobic region (green area) is defined by its heavy atoms' *Z*

coordinates (Z_p) ranging from $-r_m$ to r_m. The protein radius (r_p) on the *XY* plane is set to the larger one between the average radius of hydrophobic region heavy atoms on *XY* and the average radius of hydrophobic region heavy atoms in -2.5 Å < *Z* < 2.5 Å on *XY*. Using this metric helps place detergent molecules for proteins with various geometries (see below). Then, the protein/micelle complex size on *XY* plane (L_{XY}) is set to $2r_m + 2r_p$. If the protein is completely buried in the micelle, i.e., $\max(|Z_p|)$ is less than r_m, then the complex size along the *Z* axis (L_Z) is set to $2r_m$. Otherwise, $L_Z = 2 \max(|Z_p|)$. The system size is the larger one of L_{XY} and L_Z plus twice of water thickness (t_w) from the complex.

After the system size is determined, *Micelle Builder* provides a detailed summary of the system size as well as a model micelle system with detergent-like pseudoatoms in the head group region (step3_packing.pdb). This model protein/micelle system can provide an idea about the initial detergent distribution, i.e., the detergent packing around the protein, because the pseudoatom positions are used to place detergent molecules in *STEP 4*. Briefly, a single-type pseudoatom with a radius of 5.4 Å is used to approximate the head groups of different detergents. Initially, each pseudoatom (for selected detergents) is randomly distributed on a torus surface (surface of green and light gray regions in Figure 5.4), but not inside protein pores (if a protein has a pore) via Monte Carlo simulations using a primitive model, i.e., van der Waals and 80-scaled Coulombic interactions. It is important for a user to visually examine the system packing with this structure file (step3_packing.pdb) because it is generally not known how many detergent molecules should be around a protein experimentally. Therefore, if the number of detergent molecules appears to be smaller or larger than expected, the user needs to go back to the previous step and adjust parameters, if necessary.

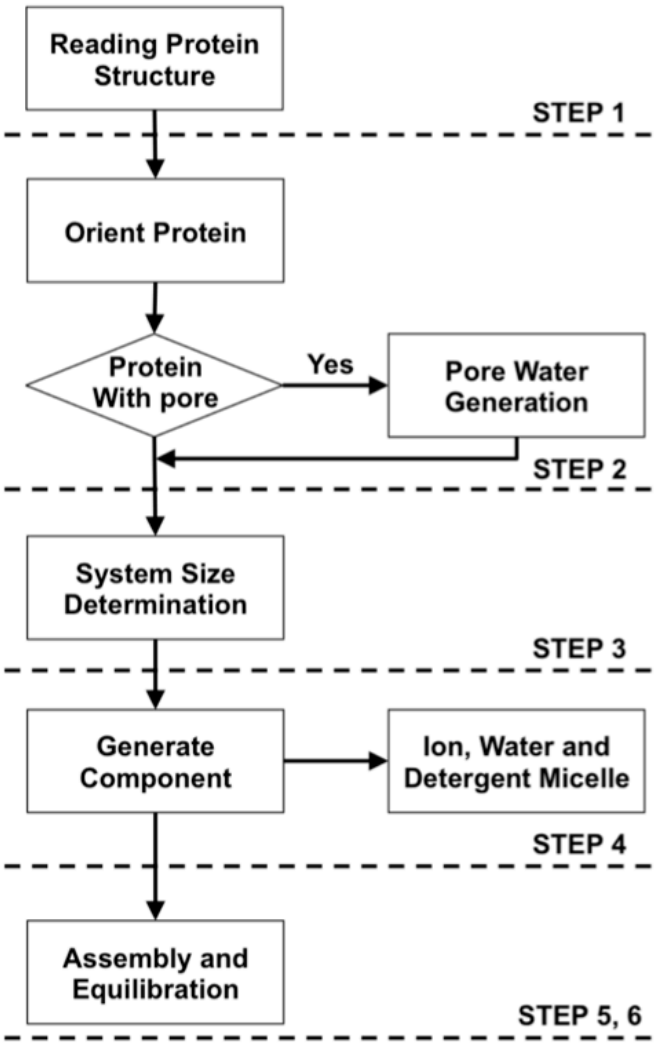

Figure 5.2 Overview of building process of protein/micelle complex simulation systems in CHARMM-GUI *Micelle Builder*.

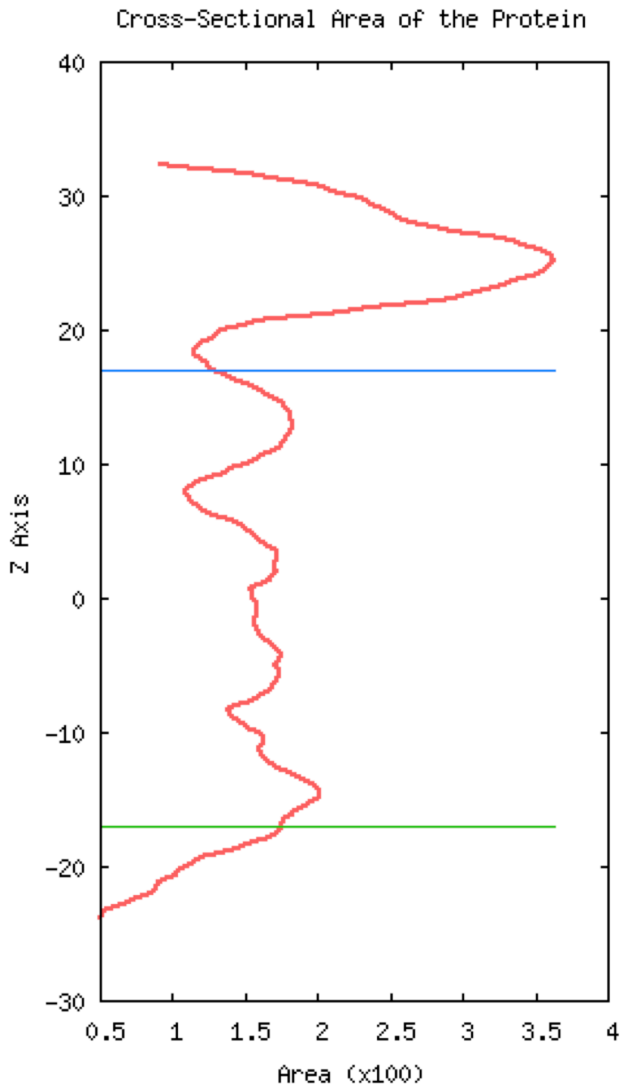

Figure 5.3 Cross-sectional area profile of Pf1 coat protein (PDB:2KSJ) along the Z-axis, generated by *Micelle Builder*. The region between the blue and green lines indicates the membrane region.

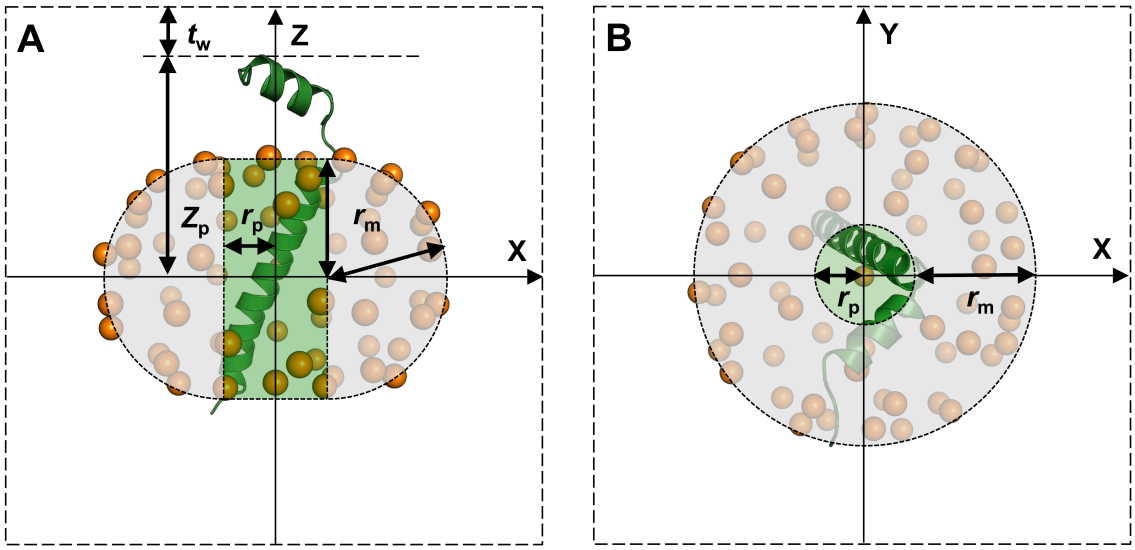

Figure 5.4 Schematic representation of a Pf1 coat protein/micelle system: (A) side and (B) top views. The protein/micelle complex size on *XY* plane (L_{XY}) is twice of the sum of protein radius (r_p) and micelle radius (r_m), i.e., $L_{XY} = 2r_m + 2r_p$. The complex size along the *Z*-axis (L_Z) is twice of the larger one between the maximum absolute protein *Z*-coordinate ($\max(|Z_p|)$) and r_m. The system size is the larger one of L_{XY} and L_Z plus twice of water thickness (t_w). The protein is shown in green cartoon representation, and the head group phosphorus atoms are shown in orange sphere. The green area indicates a protein hydrophobic region. Together with the green one, the light gray area represents the initial detergent-distributed region.

STEP 4: Build components

In this step, the system components such as a detergent micelle, bulk water, and ions are generated. The detergent micelle is built by a replacement method using the positions of detergent-like pseudoatoms in the head group region, determined in *STEP 3* (step3_packing.pdb). In this step, each pseudoatom is sequentially selected and its coordinate is used to place the head group of a randomly selected detergent molecule from the corresponding detergent structural library. A structural library of 2000 different conformations for each detergent type was generated from the homogenous micelle simulations in this study. With the head group sulfate/phosphorus atom position fixed, each detergent molecule is reoriented to ensure its carbon tail placed within the micelle hydrophobic core around the protein. To make the system neutral, *Micelle Builder* generates an appropriate number of ions, depending on the user-specified ion concentration. The initial configuration of ions is then determined around the generated protein/micelle complex via Monte Carlo simulations using the primitive model used in *STEP 3*.

STEP 5 and 6: Assemble components and equilibrate the system

Each component generated in the previous step is assembled in *STEP 5*. The user needs to examine the assembled system (step5_assembly.pdb) and verify whether the system has been built as intended. Because of the significant computing resources required for system equilibration, *Micelle Builder* does not provide the equilibrated structure. Instead, *Micelle Builder* provides the six consecutive CHARMM (134) and NAMD (135) input files for system equilibration and simulation production. As shown in Table 5.1, to assure gradual equilibration of the uncorrelated initial system, harmonic restraints are applied to the ions, heavy atoms of protein, and detergent sulfate/phosphorus atoms. These restraint forces are slowly reduced as the equilibration processes.

Table 5.1 Force constants[a] for positional harmonic restraints on each equilibration step.

STEP	Protein Backbone	Protein Sidechain	Detergent	Ion
6.1	10.0	5.0	2.5	10.0
6.2	5.0	2.5	0.0	0.0
6.3	2.5	1.0	0.0	0.0
6.4	1.0	0.5	0.0	0.0
6.5	0.5	0.1	0.0	0.0
6.6	0.1	0.0	0.0	0.0

[a]Force constants are in kcal/(mol·Å^2).

5.2.2 Detergent-only homogenous micelle systems

To illustrate the efficacy of *Micelle Builder*, we constructed and simulated detergent-only homogenous micelle systems (Figure 5.5 and Table 5.2) for four commonly used detergents: DHPC, DPC, TPC, and SDS. According to previous experimental and computational studies, we used 35 DHPC (118), 54 DPC (120), 108 TPC (119), and 62 SDS (121) molecules to characterize corresponding homogenous micelle properties. The micelle-only generation option in *Micelle Builder* was used to build these four systems. 0.15 M KCl was used for DHPC, DPC and TPC systems, while 0.15 M NaCl was used for SDS to match experiments (121-123). The molecular force field of these detergent molecules are based on the CHARMM36 (C36) lipid force field (55), and a modified TIP3P water model (56) was used. Each system was replicated and assigned with different velocities to generate five independent simulation systems, resulting in a total of 20 simulation systems. All calculations were performed in NPT (constant particle number, pressure, and temperature) ensembles (51) at 303.15 K using NAMD 2.8 (135), a parallel code designed for high-performance simulation of large biological macromolecule using the CHARMM force field (53, 55, 134). The particle mesh Ewald algorithm (59) was applied to calculate electrostatic forces, and the van der Waals interactions were smoothly switched off at 10-12 Å by a force-switching function (58). A time step of 2 fs was used in all simulations. After equilibration, a 100-ns production run was performed for each system. All average micelle properties were calculated over the five replicas using the last 40-ns simulations and presented with standard errors (Table 5.3).

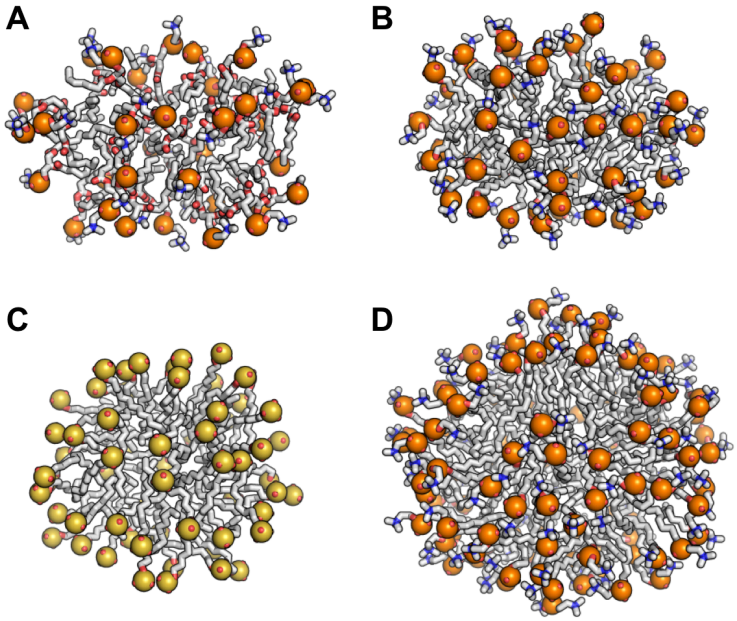

Figure 5.5 Representative structures of (A) DHPC, (B) DPC, (C) SDS and (D) TPC micelles. Water and ions are omitted for clarity. The head group phosphorus/sulfur atoms are shown in spheres and others in stick models.

Table 5.2 System information for micelle simulations.

System Name	# of Detergents	# of Water	# of Ions	# of Total Atoms
DHPC	35	11,250	20 K^+, 20 Cl^-	36,450
DPC	54	11,988	23 K^+, 23 Cl^-	39,304
TPC	108	17,505	38 K^+, 38 Cl^-	59,827
SDS	62	13,845	71 Na^+, 9 Cl^-	44,219
Pf1-DHPC50	50	13,242	43 K^+, 42 Cl^-	44,276
Pf1-DHPC75	75	19,236	62 K^+, 61 Cl^-	64,196
Pf1-DHPC100	100	26,872	85 K^+, 84 Cl^-	89,051

Table 5.3 Geometrical parameters[a] of detergent micelles.

Detergent	r_m (Å)	R_g (Å)	r_m[b] (Å)	R_g[c] (Å)	I_{max}/I_{min}	$<S>$ (Å^2)
DHPC	21.4±0.2	16.8±0.2	NR[d]	NR[d]	1.57±0.11	338.9±6.3
DPC	21.4±0.1	16.5±0.1	22.2(122), 19.5-24.5(124), 21-22(125)	17.4(125)	1.22±0.04	218.8±3.5
TPC	26.9±0.1	21.0±0.1	NR[d]	NR[d]	1.27±0.05	181.4±2.6
SDS	22.6±0.1	15.1±0.1	22.8(122), 22.3(123)	15.5±0.1(128), 15.7±0.2(126)	1.31±0.06	174.9±2.5

[a]Calculated values are the mean ± standard errors over the 5 independent simulation systems.

[b]Micelle radius (r_m) and [c]micelle radius of gyration were taken from the literature.

[d]NR, no reference was available.

5.2.3 Pf1 protein/micelle system

The average NMR structure of Pf1 protein (PDB:2KSJ) was chosen as the starting protein structure. Since the aggregation number of DHPC around Pf1 protein is not known, we used three different aggregation numbers (50, 75, and 100 DHPC molecules) to distribute DHPC molecules around the protein surface to mimic the NMR experimental conditions (75). The protein/micelle generation option in *Micelle Builder* was used to build these systems including 0.15 M KCl (Table Figure 5.1 and Figure 5.6). Each system was replicated and assigned with different velocities to generate three independent simulation systems. All calculations were performed in NPT ensembles at 303.15 K using NAMD 2.8 and the CHARMM force field. A time step of 2 fs was used and a 100-ns production run was performed for each system.

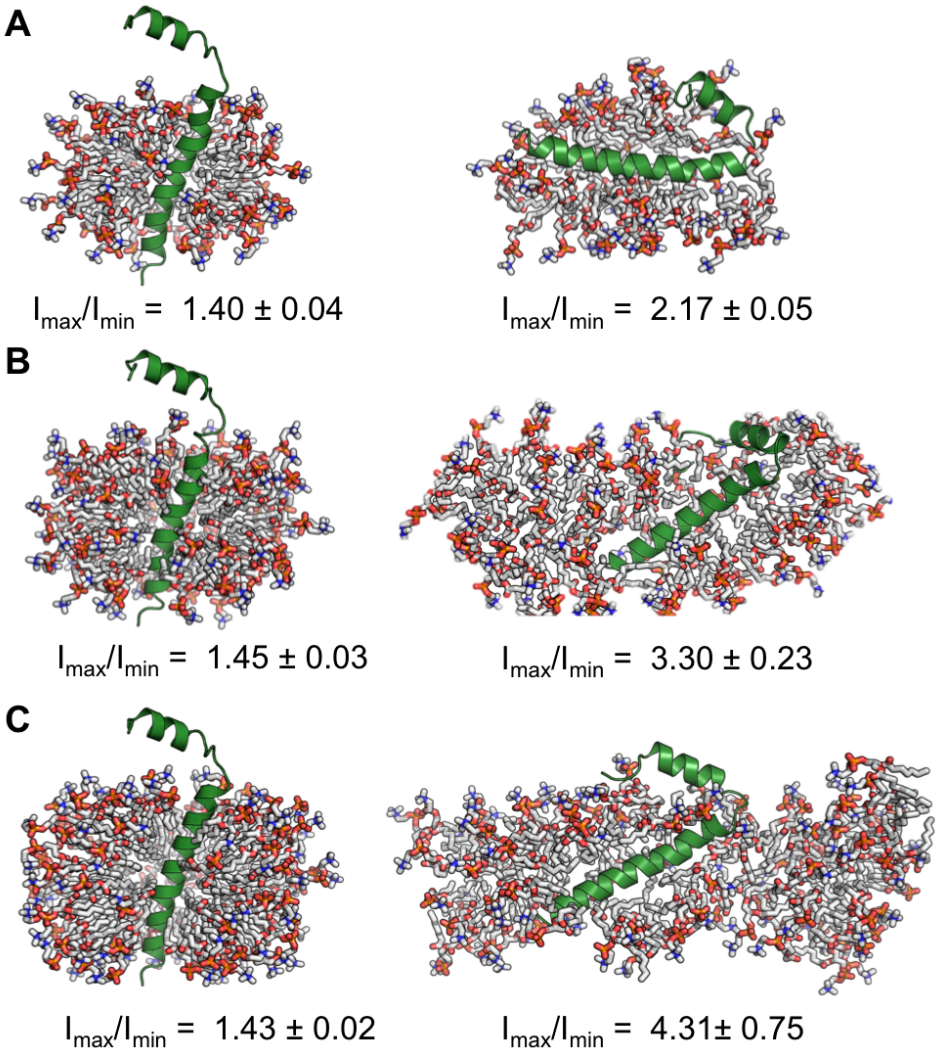

Figure 5.6 Initial (left) and equilibrated (right) structures of Pf1 coat protein in the DHPC micelles with (A) 50, (B) 75, and (C) 100 aggregation numbers. The ratio between maximum and minimum moment of inertia (I_{max}/I_{min}) for the initial (first 2 ns) and equilibrated (last 2 ns) structures are list below their corresponding snapshots. The protein is presented in green and the detergents are shown with sticks. Ions, water, and detergents blocking the view of the protein are omitted for clarity.

5.3 Results and discussion

5.3.1 Micelle size and shape

The most characteristic measurement of the micelle size is its radius. As shown in Figure 5.7, the micelle radius (r_m) is defined as the average distance (r_h) of the sulfur/phosphorus atoms of the head group from the micelle's center of mass (COM) plus the distance (d) at the first peak of the sulfur/phosphorus to water oxygen radial distribution function minus the radius of water (r_w = 1.4 Å); i.e., $r_m = r_h + d - r_w$ (127). This definition represents an effective micelle radius. In addition, the micelle radius of gyration was calculated by

$$R_g = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(|\mathbf{r}_i| - \langle r \rangle)^2} \tag{5-1}$$

where $\langle r \rangle$ is the mean distance of the heavy atoms ($\mathbf{r}_i$) from the micelle COM. R_g are stable as a function of time in all systems (Figure 5.8), indicating the convergence of the system, although less than three detergents dissociates from the micelles in a few simulation systems. As shown in Table 5.3, the average values of r_m and R_g in this work quantitatively agree with those observed in the previous experiments (122-124) and simulations (115, 125, 126), suggesting the efficacy of our models and simulation set up in *Micelle Builder*.

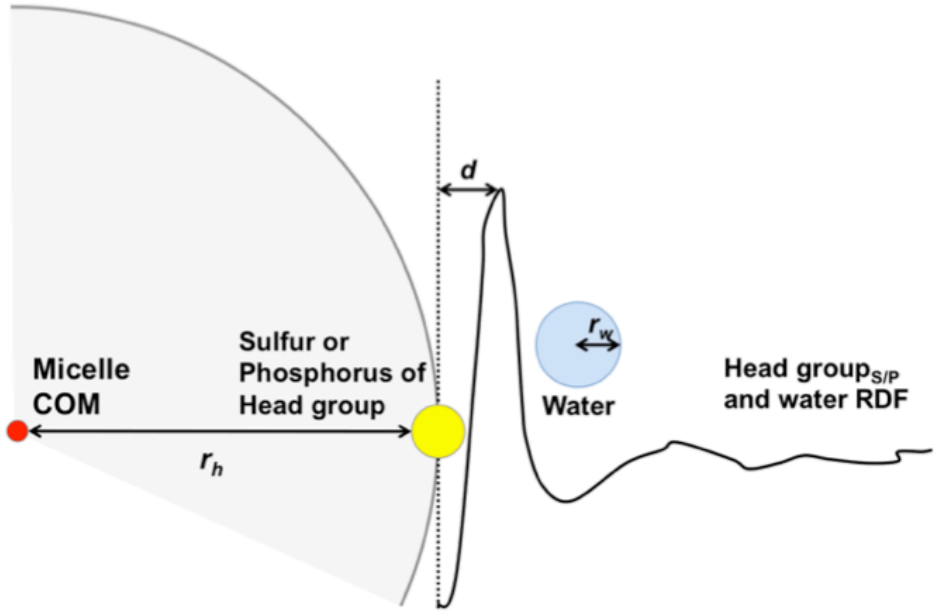

Figure 5.7 Schematic representation of the micelle radius (r_m) that is defined as the average distance (r_h) of the sulfur/phosphorus atoms of the head group from the micelle's center of mass plus the distance (d) at the first peak of the sulfur/phosphorus to water oxygen radial distribution function minus the radius of water (r_w = 1.4 Å); i.e., $r_m = r_h + d - r_w$.

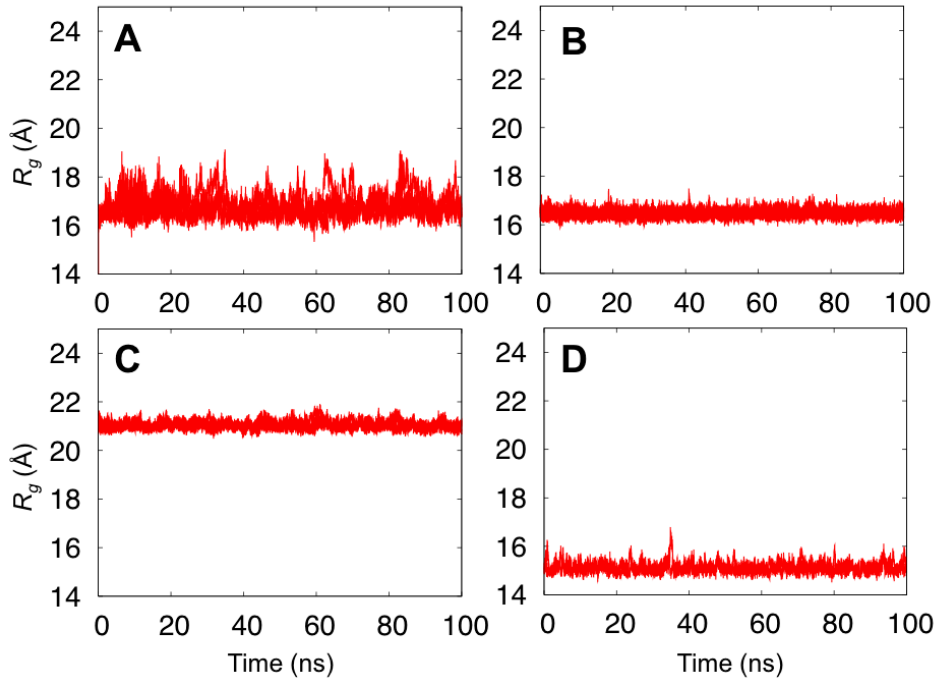

Figure 5.8 Micelle radius of gyration of all replicas as a function of time in (A) DHPC, (B) DPC, (C) TPC, and (D) SDS micelle systems.

The shape and stability of a micelle can be characterized by its moment of inertia (I) along the X, Y or Z-axis, defined by

$$I = \sum_{i=1}^{N} m_i r_i^2 \tag{5-2}$$

where m_i is the mass of an atom i and r_i is the distance of the atom from the axis. I_{max}/I_{min} is the ratio between the largest moment of inertia and the smallest one. For a perfect sphere, I_{max}/I_{min} = 1. As shown in Table 5.3, none of these micelles are perfectly spherical. The average values of I_{max}/I_{min} in DPC and SDS are 1.22 and 1.31, which are in the range reported for previous DPC simulations (1.2(117) – 1.24(125)) and SDS simulations (1.05(127) – 1.39(126)). Compared to the other micelles, DHPC is least spherical (I_{max}/I_{min} = 1.57) and such a prolate shape was also observed by small-angle X-ray scattering (136).

5.3.2 Micelle structure

Micelle structure can be analyzed in terms of radial densities of different components from the micelle COM (Figure 5.9). The shapes of the density distributions are in good agreement with those of previous simulations (117, 126, 128). The interior of micelle is void of solvent, as observed in experiments (129, 130) and other simulations (117, 126, 128). The solvent-detergent interface, i.e., the overlapping area of detergent carbon tail and water distributions in Figure 5.9, is broader in DHPC micelles due to its non-spherical nature described in the previous section. In SDS systems, some sodium ions stay close to the micelle surface because of electrostatic interactions with negatively charged SDS head groups, while potassium ions do not associate with the micelle in the other systems. As an additional analysis to characterize the structural properties of the micelle, the accessible molecular surface area (*SA*) was calculated as follows. All of the ions and water molecules were removed from the system, a probe molecule with a radius of 1.4 Å was rolled over the surface of the micelle, and then the contact area was summed to quantify the total *SA*. The accessible surface area per detergent (<*S*>) calculated by dividing the total *SA* by the number of detergents in each system is listed in Table 5.3. Bruce et al. reported a value of 176 $Å^2$ for SDS micelles (127), which is quite similar to our calculated value (174.9 ± 2.5 $Å^2$).

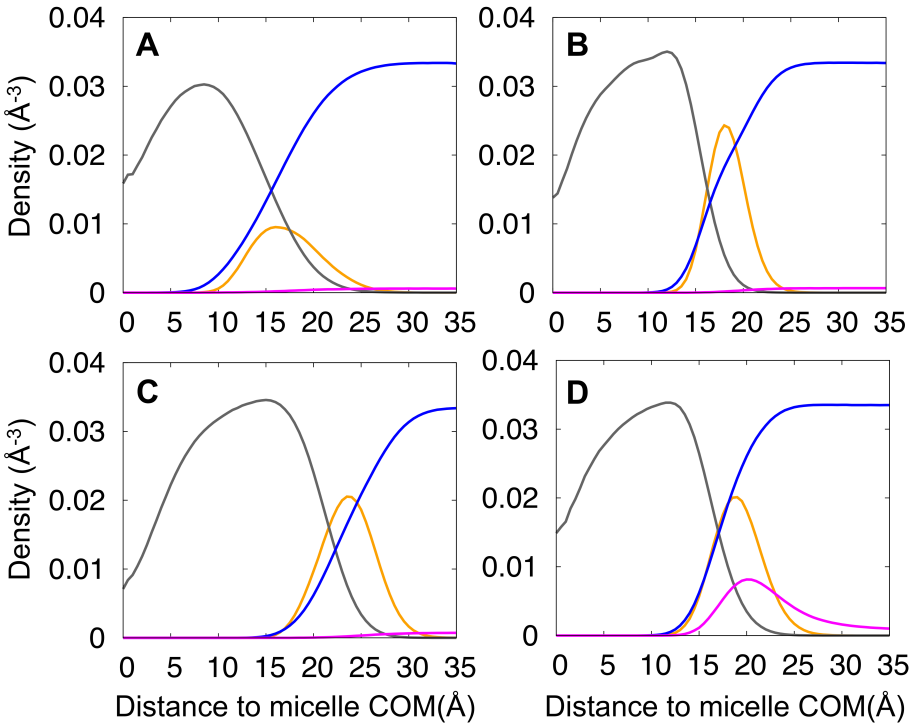

Figure 5.9 Density profiles for the specific components from the micelle center of mass: micelle carbon tails (*gray*), head groups (*orange*), water (*blue*) and positive ions (*magenta*) in (A) DHPC, (B) DPC, (C) TPC, (D) SDS micelles. These plots are produced by counting the number of selected atoms that are within 0.5 Å shells along the radial distance from the micelle center of mass. For clarity, the densities of micelle head groups and positive ions are multiplied by 10.

5.3.3 Distribution and mobility of detergents in protein/micelle systems

In addition to the Pf1 protein/micelle systems in Figure 5.6, we have built various protein/micelle complex structures to examine/validate the efficacy of the building process in *Micelle Builder*. As shown in Figure 5.10 *A-C*, *Micelle Builder* is able to build the protein/micelle initial models for various topologies of α-helical membrane proteins (137, 138), even for the conical shape of KcsA K^+ channel tetramer (139). As shown in Figure 5.10 *D-E*, β-barrel proteins such as OmpA (140) and OmpF (141) can be well solvated in micelles through *Micelle Builder*. In the case of OmpF, the option for generating pore water molecules in *STEP 2* was used to properly solvate the barrel interior. These initial structures can be relaxed with the rearrangement of detergent molecules in the course of simulation, as illustrated with Pf1 in the following paragraph.

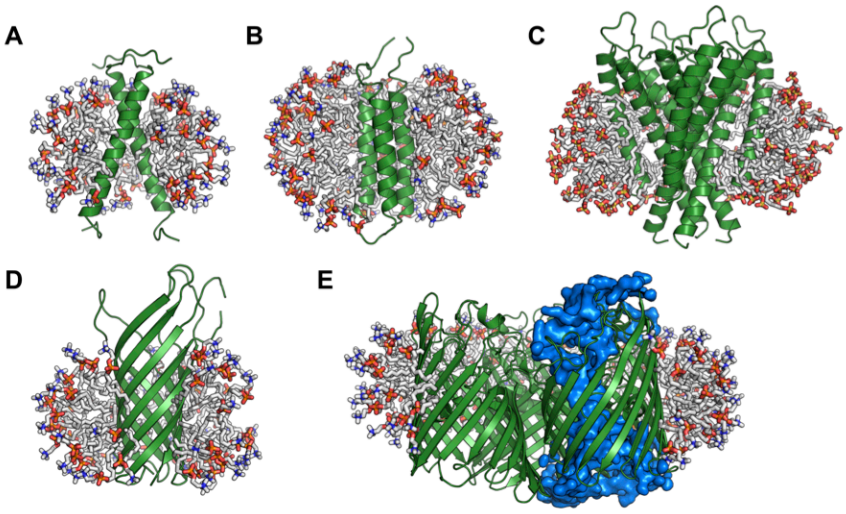

Figure 5.10 Protein/micelle complex structures for (A) GpA, (B) DAP12-NKG2C complex, (C) KcsA tetramer, (D) OmpA and (E) OmpF trimer. The protein is presented in green, the detergents are shown as gray sticks and pore water is shown in blue surface representation.

As a representative model, the Pf1 protein/micelle complex systems with different DHPC aggregation numbers (50, 75 and 100) are further characterized with the 100-ns simulations (Figure 5.6). Interestingly, even though the total number of detergents varies in the different systems (depending on its initial aggregation number), the number of detergents in direct contact with the protein becomes very similar within 10 ns (Figure 5.11 *A*-*B*): 36 ± 3 for Pf1/DHPC50, 38 ± 3 for Pf1/DHPC75 and 38 ± 4 for Pf1/DHPC100. Clearly, the torus of initial detergent molecules around the Pf1 adjusts its shape in concert with Pf1, and adapts less spherical geometry as more detergents involved in the complex during the simulation (Figure 5.6). In this process of redistribution, some detergents dissociate from the complex and come back afterwards (Figure 5.11 *C*-*E*). Compared to Pf1-DHPC50, Pf1-DHPC75 and Pf1-DHPC100 have larger numbers of dissociating detergents (Figure 5.11 *B* and Figure 5.12). Even with this variation, the current 100-ns simulations do not provide any information of possible aggregation numbers of DHPC around Pf1. Nonetheless, the simulations do provide a good estimation of the Pf1-contact detergent number and demonstrate that the smallest system of 50 DHPC is sufficient to solvate Pf1 protein. Therefore, we will focus on this Pf1-DHPC50 system for further analysis.

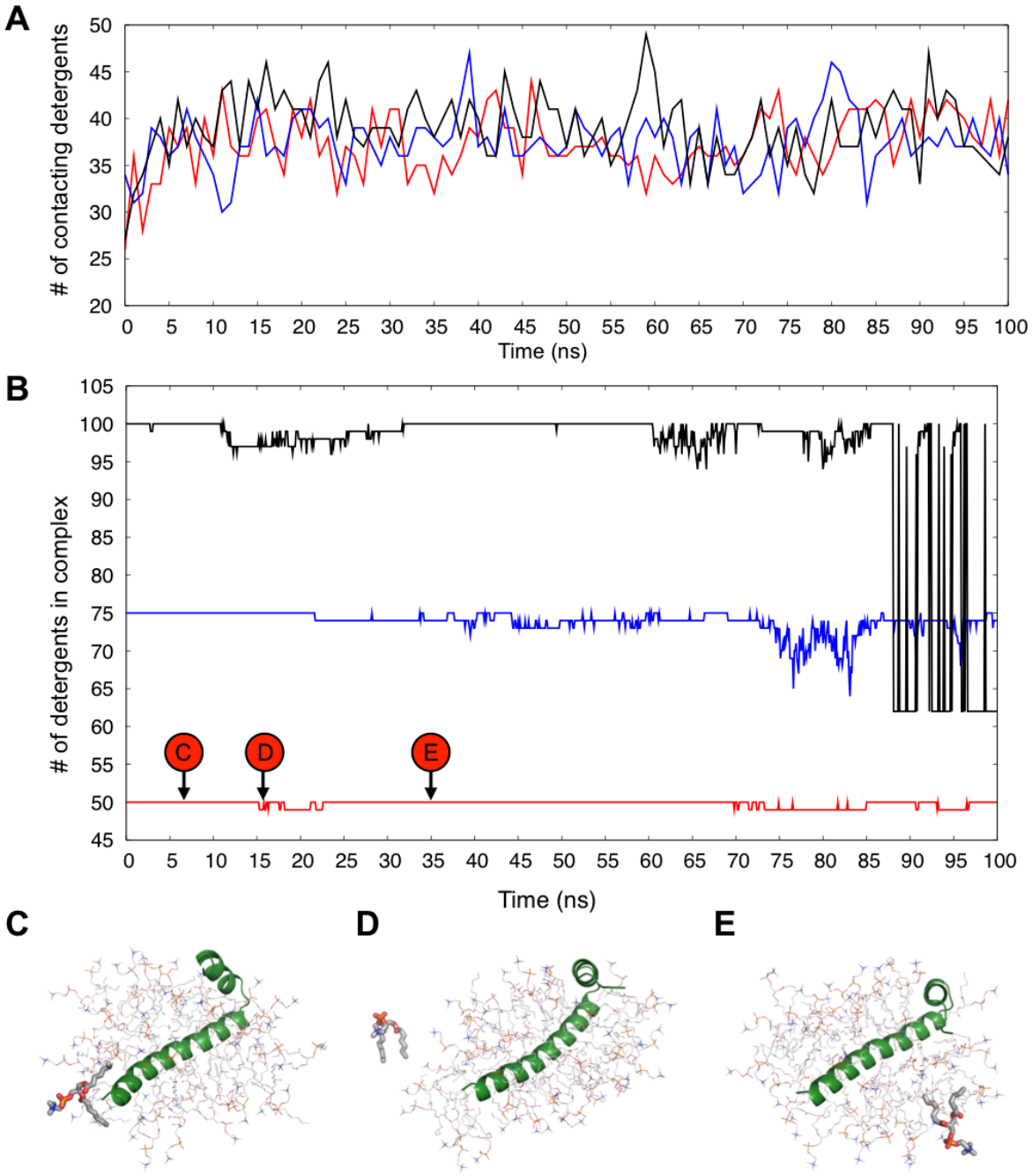

Figure 5.11 (A) The number of Pf1-contact detergents as a function of time in Pf1-DHPC50 (*red*), Pf1-DHPC75 (*blue*), and Pf1-DHPC100 (*black*) systems. A detergent is counted as a contact detergent when any of its heavy atoms is within 4 Å from the Pf1 protein heavy atoms.

(B) The number of detergents in the protein/micelle complex as a function of time in Pf1-DHPC50 (*red*), Pf1-DHPC75 (*blue*) and Pf1-DHPC100 (*black*) systems. A detergent is counted as a complex detergent when a detergent is in contact with the protein or the detergents embedding the protein. In Pf1-DHPC50, three snapshots were taken at (B) 7 ns, (C) 16 ns and (D) 35 ns to show the dynamics behavior of one detergent molecule leaving and coming back to the complex. The protein is presented in green, the detergents are shown as gray lines, and the dissociating detergents are in stick models. Ions and water blocking the view of the protein are omitted for clarity.

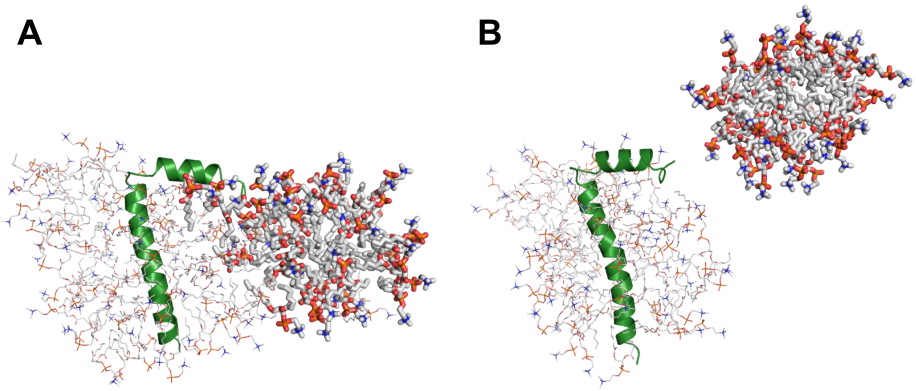

Figure 5.12 In Pf1-DHPC100, two snapshots were taken at (A) 87 ns and (B) 90 ns, showing a micelle with 36 DHPC molecules dissociating from the protein/micelle complex. The protein is presented in green, the detergents are shown as gray lines, and the dissociating detergents are in stick models. Ions and water blocking the view of the protein are omitted for clarity.

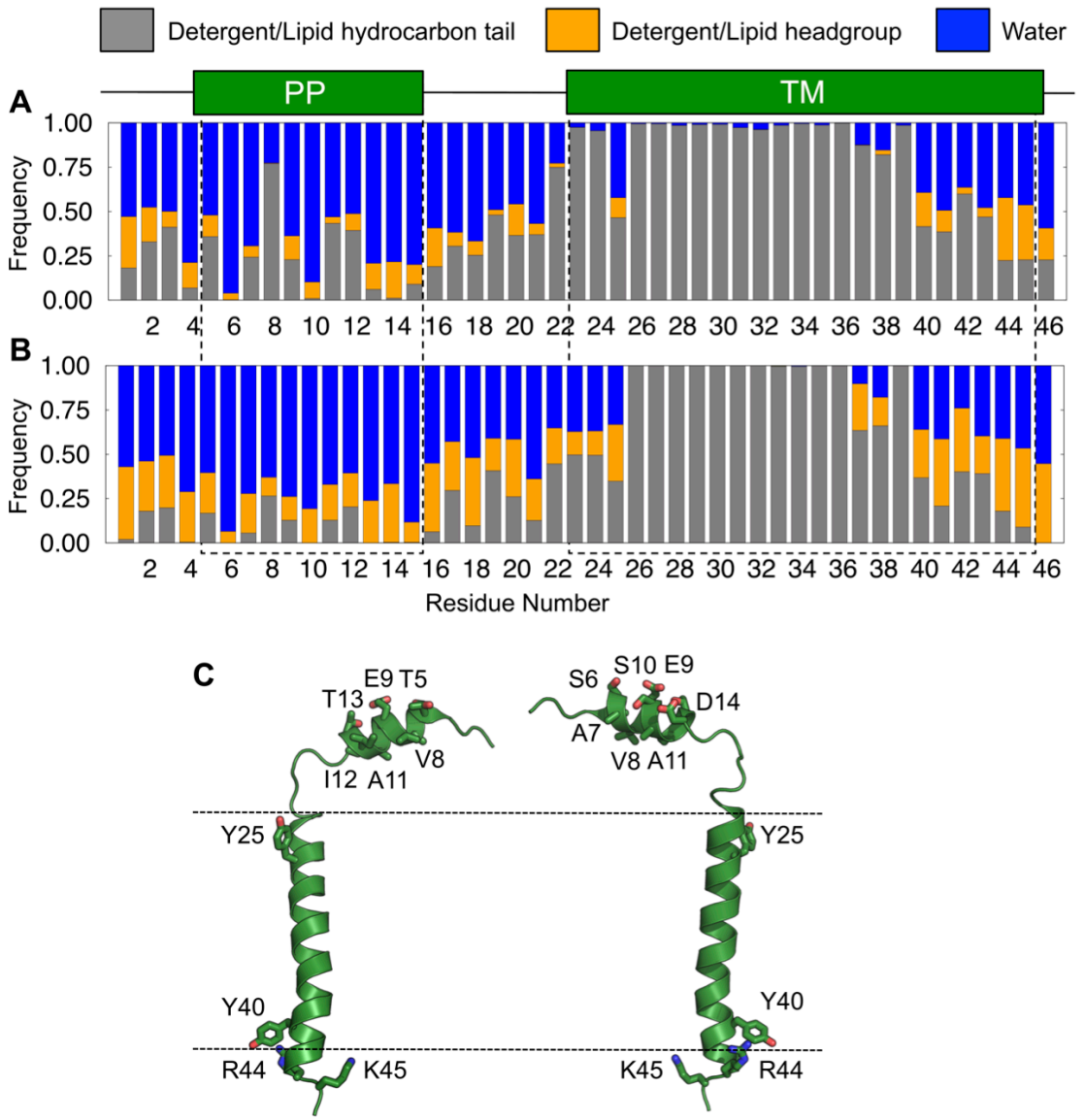

Figure 5.13 Interactions between Pf1 residues and various components in (A) Pf1-DHPC50 (micelle) and (B) Pf1-DOPC/DOPG (bilayer) systems. The graph shows the frequency with which any heavy atom of each residue is found within 4 Å of detergent or lipid carbon tails, detergent or lipid head groups, and water. The green rectangles indicate transmembrane (TM) and periplasmic (PP) helical residues. (C) Residues important for stabilizing transmemrbane and periplasmic helix orientations relative to the membrane.

5.3.4 Protein-detergents interactions and its effect on protein structure

The previous bilayer simulations of Pf1 protein show that protein-lipid interactions, especially the hydrophobic interactions involving residues A7, V8, A11, and I12 in the periplasmic helix, affect the orientation of the periplasmic helix (Figure 5.13) (142). In the micelle simulations, protein-solvent interacting patterns are quite similar to those in bilayers (Figure 5.13). However, more frequent and specific hydrophobic interactions between hydrophobic residues (A7, V8, A11 and I12) in the periplasmic helix and detergents were observed in the micelles compared to those between the residues and lipids in the bilayer system (Figure 5.14).

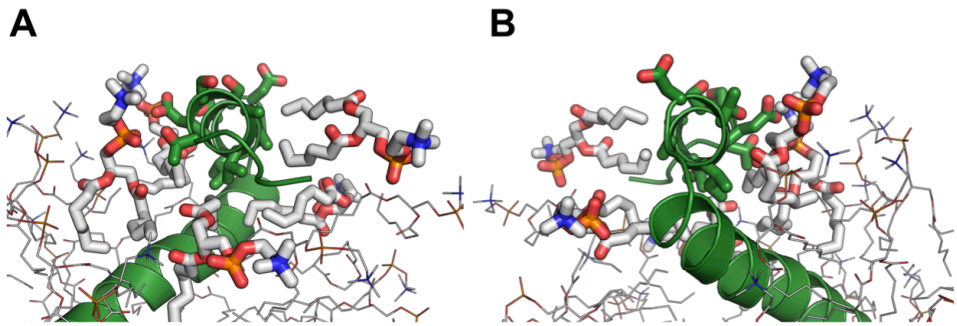

Figure 5.14 Snapshots showing the interactions between Pf1 coat protein periplasmic helix and detergent molecules in (A) front and (B) back views. The protein is shown in green. Residues in the periplasmic helix are shown as sticks. Detergents are shown as gray lines; those contacting with the periplasmic helix are drawn as sticks. Water molecules and lipids blocking the views of protein are omitted for clarity.

The Pf1 coat protein consists of two helices. Both transmembrane and periplasmic helices have root-mean-squared deviation (RMSD) from the average NMR structure (PDB:2KSJ) around 1 Å during the Pf1-DHPC50 simulations. However, as observed in the previous experiments (75) and bilayer simulations (142), the periplasmic helix orientation is highly flexible (Figure 5.15), which contributes to large overall RMSD of 7.3 ± 1.7 Å in Pf1-DHPC50. As mentioned above, the interactions between the periplasmic helix and detergents/lipids contribute to its orientation. Considering this important association as well as the distinctive geometries between micelle and bilayer, Pf1 protein (with the periplasmic helix) may present different structures in the different environments (Figure 5.16). In terms of the angle between transmembrane and periplasmic helices, the Pf1 protein essentially has no angle smaller than 60° in the bilayers, while it have large population with angles 20-60° in the micelles (Figure 5.15). Our simulation results suggest that, while micelles are effective mimetic for membrane bilayers and could hold the membrane protein transmembrane domain structure well (74), they may not hold other structural elements (and their orientations) existing on the outside or periphery of membrane bilayers well.

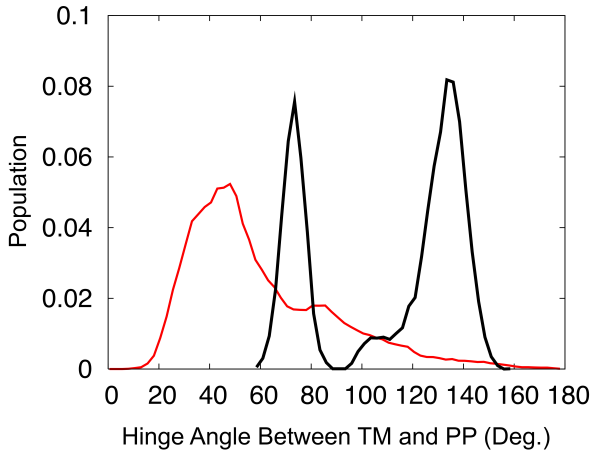

Figure 5.15 Distribution of angles between the transmembrane (TM) and periplasmic (PP) helices of Pf1 protein in the DHPC micelle (*red*) and DOPC/DOPG bilayer (*black*).

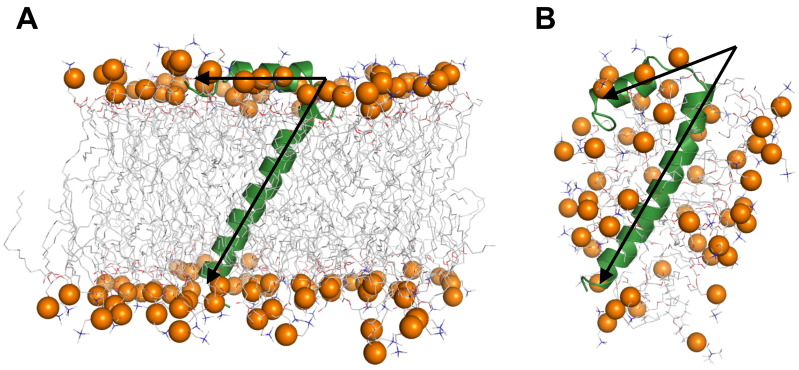

Figure 5.16 Snapshots of Pf1 protein in (A) Pf1-DOPC/DOPG bilayer and (B) Pf1-DHPC50 micelle systems. The black arrows indicate the principal axis of helix in the protein. The protein is presented in green; the detergents are shown as gray lines; the head group phosphorus atoms are shown in spheres. Ions and water blocking the views of protein are omitted for clarity.

5.4 Conclusions

We have described the generalized and automated procedure to build a protein/micelle complex system for MD simulation using *Micelle Builder* in CHARMM-GUI. Its efficacy was first illustrated by building and simulating four representative homogenous micelle systems of DHPC, DPC, TPC, and SDS. During the simulations, all micelle systems are stable. The size, shape, and structure of resulting micelles are in quantitative agreement with available experimental data and other simulation studies.

Instead of using a spherical model, the torus-shaped micelle building method in *Micelle Builder* was developed to generate protein/micelle complex structures. As illustrated with distinct topologies of α-helical and β-barrel membrane proteins, this method can be applied to various proteins with different geometries. As a representative, Pf1 protein/micelle systems were built, simulated, and characterized. Due to the larger mobility of detergents in a micelle, the detergents can interact with specific sites on the protein than their lipid counterparts. In the Pf1 case, the different geometries of micelle and bilayer could affect the structure and orientation of domains outside or at the periphery of the membrane in a membrane protein, and thus lead to different overall protein structures. It is our hope that CHARMM-GUI *Micelle Builder* is used for simulation studies for various protein/micelle systems to better understand the protein structure and dynamics in micelles as well as distribution of detergents and their dynamics around protein.

Reference

1. Krogh A, Larsson B, von Heijne G, & Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* 305(3):567-580.

2. Fujita M*, et al.* (2009) Hepatic uptake of gamma-butyrobetaine, a precursor of carnitine biosynthesis, in rats. *American journal of physiology. Gastrointestinal and liver physiology* 297(4):G681-686.

3. Ullrich A & Schlessinger J (1990) Signal transduction by receptors with tyrosine kinase activity. *Cell* 61(2):203-212.

4. Andre B (1995) An overview of membrane transport proteins in Saccharomyces cerevisiae. *Yeast* 11(16):1575-1611.

5. Freemont AJ & Hoyland JA (1996) Cell adhesion molecules. *Clinical molecular pathology* 49(6):M321-330.

6. Overington JP, Al-Lazikani B, & Hopkins AL (2006) How many drug targets are there? *Nature reviews. Drug discovery* 5(12):993-996.

7. Kim T & Im W (2010) Revisiting hydrophobic mismatch with free energy simulation studies of transmembrane helix tilt and rotation. *Biophysical journal* 99(1):175-183.

8. Torres J, Stevens TJ, & Samso M (2003) Membrane proteins: the 'Wild West' of structural biology. *Trends in biochemical sciences* 28(3):137-144.

9. Marassi FM & Opella SJ (1998) NMR structural studies of membrane proteins. *Current opinion in structural biology* 8(5):640-648.

10. Opella SJ (1997) NMR and membrane proteins. *Nature structural biology* 4 Suppl:845-848.

11. Anet FAL & Bourn AJR (1965) Nuclear Magnetic Resonance Spectral Assignments from Nuclear Overhauser Effects. *J Am Chem Soc* 87(22):5250-&.

12. Gayathri C, Bothnerby AA, Vanzijl PCM, & Maclean C (1982) Dipolar Magnetic-Field Effects in Nmr-Spectra of Liquids. *Chem Phys Lett* 87(2):192-196.

13. Tolman JR, Flanagan JM, Kennedy MA, & Prestegard JH (1995) Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information for Structure Determination in Solution. *P Natl Acad Sci USA* 92(20):9279-9283.

14. Clore GM, Starich MR, & Gronenborn AM (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J Am Chem Soc* 120(40):10571-10572.

15. Tjandra N & Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278(5340):1111-1114.

16. Brunner E (2001) Residual dipolar couplings in protein NMR. *Concept Magnetic Res* 13(4):238-259.

17. Rathinavelan T & Im W (2007) Explicit treatment of force contribution from alignment tensor using overdetermined linear equations and its application in NMR structure determination. *J Comput Chem* 28(11):1858-1864.

18. Opella SJ, Stewart PL, & Valentine KG (1987) Protein structure by solid-state NMR spectroscopy. *Quarterly reviews of biophysics* 19(1-2):7-49.

19. Lee J, Chen J, Brooks CL, 3rd, & Im W (2008) Application of solid-state NMR restraint potentials in membrane protein modeling. *Journal of magnetic resonance* 193(1):68-76.

20. Ketchem RR, Hu W, & Cross TA (1993) High-resolution conformation of gramicidin A in a lipid bilayer by solid-state NMR. *Science* 261(5127):1457-1460.

21. Im W, Jo S, & Kim T (2012) An ensemble dynamics approach to decipher solid-state NMR observables of membrane proteins. *Biochimica et biophysica acta* 1818(2):252-262.

22. Pluhackova K, Wassenaar TA, & Bockmann RA (2013) Molecular dynamics simulations of membrane proteins. *Methods in molecular biology* 1033:85-101.

23. Palmer JC & Gubbins KE (2012) Atomistic models for disordered nanoporous carbons using reactive force fields. *Micropor Mesopor Mat* 154:24-37.

24. Rathinavelan T & Im W (2008) A novel strategy to determine protein structures using exclusively residual dipolar coupling. *J Comput Chem* 29(10):1640-1649.

25. Roux B & Islam SM (2013) Restrained-ensemble molecular dynamics simulations based on distance histograms from double electron-electron resonance spectroscopy. *The journal of physical chemistry. B* 117(17):4733-4739.

26. Jo S & Im W (2011) Transmembrane Helix Orientation and Dynamics: Insights from Ensemble Dynamics with Solid-State NMR Observables. *Biophysical journal* 100(12):2913-2921.

27. Im W, Jo S, & Kim T (2012) An ensemble dynamics approach to decipher solid-state NMR observables of membrane proteins. *Bba-Biomembranes* 1818(2):252-262.

28. Kim T, Jo S, & Im W (2011) Solid-state NMR ensemble dynamics as a mediator between experiment and simulation. *Biophys. J.* 100(12):2922-2928.

29. Kim T, Jo S, & Im W (2011) Solid-State NMR Ensemble Dynamics as a Mediator between Experiment and Simulation. *Biophysical journal* 100(12):2922-2928.

30. Im W, Feig M, & Brooks CL, 3rd (2003) An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins. *Biophys. J.* 85(5):2900-2918.

31. Lazaridis T (2003) Effective energy function for proteins in lipid membranes. *Proteins* 52(2):176-192.

32. Im W & Brooks CL, 3rd (2005) Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations. *Proc Natl Acad Sci U S A* 102(19):6771-6776.

33. Killian JA & Nyholm TK (2006) Peptides in lipid bilayers: the power of simple models. *Curr Opin Struct Biol* 16(4):473-479.

34. Strandberg E, Esteban-Martin S, Ulrich AS, & Salgado J (2012) Hydrophobic mismatch of mobile transmembrane helices: Merging theory and experiments. *Biochim Biophys Acta* 1818(5):1242-1249.

35. Ulmschneider MB, Sansom MS, & Di Nola A (2005) Properties of integral membrane protein structures: derivation of an implicit membrane potential. *Proteins* 59(2):252-265.

36. Andersen OS & Koeppe RE (2007) Bilayer thickness and membrane protein function: An energetic perspective. *Annual review of biophysics and biomolecular structure* 36:107-130.

37. Kim T*, et al.* (2012) Influence of Hydrophobic Mismatch on Structures and Dynamics of Gramicidin A and Lipid Bilayers. *Biophysical journal* 102(7):1551-1560.

38. Dorairaj S & Allen TW (2007) On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. *P Natl Acad Sci USA* 104(12):4943-4948.

39. Kim T & Im W (2010) Revisiting Hydrophobic Mismatch with Free Energy Simulation Studies of Transmembrane Helix Tilt and Rotation. *Biophysical journal* 99(1):175-183.

40. Mori T, Ogushi F, & Sugita Y (2012) Analysis of lipid surface area in protein-membrane systems combining voronoi tessellation and monte carlo integration methods. *Journal of Computational Chemistry* 33(3):286-293.

41. Ru H, Kumar R, & Im W (2011) Membrane Tension, Lipid Adaptation, Conformational Changes, and Energetics in MscL Gating. *Biophysical journal* 101(3):671-679.

42. Rui H, Lee J, & Im W (2009) Comparative Molecular Dynamics Simulation Studies of Protegrin-1 Monomer and Dimer in Two Different Lipid Bilayers. *Biophysical journal* 97(3):787-795.

43. Rui H, Root KT, Lee J, Glover KJ, & Im W (2014) Probing the U-Shaped Conformation of Caveolin-1 in a Bilayer. *Biophysical journal* 106(6):1371-1380.

44. Kim HJ, Howell SC, Van Horn WD, Jeon YH, & Sanders CR (2009) Recent Advances in the Application of Solution NMR Spectroscopy to Multi-Span Integral Membrane Proteins. *Progress in nuclear magnetic resonance spectroscopy* 55(4):335-360.

45. Opella SJ & Marassi FM (2004) Structure determination of membrane proteins by NMR spectroscopy. *Chemical reviews* 104(8):3587-3606.

46. Call ME, Wucherpfennig KW, & Chou JJ (2010) The structural basis for intramembrane assembly of an activating immunoreceptor complex. *Nature immunology* 11(11):1023-1029.

47. Lanier LL (2008) Up on the tightrope: natural killer cell activation and inhibition. *Nature immunology* 9(5):495-502.

48. Call ME & Wucherpfennig KW (2007) Common themes in the assembly and architecture of activating immune receptors. *Nature reviews. Immunology* 7(11):841-850.

49. Jo S, Kim T, & Im W (2007) Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PloS one* 2(9):e880.

50. Jo S, Lim JB, Klauda JB, & Im W (2009) CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophysical journal* 97(1):50-58.

51. Feller SE, Zhang YH, & Pastor RW (1995) Computer-Simulation of Liquid/Liquid Interfaces .2. Surface-Tension Area Dependence of a Bilayer and Monolayer. *J Chem Phys* 103(23):10267-10276.

52. Brooks BR*, et al.* (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30(10):1545-1614.

53. MacKerell AD*, et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* 102(18):3586-3616.

54. Mackerell AD, Feig M, & Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25(11):1400-1415.

55. Klauda JB*, et al.* (2010) Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *Journal of Physical Chemistry B* 114(23):7830-7843.

56. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* 79(2):926-935.

57. Ryckaert JP, Ciccotti G, & Berendsen HJC (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J Comput Phys* 23(3):327-341.

58. Steinbach PJ & Brooks BR (1994) New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. *J Comput Chem* 15(7):667-683.

59. Essmann U*, et al.* (1995) A Smooth Particle Mesh Ewald Method. *J Chem Phys* 103(19):8577-8593.

60. Kim HJ, Howell SC, Van Horn WD, Jeon YH, & Sanders CR (2009) Recent advances in the application of solution NMR spectroscopy to multi-span integral membrane proteins. *Progress in nuclear magnetic resonance spectroscopy* 55(4):335-360.

61. Opella SJ & Marassi FM (2004) Structure determination of membrane proteins by NMR spectroscopy. *Chemical reviews* 104(8):3587-3606.

62. Tamm LK & Liang BY (2006) NMR of membrane proteins in solution. *Progress in nuclear magnetic resonance spectroscopy* 48(4):201-210.

63. Bonvin AMJJ & Brunger AT (1995) Conformational Variability of Solution Nuclear-Magnetic-Resonance Structures. *Journal of molecular biology* 250(1):80-93.

64. Lange OF*, et al.* (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320(5882):1471-1475.

65. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, & Vendruscolo M (2005) Simultaneous determination of protein structure and dynamics. *Nature* 433(7022):128-132.

66. Richter B, Gsponer J, Varnai P, Salvatella X, & Vendruscolo M (2007) The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J Biomol Nmr* 37(2):117-135.

67. Sharma M*, et al.* (2010) Insight into the Mechanism of the Influenza A Proton Channel from a Structure in a Lipid Bilayer. *Science* 330(6003):509-512.

68. Park SH*, et al.* (2012) Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* 491(7426):779-+.

69. Durr UHN, Yamamoto K, Im SC, Waskell L, & Ramamoorthy A (2007) Solid-state NMR reveals structural and dynamical properties of a membrane-anchored electron-carrier protein, cytochrome b(5). *J Am Chem Soc* 129(21):6670-+.

70. Hagn F, Etzkorn M, Raschle T, & Wagner G (2013) Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins. *J Am Chem Soc* 135(5):1919-1925.

71. Schwieters CD, Kuszewski JJ, Tjandra N, & Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *Journal of magnetic resonance* 160(1):65-73.

72. Dowhan W & Bogdanov M (2011) Lipid-protein interactions as determinants of membrane protein structure and function. *Biochem Soc T* 39:767-774.

73. Xia B, Tsui V, Case DA, Dyson HJ, & Wright PE (2002) Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water. *J Biomol Nmr* 22(4):317-331.

74. Cheng X & Im W (2012) NMR Observable-Based Structure Refinement of DAP12-NKG2C Activating Immunoreceptor Complex in Explicit Membranes. *Biophysical journal* 102(7):L27-L29.

75. Park SH, Marassi FM, Black D, & Opella SJ (2010) Structure and Dynamics of the Membrane-Bound Form of Pf1 Coat Protein: Implications of Structural Rearrangement for Virus Assembly. *Biophysical journal* 99(5):1465-1474.

76. Shi L*, et al.* (2009) A refinement protocol to determine structure, topology, and depth of insertion of membrane proteins using hybrid solution and solid-state NMR restraints. *J Biomol Nmr* 44(4):195-205.

77. Tian Y, Schwieters CD, Opella SJ, & Marassi FM (2012) AssignFit: A program for simultaneous assignment and structure refinement from solid-state NMR spectra. *Journal of magnetic resonance* 214:42-50.

78. Jo S, Kim T, Iyer VG, & Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* 29(11):1859-1865.

79. Dolan EA, Venable RM, Pastor RW, & Brooks BR (2002) Simulations of membranes and other interfacial systems using P2(1) and pc periodic boundary conditions. *Biophysical journal* 82(5):2317-2325.

80. Kandasamy SK*, et al.* (2009) Solid-state NMR and molecular dynamics simulations reveal the oligomeric ion-channels of TM2-GABA(A) stabilized by intermolecular hydrogen bonding. *Bba-Biomembranes* 1788(3):686-695.

81. Ramamoorthy A, Kandasamy SK, Lee DK, Kidambi S, & Larson RG (2007) Structure, topology, and tilt of cell-signaling peptides containing nuclear localization sequences in membrane bilayers determined by solid-state NMR and molecular dynamics simulation studies. *Biochemistry-Us* 46(4):965-975.

82. Bond PJ & Sansom MS (2003) Membrane protein dynamics versus environment: simulations of OmpA in a micelle and in a bilayer. *Journal of molecular biology* 329(5):1035-1053.

83. Cross TA, Sharma M, Yi M, & Zhou HX (2011) Influence of solubilizing environments on membrane protein structures. *Trends in biochemical sciences* 36(2):117-125.

84. Ramamoorthy A, Lee DK, Narasimhaswamy T, & Nanga RP (2010) Cholesterol reduces pardaxin's dynamics-a barrel-stave mechanism of membrane disruption investigated by solid-state NMR. *Biochimica et biophysica acta* 1798(2):223-227.

85. Cross TA, Ekanayake V, Paulino J, & Wright A (2014) Solid state NMR: The essential technology for helical membrane protein structural characterization. *J. Magn. Reson.* 239:100-109.

86. Wang S & Ladizhansky V (2014) Recent advances in magic angle spinning solid state NMR of membrane proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* 82(0):1-26.

87. Bonvin AM & Brunger AT (1995) Conformational variability of solution nuclear magnetic resonance structures. *J Mol Biol* 250(1):80-93.

88. Lange OF*, et al.* (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320(5882):1471-1475.

89. Richter B, Gsponer J, Varnai P, Salvatella X, & Vendruscolo M (2007) The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J Biomol Nmr* 37(2):117-135.

90. Cheng X, Jo S, Marassi FM, & Im W (2013) NMR-Based Simulation Studies of Pf1 Coat Protein in Explicit Membranes. *Biophysical journal* 105(3):691-698.

91. Marassi FM & Opella SJ (2003) Simultaneous assignment and structure determination of a membrane protein from NMR orientational restraints. *Protein Sci* 12(3):403-411.

92. Zeri AC, Mesleh MF, Nevzorov AA, & Opella SJ (2003) Structure of the coat protein in fd filamentous bacteriophage particles determined by solid-state NMR spectroscopy. *P Natl Acad Sci USA* 100(11):6458-6463.

93. Jo S & Im W (2011) Transmembrane helix orientation and dynamics: insights from ensemble dynamics with solid-state NMR observables. *Biophys. J.* 100(12):2913-2921.

94. Jo S, Lim JB, Klauda JB, & Im W (2009) CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophysical journal* 97(1):50-58.

95. Shaw DE*, et al.* (2008) Anton, a special-purpose machine for molecular dynamics simulation. *Commun Acm* 51(7):91-97.

96. Martyna GJ, Tobias DJ, & Klein ML (1994) Constant-Pressure Molecular-Dynamics Algorithms. *J Chem Phys* 101(5):4177-4189.

97. Martyna GJ, Tuckerman ME, Tobias DJ, & Klein ML (1996) Explicit reversible integrators for extended systems dynamics. *Mol Phys* 87(5):1117-1157.

98. Shan YB, Klepeis JL, Eastwood MP, Dror RO, & Shaw DE (2005) Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. *J Chem Phys* 122(5).

99. Tian Y, Schwieters CD, Opella SJ, & Marassi FM (2012) AssignFit: a program for simultaneous assignment and structure refinement from solid-state NMR spectra. *J Magn Reson* 214(1):42-50.

100. Pitera JW & Chodera JD (2012) On the Use of Experimental Observations to Bias Simulated Ensembles. *J Chem Theory Comput* 8(10):3445-3451.

101. Islam SM, Stein RA, Mchaourab HS, & Roux B (2013) Structural Refinement from Restrained-Ensemble Simulations Based on EPR/DEER Data: Application to T4 Lysozyme. *Journal of Physical Chemistry B* 117(17):4740-4754.

102. Chen J, Won HS, Im W, Dyson HJ, & Brooks CL, 3rd (2005) Generation of native-like protein structures from limited NMR data, modern force fields and advanced conformational sampling. *J Biomol Nmr* 31(1):59-64.

103. Xia B, Tsui V, Case DA, Dyson HJ, & Wright PE (2002) Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water. *J Biomol Nmr* 22(4):317-331.

104. Cheng X & Im W (2012) NMR observable-based structure refinement of DAP12-NKG2C activating immunoreceptor complex in explicit membranes. *Biophys. J.* 102(7):L27-29.

105. Chen J, Im W, & Brooks CL, 3rd (2004) Refinement of NMR structures using implicit solvent and advanced sampling techniques. *J Am Chem Soc* 126(49):16038-16047.

106. Almeida FC & Opella SJ (1997) fd coat protein structure in membrane environments: structural dynamics of the loop between the hydrophobic trans-membrane helix and the amphipathic in-plane helix. *J Mol Biol* 270(3):481-495.

107. Choe S, Hecht KA, & Grabe M (2008) A continuum method for determining membrane protein insertion energies and the problem of charged residues. *J Gen Physiol* 131(6):563-573.

108. Panahi A & Feig M (2013) Dynamic Heterogeneous Dielectric Generalized Born (DHDGB): An Implicit Membrane Model with a Dynamically Varying Bilayer Thickness. *J Chem Theory Comput* 9(3):1709-1719.

109. Zhou YC, Lu BZ, & Gorfe AA (2010) Continuum electromechanical modeling of protein-membrane interactions. *Phys Rev E* 82(4).

110. Columbus L*, et al.* (2006) Expression, purification, and characterization of Thermotoga maritima membrane proteins for structure determination. *Protein Sci* 15(5):961-975.

111. Eshaghi S*, et al.* (2005) An efficient strategy for high-throughput expression screening of recombinant integral membrane proteins. *Protein Sci* 14(3):676-683.

112. Sanders CR & Sonnichsen F (2006) Solution NMR of membrane proteins: practice and challenges. *Magn Reson Chem* 44 Spec No:S24-40.

113. Berger BW, Garcia RY, Lenhoff AM, Kaler EW, & Robinson CR (2005) Relating surfactant properties to activity and solubilization of the human adenosine a3 receptor. *Biophys J* 89(1):452-464.

114. Wiener MC (2004) A pedestrian guide to membrane protein crystallization. *Methods* 34(3):364-372.

115. Pires JM, de Moura AF, & Freitas LCG (2012) Investigating the Spontaneous Formation of Sds Micelle in Aqueous Solution Using a Coarse-Grained Force Field. *Quim Nova* 35(5):978-U352.

116. Bockmann RA & Caflisch A (2005) Spontaneous formation of detergent micelles around the outer membrane protein OmpX. *Biophys J* 88(5):3191-3204.

117. Bond PJ & Sansom MSP (2003) Membrane protein dynamics versus environment: Simulations of OmpA in a micelle and in a bilayer. *J Mol Biol* 329(5):1035-1053.

118. Columbus L*, et al.* (2009) Mixing and matching detergents for membrane protein NMR structure determination. *J Am Chem Soc* 131(21):7320-7326.

119. Strop P & Brunger AT (2005) Refractive index-based determination of detergent concentration and its application to the study of membrane proteins. *Protein Sci* 14(8):2207-2211.

120. Tieleman DP, van der Spoel D, & Berendsen HJC (2000) Molecular dynamics simulations of dodecylphosphocholine micelles at three different aggregate sizes: Micellar structure and chain relaxation. *J Phys Chem B* 104(27):6380-6388.

121. Turro NJ & Yekta A (1978) Luminescent Probes for Detergent Solutions - Simple Procedure for Determination of Mean Aggregation Number of Micelles. *J Am Chem Soc* 100(18):5951-5952.

122. Gao X & Wong TC (1998) Studies of the binding and structure of adrenocorticotropin peptides in membrane mimics by NMR spectroscopy and pulsed-field gradient diffusion. *Biophys J* 74(4):1871-1888.

123. Itri R & Amaral LQ (1991) Distance Distribution Function of Sodium Dodecyl-Sulfate Micelles by X-Ray-Scattering. *J Phys Chem-Us* 95(1):423-427.

124. Lauterwein J, Bosch C, Brown LR, & Wuthrich K (1979) Physicochemical Studies of the Protein-Lipid Interactions in Melittin-Containing Micelles. *Biochim Biophys Acta* 556(2):244-264.

125. Wymore T, Gao XF, & Wong TC (1999) Molecular dynamics simulation of the structure and dynamics of a dodecylphosphocholine micelle in aqueous solution. *J Mol Struct* 485:195-210.

126. Jalili S & Akhavan M (2009) A coarse-grained molecular dynamics simulation of a sodium dodecyl sulfate micelle in aqueous solution. *Colloid Surface A* 352(1-3):99-102.

127. Bruce CD, Berkowitz ML, Perera L, & Forbes MDE (2002) Molecular dynamics simulation of sodium dodecyl sulfate micelle in water: Micellar structural characteristics and counterion distribution. *J Phys Chem B* 106(15):3788-3793.

128. Gao J, Ge W, Hu GH, & Li JH (2005) From homogeneous dispersion to micelles - A molecular dynamics simulation on the compromise of the hydrophilic and hydrophobic effects of sodium dodecyl sulfate in aqueous solution. *Langmuir* 21(11):5223-5229.

129. Bendedouch D, Chen SH, & Koehler WC (1983) Structure of Ionic Micelles from Small-Angle Neutron-Scattering. *J Phys Chem-Us* 87(1):153-159.

130. Jones RRM, Maldonado R, Szajdzinskapietek E, & Kevan L (1986) Electron-Spin Echo Modulation of Doxylstearic Acid Probes of the Surface and Internal Structure of Lithium Dodecyl-Sulfate Micelles - Comparison with Sodium Dodecyl-Sulfate and Tetramethylammonium Dodecyl-Sulfate Micelles. *J Phys Chem-Us* 90(6):1126-1129.

131. Berman HM*, et al.* (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235-242.

132. Lomize MA, Lomize AL, Pogozheva ID, & Mosberg HI (2006) OPM: orientations of proteins in membranes database. *Bioinformatics* 22(5):623-625.

133. Berman HM*, et al.* (2002) The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* 58(Pt 6 No 1):899-907.

134. Brooks BR*, et al.* (2009) CHARMM: The Biomolecular Simulation Program. *J Comput Chem* 30(10):1545-1614.

135. Phillips JC*, et al.* (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781-1802.

136. Lipfert J, Columbus L, Chu VB, Lesley SA, & Doniach S (2007) Size and shape of detergent micelles determined by small-angle x-ray scattering. *J Phys Chem B* 111(43):12427-12438.

137. MacKenzie KR, Prestegard JH, & Engelman DM (1997) A transmembrane helix dimer: structure and implications. *Science* 276(5309):131-133.

138. Call ME, Wucherpfennig KW, & Chou JJ (2010) The structural basis for intramembrane assembly of an activating immunoreceptor complex. *Nature Immunology* 11(11):1023-U1073.

139. Zhou Y & MacKinnon R (2003) The occupancy of ions in the K+ selectivity filter: charge balance and coupling of ion binding to a protein conformational change underlie high conduction rates. *Journal of molecular biology* 333(5):965-975.

140. Pautsch A & Schulz GE (1998) Structure of the outer membrane protein A transmembrane domain. *Nature structural biology* 5(11):1013-1017.

141. Efremov RG & Sazanov LA (2012) Structure of Escherichia coli OmpF porin from lipidic mesophase. *Journal of structural biology* 178(3):311-318.

142. Cheng X, Jo S, Marassi FM, & Im W (2013 (submitted)) NMR-Based simulation studies of pf1 coat protein in explicit membranes. *Biophys J*.

Appendix

A list of publications

1. **X. Cheng**, S. Jo, Y. Qi, F.M. Marassi, and W. Im. Solid-State NMR based Ensemble Dynamics Simulations of Membrane Protein in Explicit Membranes. *(In press)*

2. S. Jo, **X. Cheng**, S.M. Islam, L. Huang, H. Rui, A. Zhu, H.S. Lee, Y. Qi, W. Han, K. Vanommeslaeghe, A.D. MacKerell, Jr., B. Roux, and W. Im CHARMM-GUI PDB Manipulator for Advanced Modeling and Simulations of Proteins Containing Nonstandard Residues. *Adv. Protein Chem. Struct. Biol. 96:235-265 (2014)*

3. E.L. Wu*, **X. Cheng***, S. Jo*, H. Rui*, K.C. Song*, E.M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R.M. Venable, J.B. Klauda, and W. Im CHARMM-GUI Membrane Builder Toward Realistic Biological Membrane Simulations. *J. Comput. Chem. 35:1997-2004 (2014) [cover]*

4. Y. Qi, **X. Cheng**, W. Han, S. Jo, K. Schulten, and W. Im CHARMM-GUI PACE CG Builder for Solution, Micelle, and Bilayer Coarse-Grained Simulations. *J. Chem. Inf. Model. 54:1003-1009 (2014)*

5. **X. Cheng**, S. Jo, H.S. Lee, J.B. Klauda, and W. Im CHARMM-GUI Micelle Builder for Pure/Mixed Micelle and Protein/Micelle Complex Systems. *J. Chem. Inf. Model. 53:2171-2180 (2013)*

6. **X. Cheng**, S. Jo, F.M. Marassi, and W. Im NMR-Based Simulation Studies of Pf1 Coat Protein in Explicit Membranes. *Biophys. J. 105:691-698 (2013)*

7. **X. Cheng** and W. Im NMR Observable-Based Structure Refinement of DAP12-NKG2C Activating Immunoreceptor Complex in Explicit Membranes. *Biophys. J. 102:L27-L29 (2012)*

* Authors contribute equally to the work.