

APPLICATIONS AND DEVELOPMENTS OF AN IMPLICIT MEMBRANE MODEL IN
SIMULATIONS OF MEMBRANE BOUND PEPTIDES

By

Afra Panahi

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry-Doctor of Philosophy

2013

ABSTRACT

APPLICATIONS AND DEVELOPMENTS OF AN IMPLICIT MEMBRANE MODEL IN SIMULATIONS OF MEMBRANE BOUND PEPTIDES

By

Afra Panahi

Heterogeneous dielectric generalized Born (HDGB) model has been applied to study the conformational sampling of the influenza fusion peptide (IFP) in the context of membrane bilayers. The dominant structure of the IFP shows a tight break in the middle region and interfacial insertion in the membrane bilayer. Different protonation states of titratable residues in the IFP sequence were also simulated and the results show that the pH has little effect on the configuration of the peptide in the membrane.

The conformational sampling of the Ebola fusion peptide (EFP) in the membrane environment and water was also studied using the implicit membrane/solvent model. The EFP shows to be strongly sensitive to its neighboring residues and it adopts different conformation in their presence. It also tends to have more extended structure in water compared to the membrane environment. However, its overall structure in the membrane bilayer was observed to be helical in the middle region and extended in the N and C terminal parts.

The overall comparison between the results obtained from simulations and experimental observations for these peptides proves that the HDGB model is capable of capturing the peptide-membrane interactions. However, in all current implicit membrane models, including HDGB, the static structure of the membrane bilayer and lack of fluctuations in the local thickness and curvature of the membrane does not allow the bilayer to respond to the presence of a charged particle or any hydrophobic mismatch in general. This would in particular, lead to overestimation

of insertion energy barriers of charged particles. In the improved implicit membrane model called dynamic heterogeneous dielectric generalized Born (DHDGB), these problems are being addressed by introducing extra degrees of freedom to represent the membrane fluctuations. The effect of membrane dynamics on different test systems has also been studied and the results are being presented and discussed.

ACKNOWLEDGEMENTS

I consider my years at Michigan State University to be the most exciting and fruitful era of my life. During these years I was a part of a thriving research group which helped me grow and prepared me for my next step. However, my great motivation to start this path was my beloved mother who is the best teacher I have ever had. I also consider myself as one of the most fortunate students in my hometown for having the best mentors during my high school years.

I am also grateful to my professors in Sharif University of Technology during my undergraduate and graduate years especially Dr. Gholamabbas Parsafar and Dr. Mehdi Jalali for showing me the joy in learning and to my wonderful friends who kept me motivated even through the most difficult times.

My deepest gratitude goes to my PhD advisor, Dr. Michael Feig. His constant search for new and improved methods taught me that there is no limit to learning and I should always keep my mind open for better alternatives to what might occur to me as the “best” solution. What I have accomplished during my PhD wouldn’t have been possible without his endless support and patience.

I also want to use this opportunity to thank Feig’s lab members Dr. Sean Law, Dr. Maryam Sayadi, Vahid Mirjalili, Dr. Shayantani Mukherji, Dr. Sirinivasa Gopal, Dr. Alex Predeus, Dr. Liang Fang, Dr. Beibei Wang, Dr. Parimal Kar, Dr. Munika Sharma, Seref Gul and Adam Jaskierny for providing a friendly environment where questions were always welcome and for our insightful discussions. Especial thanks go to my committee members, Dr. Robert Cukier, Dr. David Weliky and Dr. Daniel Jones for their critical comments on my dissertation and on my

work and to my dear friends Dr. Erica Vogel and Dr. Yan Sun for the helpful discussions which expanded my knowledge on experimental techniques.

I would like to express my sincere gratitude to my great friends in East Lansing especially Dr. Atefeh Garzan and Dr. Elham Attaran for the wonderful time we had together as roommates.

Finally, I wouldn't have been able to start and continue the long journey that I am about to complete if I didn't feel the unconditional support and love of my family every step of the way. Their words and their kind faces constantly remind me that they are always close to me no matter how far we are.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
1 INTRODUCTION	1
1.1 Background	2
1.2 MD: theory	3
1.3 Implicit solvent models	7
1.3.1 Electrostatic energy based on the Poisson-Boltzmann equation	9
1.3.2 Non-polar contribution based on solvent-accessible surface area	16
1.3.3 Limitations of implicit solvent models	17
1.4 Implicit treatment of the membrane	17
1.4.1 Implicit membrane model; extension to GBMV2	18
1.4.2 Elastic implicit membrane model	21
1.5 Conformational sampling of fusion peptides in the context of the membrane	24
1.5.1 Influenza fusion peptide in the context of the membrane bilayer	25
1.5.2 Ebola fusion peptide	28
2 CONFORMATIONAL SAMPLING OF INFLUENZA FUSION PEPTIDE IN MEMBRANE BILAYERS AS A FUNCTION OF TERMINI AND PROTONATION	30
2.1 Abstract	31
2.2 Introduction	32
2.3 Computational methodology	35
2.4 Results and discussion	42
2.4.1 Conformational sampling as a function of different termini	42
2.4.2 Conformational sampling as a result of different protonation states	54
2.5 Conclusions	61
2.6 Acknowledgements	63
3 EFFECT OF FLANKING RESIDUES ON THE CONFORMATIONAL SAMPLING OF THE INTERNAL FUSION PEPTIDE FROM EBOLA VIRUS	64
3.1 Abstract	65
3.2 Introduction	66
3.3 Methods	69
3.4 Results	73
3.4.1 Conformational sampling of EFP, W8A, and EFP-X in membrane bilayer	73
3.4.2 EFP in aqueous solvent	77
3.4.3 Helical propensity in EFP	78
3.5 Discussion and conclusion	80
3.6 Acknowledgements	84

4	DYNAMIC HETEROGENEOUS DIELECTRIC GENERALIZED BORN (DHDGB): AN IMPLICIT MEMBRANE MODEL WITH A DYNAMICALLY VARYING BILAYER THICKNESS.....	85
4.1	Abstract.....	86
4.2	Introduction.....	87
4.3	Theory.....	90
4.3.1	Membrane deformation energy (ΔG_{mem}).....	91
4.3.2	Electrostatic solvation free energy (ΔG_{GB}).....	98
4.3.3	Non-polar solvation free energy (ΔG_{np}).....	102
4.4	Computational Methodology.....	104
4.4.1	Membrane Insertion of Small Molecules.....	106
4.4.2	Arg-Containing TM Helix.....	106
4.4.3	WALP23 Peptide.....	107
4.4.4	Gramicidin A.....	108
4.5	Results and Discussion.....	109
4.5.1	Amino acid side chain analogs.....	109
4.5.2	TM Helix with a Central Arg(+).....	110
4.5.3	MD simulation of WALP23 in a membrane bilayer.....	114
4.5.4	MD simulation of gramicidin A in a membrane bilayer.....	118
4.6	Conclusions.....	120
4.7	Acknowledgements.....	122
5	CONCLUSIONS AND PRESPECTIVES.....	123
5.1	Study of the membrane fusion process and the role of fusion peptides.....	124
5.2	Study of membrane deformation.....	126
6	SUPPLEMENTAL CHAPTER: CONFORMATIONAL SAMPLING OF THE DIMERIC FORM OF THE INFLUENZA FUSION PEPTIDE IN THE CONTEXT OF THE MEMBRANE BILAYER.....	130
6.1	Introduction.....	131
6.2	Computational methodology.....	132
6.3	Results and discussion.....	133
6.4	Conclusions.....	137
	REFERENCES.....	138

LIST OF TABLES

Table 2-1: Simulated IFP systems with different C- and N-termini and protonation states of acidic residues.	36
Table 2-2: Average COFM, membrane insertion angles, and helicity for IFP peptides with different termini from structures at 300K with COFM values of less than 25 Å. The tag in NH_3^+ /tag/CONH ₂ is excluded in calculating the average values reported here. Statistical errors are given in parentheses.	43
Table 3-1: Overview of EFP simulations.....	71
Table 3-2: Population of inserted peptides, average center of mass (COM) insertion depth, insertion angles, and helicity fractions from simulated EFP structures at 300 K for membrane-associated conformations (defined as having COM insertion depths of less than 25 Å). Analysis of EFP-X was only applied to the core EFP sequence. Standard deviations are reported in parentheses. The standard errors based on block averaging are about 0.3 Å for COM, 10.5 degrees for insertion angles and 0.03 for helical fractions.	75

LIST OF FIGURES

Figure 1-1: Cartoon representation of a deformed membrane.....	23
Figure 2-1: Dielectric and non-polar profiles for periodic membrane along z axis..	38
Figure 2-2: PMFs for simulations of IFP with different termini	46
Figure 2-3: Dominant structures for simulations of IFP with different termini	48
Figure 2-4: Average fraction of helical conformations as a function of residue for IFP with different termini.	51
Figure 2-5: PMFs from WHAM for simulations of IFP with $\text{NH}_3^+/\text{tag}/\text{CONH}_2$	53
Figure 2-6: Dominant structures for simulations of IFP with $\text{NH}_3^+/\text{tag}/\text{CONH}_2$	54
Figure 2-7: PMFs for simulations of IFP with different protonation states.....	56
Figure 2-8: Dominant structures for simulations of IFP with different protonation states.....	58
Figure 2-9: pK_a shifts of E11, E15 and D19 as a function of insertion depth.....	60
Figure 2-10: PMFs for mixed ensembles.....	62
Figure 3-1: Schematic representation of Ebola fusion protein in fusogenic state..	67
Figure 3-2: Left panel: PMFs for EFP, W8A, and EFP-X, Right panel: Dominant structures for simulations with EFP, W8A and EFP-X from clustering.	74
Figure 3-3: Left panel: Potential of mean force (PMF) of EFP-AQ. Right panel: Dominant structures for simulations with EFP-AQ from clustering.	78
Figure 3-4: Average fraction of helical conformations as a function of residue for EFP, W8A, and EFP-X.....	82
Figure 4-1: A schematic representation of an arbitrary deformation around a cylindrical solute inclusion.....	94
Figure 4-2: Schematic illustration of the dielectric profile calculation for a symmetric deformation of -15 Å in the top leaflet..	99
Figure 4-3: Optimized coefficients for the dielectric profile calculations.....	105

Figure 4-4: Amino acid analog insertion free energy profiles.....	111
Figure 4-5: Membrane insertion energies for a poly-leucine TM helix with a central Arg(+)..	112
Figure 4-6: Variation of deformations of top $\{S_1, \dots, S_5\}$ and bottom $\{S_6, \dots, S_{10}\}$ leaflets.....	113
Figure 4-7: Probability distribution of the membrane insertion angle of WALP23 in implicit membrane environment	116
Figure 4-8: Probability distribution of the membrane insertion angle for the gA dimer in the membrane bilayer.....	120
Figure 6-1: initial monomeric structures of IFP at pH=5.0(1bn) and pH=7.4(1bo)[115].	132
Figure 6-2: Top panel: 2D PMF of dimers obtained form TRMD of 1bn monomers. Bottom panel: the most visited structures in the ensemble of dimers of 1bn..	134
Figure 6-3: Top panel: 2D PMF of dimers obtained form TRMD of 1bo monomers. Bottom panel: the most visited structures in the ensemble of dimers of 1bo.	136

1 INTRODUCTION

1.1 Background

The first high resolution crystal structure of a protein published in 1958[1] contributed to the initial view of proteins as rigid bodies rather than dynamic systems[2, 3] that can sample different conformations. However, the first molecular dynamic simulation (MD) of a small protein challenged this idea by capturing motions of the atoms during the course of the simulation. Although this simulation was only several picoseconds long, it could clearly show that proteins are dynamic macromolecules with conformational fluctuations[4]. Today, along with the experimental techniques, molecular dynamics is a strong tool for studying dynamics and energies of biological systems [3, 5].

Molecular dynamics simulation refers to one of the families of computational methods in which a potential energy function is used to model macromolecules and investigate their energetics and evolution with time[6]. In the simplified model of the system used in MD, molecules are considered as assemblies of spherical atoms connected via springs that represent the chemical bonds. The interactions between atoms of molecules are modeled via mathematical functions that represent the fluctuations around the equilibrium values obtained from spectroscopy. Once a potential energy function, that accounts for all atomic interactions, is established, the system dynamics can be studied by numerical integration of Newton's equation of motion at each step of the simulation[6]. Therefore, detailed structural fluctuations [7] that may not be captured in experimental techniques can be observed in MD simulations. In nuclear magnetic resonance (NMR) for instance, measured properties are averages of possible states that the system can sample in the given experimental timescale and conditions. While in X-ray crystal structures on the other hand the system is trapped in a crystal structure which is only one of the numerous possible states that can be adopted by the system and not necessarily the most

favorable configuration. In MD, however, one can obtain average properties without losing the structural details. This capability has put MD among the most widely applied methods to investigate structure and dynamics of biological macromolecules such as proteins, DNA and RNA [3, 8-10].

The first MD simulation was performed in the early 50's on a system of simple gasses. In this simulation, atoms were considered as hard spheres and all the interactions were modeled by perfect collisions[11]. Later, more sophisticated potential energy functions were developed which were suitable for investigation of biological systems. The first protein studied using MD was the small bovine pancreatic trypsin inhibitor (BPTI). This simulation was only 9.2 ps long and was performed in vacuum[4]. Today, due to theoretical and methodological improvements in the field of molecular dynamics along with immense increase in computing power, MD simulations can now be performed for as long as sub-millisecond time scale [12, 13] and on systems containing as many as 10^6 - 10^8 atoms[14-17].

In the following chapter, first a brief introduction on the theory of MD will be presented. Then Poisson-Boltzmann (PB) equation as the benchmark of electrostatic calculations will be introduced and some implicit solvent models stemming from PB will be discussed. Finally an implicit membrane model along with some applications and limitations will be presented.

1.2 MD: theory

The majority of computational techniques that are based on classical dynamics depend upon a suitable potential energy function that describes the interactions between the components of the system. Normally the energy functions are composed of different terms each parameterized based on experimental measurements or quantum mechanical calculations of small molecules or

fragments of macromolecules. These parameters are assumed to be transferable to large molecules such as proteins. The set of functions accompanied by the appropriate set of parameters is called a force field. Different energy terms of the force fields are assumed to be independent of each other. For instance, bond lengths and bond angles are contributing to two independent energy components and variation in one will not affect the parameters of the other. This property insures the additivity in different force fields[6].

Some of the most common force fields that are employed to study biological systems are CHARMM [18, 19], OPLS[20], GROMOS[21, 22] and AMBER[23].

In the following, the CHARMM force field [18, 19] is briefly introduced, and later throughout this dissertation its application in studying several membrane bound peptides will be presented. The CHARMM force field is composed of discrete terms to model different interactions within the system. These terms are normally divided into bonded and non-bonded interactions. For bonded interactions, simple two-body, three-body, and four-body terms are considered to describe bonds, angles, and dihedrals, respectively. According to the CHARMM force field [18, 19] potential energy can be formulated using Eq. 1.2.1

$$\begin{aligned}
 V(\vec{R}) = & \sum_{bonds} K_d (d - d_0)^2 + \sum_{Uray-Bradley} K_{UB} (S - S_0)^2 + \\
 & \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi (1 + \cos(n\chi - \delta)) + \\
 & \sum_{impropers} K_\phi (\phi - \phi_0)^2 + \sum_{nonbonded} \left\{ \epsilon_{ij} \left[\left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{12} - \left(\frac{R_{ij}^{min}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon_l r_{ij}} \right\}
 \end{aligned} \tag{1.2.1}$$

In this equation, K_d , K_{UB} , K_θ , K_χ , and K_ϕ are the bond length, Urey-Bradley (1-3 bond length), bond angle, dihedral angle, and improper dihedral angle force constants respectively. Also d , S , θ , χ , and ϕ are bond length, Urey-Bradley, bond angle, dihedral angle, and improper dihedral values of any random configuration of the system, and the zero subscripts denote the equilibrium values of each of those[6]. In addition to the terms mentioned above that are common among different force fields, CHARMM utilizes a further a grid base cross-correlation term called CMAP to modify the torsion angles of the protein back bone atoms[24].

The non-bonded interactions are calculated in the last summation in Eq. 1.2.1. It incorporates Lennard-Jones potential which models hard-sphere repulsion and van der Waals attractions and Coulombic interactions. For Lennard-Jones potential, the first term of the summation, ϵ_{ij} 's are the geometrical averages of ϵ_{ii} and ϵ_{jj} which are the Lennard-Jones well depth of two identical interacting atoms and r_{ij} is the distance between atoms i and j . Finally R_{ij}^{min} is the arithmetic average of R_{ii}^{min} and R_{jj}^{min} which are the distances at which Lennard Jones potentials of two identical interacting atoms are zero. The Coulombic interactions are modeled using q_i and q_j for partial atomic charges and ϵ_l the dielectric constant relative to vacuum. ϵ_l is commonly chosen to be 1 inside protein and 80 for water.

The next step to investigate the system dynamics is to numerically integrate the Newton equation of motion to obtain the atomic positions at each given time step. For each atom i , with mass m_i , position $\vec{r}_i = (x_i, y_i, z_i)$ and momentum p_i ,

$$\frac{dr_i}{dt} = \frac{p_i}{m_i} \quad 1.2.2$$

The net force acting on atom i is the negative gradient of the potential energy resulting from the interaction of the rest of the system with atom i with respect to the atomic coordinates of i .

$$F_i = -\nabla_i V \quad 1.2.3$$

The Newton equation of motion for atom i will be

$$\frac{dp_i}{dt} = F_i \quad 1.2.4$$

Given the atomic position of an arbitrary atom i along any direction x at a specific time t , the position after a short finite time Δt can be calculated from the Taylor series of x around t

$$x_i(t + \Delta t) = x_i(t) + \frac{dx_i}{dt} \Delta t + \frac{d^2x_i}{dt^2} \frac{\Delta t^2}{2} + \dots \quad 1.2.5$$

According to the Newton's second law,

$$\frac{d^2x_i}{dt^2} = \frac{F_{x,i}}{m} \quad 1.2.6$$

where $F_{x,i}$ is the net force acting on the atom i along direction x . In Eq. 1.2.5 only up to the second order derivatives are considered. The summation of the higher order term is referred to as the order of accuracy.

Several integration methods have been devised to numerically solve the Newton equation of motion. In most of these methods, the integration takes place at finite time intervals specified by

Δt . The simple Verlet algorithm[25], leapfrog Verlet[26] and velocity Verlet[27] are among the most widely used integration schemes. However, discussion about the advantages and disadvantages of each of the integrators listed above is beyond the scope of this introduction. For more detailed comparison the reader is referred to ref[7].

1.3 Implicit solvent models

The early MD simulations were performed on isolated proteins in vacuum [4, 28]. However, most proteins exist in fully or partially aqueous environment, which has considerable impacts on their structure and dynamics. Therefore, a realistic representation of the proteins surroundings is important to fully characterize their properties.

In MD simulations of water soluble proteins, proteins are commonly solvated using water molecules, and the solvated protein along with the solvent molecules are duplicated periodically in all directions to represent an infinite system and to ensure that each atom is surrounded by its neighboring molecules or their images[7]. To fully characterize solvent properties, several water models have been developed and employed in MD simulations. Some of the most popular water models are TIP3P, TIP4P[29], TIP5P[30], SPC, and SPC/E[31].

Water molecules can form hydrogen bonds with the solute. The dipole moment of water molecules also affect the charge-charge interactions between the solute atoms. However, explicit inclusion of water molecules can drastically increase the computational cost. In a recent MD study of the complete satellite tobacco mosaic virus, in order to fully solvate 165,290 atoms of proteins and nucleic acids, 299,855 water molecules had to be included in the system[32]. Inclusion of this many water molecules increases the non-bonded interactions. In addition, a significant part of the computational time would be spent on the sampling of degrees of freedom

that normally belong to the solvent and not the solute and generally have minimal effect on the overall conformational sampling of the protein. In order to reduce the computational complexity of the system, one alternative method is to substitute explicit water molecules by a mean-field model which is often referred to as implicit solvent[33]. Some implicit solvent models are very fast and can perform with a cost comparable to vacuum. For example, a simple model developed by Fraternali *et. al* can calculate the solvation free energy with almost 70 percent of the speed of vacuum[34]. However, there is a trade-off between how fast and how accurately a given implicit model can approximate the solvation free energy compared to explicit solvent [35, 36].

During the past years, implicit solvent models have been the target of many studies and many attempts have been made to improve their performance and their accuracy [37-42].

Usually the solvation free energy, ΔG_{solv} , in implicit solvent models is decomposed into two main parts: electrostatic, ΔG_{elec} , and non-polar contribution, ΔG_{np} .

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{np} \tag{1.3.1}$$

Different approaches have been proposed to calculate each term efficiently and accurately [37-42]. However, a detailed discussion on different implicit solvent models and their performance is beyond the scope of this introduction. For more information and details on implicit solvent models the reader is referred to ref [43] and the references therein. In the following chapter one of the most common techniques in calculating the solvation free energy is introduced and briefly discussed.

1.3.1 Electrostatic energy based on the Poisson-Boltzmann equation

The Poisson-Boltzmann (PB) equation is known to be the most accurate way of obtaining electrostatic potential for any protein with a given structure. Multiple dielectric constants can be employed to describe the solute and its environment and the electrostatic potential can be obtained by solving Eq. 1.3.2. This equation is numerically solved with finite difference or boundary elements methods[44].

$$\nabla(\epsilon(r)\nabla\phi(r))=-4\pi\rho(r) \quad 1.3.2$$

In the PB equation, ϵ is the dielectric constant at a particular grid point in space, ϕ is the electrostatic potential, ρ is the solute charge distribution and r is any given position in Cartesian coordinates. In order to obtain the electrostatic contribution to solvation free energy, PB should be solved twice: once in vacuum where the solvent dielectric is one (ϕ_{vac}) and once with the dielectric of interest, i.e. 80 for water, for the solvent region (ϕ_{solv}) and the resulting difference $\phi_{diff} = \phi_{solv} - \phi_{vac}$ will be the variation in the electrostatic potential due to solvation. The polar contribution to solvation free energy is then calculated according to the following equation[44]

$$\Delta G_{PB} = \frac{1}{2} \int \phi_{diff} \rho(r) dV \quad 1.3.3$$

where dV is the volume element and the integral is taken over the entire space.

For a set of discrete point charges (q_i) Eq. 1.3.3 can be rewritten as

$$\Delta G_{pol} = \frac{1}{2} \sum_i q_i \phi_{diff}(r_i) \quad 1.3.4$$

The Poisson-Boltzmann solvation free energy can be efficiently calculated for small molecules but it can get expensive for large proteins. In addition, obtaining smooth derivatives from numerical solution of the PB equation is difficult[44]. These problems call for alternative methods which can calculate solvation free energy with smaller computational costs and comparable accuracy[44]. One of the simplifications to PB is Generalized Born equation (GB)[45, 46] . In general, solvation free energy of a spherical ion (ΔG_{Born}) with radius a and charge q can be written as

$$\Delta G_{Born} = -\frac{q^2}{2a} \left(1 - \frac{1}{\epsilon_w} \right) \quad 1.3.5$$

Eq. 1.3.5 is the well known Born formula. For an assembly of charges ($q_1, \dots, q_N$) which each is embedded in a sphere with radius ($a_1, \dots, a_N$), and the separation between the i^{th} and the j^{th} particle is taken to be r_{ij} , the solvation free energy is given by a sum of individual solvation energy terms which are obtained by the Born equation (Eq. 1.3.5) and the pair-wise terms:

$$\Delta G_{pol} = \sum_{i=1}^N \frac{q_i^2}{2a_i} \left(\frac{1}{\epsilon_w} - 1 \right) + \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i}^N \frac{q_i q_j}{r_{ij}} \left(\frac{1}{\epsilon_w} - 1 \right) \quad 1.3.6$$

The second sum in Eq. 1.3.6, represents the pair-wise Coulombic interactions which are scaled by the factor $(1/\epsilon_w - 1)$ to account for the change in dielectric constant in the process of transferring from vacuum to solvent.

The purpose of the GB model is to find a relatively simple analytical equation, which can capture the physics of PB for any given geometry of a molecule. In fact, it seeks a function that can act as effective interaction distance for the pair-wise interactions and converges to the Born radius for individual atoms. The most common form of this function proposed by Still *et al.*[45] is:

$$f_{GB} = \left(r_{ij}^2 + \alpha_i \alpha_j \exp \left(-\frac{r_{ij}^2}{4\alpha_i \alpha_j} \right) \right)^{\frac{1}{2}} \quad 1.3.7$$

In Eq. 1.3.7, α_i and α_j are the Born radii of atoms i and j which not only depend on the atomic radius of each individual atom (a_i and a_j) but also on the radii and relative positions of all other atoms in the system. The polar contribution to the solvation free energy can then be calculated from Eq. 1.3.8 [45]

$$\Delta G_{pol} = 166 \left(\frac{1}{\epsilon_w} - 1 \right) \sum_{i=1}^N \sum_{j=1}^N \left(\frac{q_i q_j}{f_{GB}} \right) \quad 1.3.8$$

The Born radius of a charge at the center of a spherical solute will simply be considered as the radius of the encompassing sphere. However, if the solute geometry or the position of the charge deviate from this simple model, other modifications are required[47]. In order to obtain the Born radius of atom i in a solute with an arbitrary geometry and charge distribution, one can obtain the polar contribution to the solvation free energy by solving the PB equation with all charge being turned off except for the charge on this particular atom. Born radius of atom i can then be obtained by submitting the calculated energy to Eq. 1.3.5 and solving for α_i . By repeating this calculation for all atoms present in the molecule, it is possible to collect all Born radii. This

means that the PB equation would have to be solved twice for each atom[47] which is not computationally attractive. Therefore, the challenging task for employing the GB formalism is to find a way to estimate the Born radii without the costly solution of the PB equation. It has been shown that if the Born radii calculated according to the procedure described above are substituted into the GB equation, the resulting free energy is in close agreement with the free energy obtained from the solution of the PB equation[47].

One of the most common alternative methods for calculating the Born radii is based on Coulomb field approximation (CFA) [48]. According to CFA, the atomic self energy is assumed to be only described as the interaction of each individual atomic charge with the dipole it induces into the solvent. In other words, the work required for replacing atom i with the charge q_i at the origin of a molecule with interior dielectric ϵ_{in} and exterior dielectric ϵ_{ex} can be written as[44, 49]

$$W_i = \frac{1}{8\pi} \int \left(\frac{D}{\epsilon}\right) \cdot D dV \approx \frac{1}{8\pi} \int_{in} \frac{q_i^2}{r^4 \epsilon_{in}} dV + \frac{1}{8\pi} \int_{ex} \frac{q_i^2}{r^4 \epsilon_{ex}} dV \quad 1.3.9$$

where D is the dielectric displacement.

The electrostatic component of the solvation free energy is taken as the difference in W_i once the exterior dielectric alters from 1 to ϵ_{sol} .

$$\Delta G_{pol,i} = -\frac{1}{8\pi} \left(1 - \frac{1}{\epsilon_{sol}} \right) \int_{ex} \frac{q_i^2}{r^4} dV \quad 1.3.10$$

Comparing the resulting equation (Eq. 1.3.10) to Born equation (Eq. 1.3.5) shows that the Born radius of atom i can be formulated as

$$\frac{1}{\alpha_i} = \frac{1}{4\pi} \int_{ex} \frac{1}{r^4} dV \quad 1.3.11$$

It is convenient to rewrite Eq. 1.3.11 as the following

$$\frac{1}{\alpha_i} = \frac{1}{R_i} - \frac{1}{4\pi} \int_{ex} \frac{1}{r^4} dV \quad 1.3.12$$

in which R_i is the van der Waals radius of atom i [44, 49].

As it was mentioned earlier, CFA appears to work best for systems containing a charge at the center of a spherical solute. However, it breaks down if the solute is not spherical or if the charge approaches the surface of the solute, where the polarization effects become more dominant and the single monopole-dipole interaction term by itself cannot capture the solute solvent interaction. One intuitive solution to address the oversimplification in CFA is to include higher order terms in the solvent-solute interaction potential. Lee *et al.* introduced a new GB method based on molecular volume integration with an additional empirical correction to the Born radius that mimicked the higher order terms needed to describe the solute-solvent interactions[50]. According to this model known as GBMV, the empirical correction needed to account for higher order interactions can be written as

$$A_5 = P \sqrt{\frac{1}{2R_i^2} - \frac{1}{4\pi} \int_{r>a} \frac{1}{r^5} dV} \quad 1.3.13$$

Therefore the Born radius of atom i can be formulated as

$$\alpha_i = \frac{1}{A_5 - A_4} + \alpha_0 \quad 1.3.14$$

In the new and extended Born radius equation, A_4 is the original CFA term and P and α_0 are adjustable parameters. In this model the molecular volume was mathematically defined as

$$V_{MV}(r) = \frac{1}{1 + \exp\left(\beta \left(\sum_{j=1}^N A_j(r) - \lambda\right)\right)} \quad 1.3.15$$

$$A_j(r) = \exp\left(\frac{\gamma_0 \ln \lambda}{R_j^2} |r_j|^4\right) \quad 1.3.16$$

with γ_0 , λ and β being adjustable parameters. For a protein with 60 residues this model is reported to be 10 times slower than the vacuum simulation. However, the correlation coefficient between the calculated Born radii using GBMV and the values obtained from the PB equation shows significant improvement compared to the GB model[50].

This model was further improved by Lee *et al.* by including higher order terms in the Born radii calculation as shown in the following equations[51].

$$A_7 = \sqrt[4]{\frac{1}{4R_i^4} - \frac{1}{4\pi} \int_{solute, r > R_i} \frac{1}{r^7} dV} \quad 1.3.17$$

$$\alpha_i = \frac{S}{C_0 A_4 + C_1 A_7} + D \quad 1.3.18$$

where S , C_0 , C_1 and D are adjustable parameters and R_i is the van der Waals radius of atom i .

Also a new vector based method was developed to construct the standard molecular volume (SMV)[51]. In the new modification of SMV[51], a scaling factor based on the summation of atomic vectors was introduced. These vectors connect each point of interest to the center of each atom. Therefore in the vacant regions which are not surrounded by the solute atoms, the sum of the vectors is constructive. On the other hand if the vacant regions are confined with solute atoms, the vectors will add up destructively and consequently the molecular volume will be reduced [51]. The scaling factor is formulated as follows

$$S_{MV}(\vec{r}) = S_0 \left[\sum_j F_{MV}(\|\vec{t}_j\|) \right] \frac{\sum_j \|\vec{t}_j\|^2 F_{MV}^2(\|\vec{t}_j\|)}{\left\| \sum_j \vec{t}_j F_{MV}(\|\vec{t}_j\|) \right\|^2} \quad 1.3.19$$

In Eq. 1.3.19, $\vec{t}_j$ is the vector joining any point of interest $\vec{r}$ to the atomic center, and S_0 is an

adjustable empirical parameter. The F_{MV} depends on atomic positions and is defined as

$$F_{MV}(\vec{t}_j) = \frac{C_j^2}{\left(C_j + \|\vec{t}_j\|^2 - R_j^2\right)^2}, C_j = P_1 R_j + P_2 \quad 1.3.20$$

with P_1 and P_2 being empirically fitted parameters[51]. The improved GBMV, often referred to as GBMV2, is reported to be two times slower than the GBMV model. However, some technical improvements have shown to increase its performance to 1.2 times slower than GBMV[51].

1.3.2 Non-polar contribution based on solvent-accessible surface area

In some of the early implicit solvent models, the solvation free energy was simply considered as a function of the solvent accessible surface area (SASA) which is a measure of the exposure of each solute atom to the solvent[52, 53]. In these models, commonly, a linear function of SASA was used to estimate both the electrostatic and van der Waals interactions between solute and solvent.

$$\Delta G_{sol} = \sum_i \gamma_i SASA_i \quad 1.3.21$$

In Eq. 1.3.21, the atomic surface tension, γ_i , is obtained by fitting the solvation free energy to the experimental values. The linear SASA approach was used later to estimate the non-polar contribution to the solvation free energy while the electrostatic part was calculated by other methods [37, 42, 54, 55]. In the current GBMV models the linear approximation to the non-polar contribution is being implemented [50, 51] as follows.

$$\Delta G_{np} = \gamma \sum_i SASA_i \quad 1.3.22$$

where γ reflects the surface tension in water. Different values have been reported for γ ranging from 0.015 to 0.033 kcal/molÅ² [39, 56] based on experimental transfer energies for the alkane series [39, 57] or successful folding of peptides with implicit solvent model [58-60].

1.3.3 Limitations of implicit solvent models

In spite of the success in studying variety of systems using implicit solvent models [42, 61-65], the approximations introduced by these models can lead to some limitations. One of the well known problems of implicit solvent models is their inability to differentiate between positive and negative charges in the solution. In explicit solvent simulations for instance, water molecules can respond to the charge of the solute by adopting a different orientation in their vicinity [66-68]. However, this feature is not present in the implicit solvent models due to their symmetric nature. Several approaches have been proposed to address this problem. One solution could be to add extra dipoles to the first hydration shell of the solute [66]. Other recommended approaches are to add explicit water molecules to the regions with high charge densities [67] or to adjust the Born radii based on the induced charge density [68].

1.4 Implicit treatment of the membrane

In order to study membrane bound proteins, an accurate description of membrane bilayer and water molecules is required [69]. Membrane bilayers are composed of hydrophobic acyl chains that are attached to hydrophilic head groups. Bilayers are oriented in a way that the head groups are facing water while the acyl chains are shielded from the solvent. The most detailed representation of the protein environment can be achieved by explicit inclusion of lipid and water

molecules[70, 71]. However, the majority of the simulation time in this case will be spent on sampling the degrees of freedom that belong to the lipid bilayer or the solvent [33, 56]. The successful implicit representation of aqueous solvent media that was mentioned earlier [50, 51] encouraged the use of the similar approach in membrane environments and developing implicit membrane models that are capable of capturing the membrane environments' characteristics.

The simplest representation of an implicit membrane would be a slab with a low dielectric constant and the thickness that matches the hydrophobic core of the membrane [38, 55, 62]. In several models, this dielectric is assumed to be equal to the dielectric of the solute which allows for the extension of the integrated region in Eq. 1.3.12 to membrane water interface. Spassov *et al.* used an empirical term to approximate the membrane region of the integral as a function of the atomic distances from membrane center along the membrane normal[55]. Im *et al.* followed the same assumption with modified molecular volume function that extended over the membrane region [62]. However, the problem with such an approach is that it is limited to only two dielectric constants, i.e. the membrane dielectric and water, which might not be sufficient to represent the dielectric distribution along membrane normal[72].

1.4.1 Implicit membrane model; extension to GBMV2

In order to obtain a more realistic description of membrane bilayer, Tanizaki *et al.* presented an implicit membrane model as an extension to GBMV2[51]. In the new model known as heterogeneous generalized Born (HDGB), the membrane was considered as a heterogeneous environment with different dielectric constants along membrane normal[73]. In an earlier study, Feig *et al.* showed that the effect of solvent dielectric on the atomic Born radii can be modeled using Eq. 1.4.1[74]

$$\alpha_i(\varepsilon_w, \varepsilon_p) = \frac{1}{C_0 A_4 + C_1 \left(\frac{3\varepsilon_w}{3\varepsilon_w + 2\varepsilon_p} \right) A_7} + D + \frac{E}{\varepsilon_w + 1} \quad 1.4.1$$

where C_0 , C_1 , D and E are adjustable parameters which were obtained by fitting the electrostatic solvation free energies with the corresponding values obtained from PB for a set of test proteins[74]. ε_w and ε_p are the dielectric constants of the solvent and the solute respectively.

The same concept can be used to apply Eq. 1.4.1 for membrane environment[73]. In the new formalism of membrane model, the dielectric constant at each point along the membrane normal was defined based on the nearest solvent regions and the GB term was modified to include the dependency of the dielectric constant on the position along the membrane normal as follows:

$$\Delta G_{GB} = -166 \sum_i \sum_j \left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon'_{ij}(\varepsilon_i, \varepsilon_j)} \right) \times \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i(\varepsilon_i) \alpha_j(\varepsilon_j) \exp \left(-\frac{r_{ij}^2}{F \alpha_i(\varepsilon_i) \alpha_j(\varepsilon_j)} \right)}} \quad 1.4.2$$

and

$$\varepsilon'_{ij} = \frac{\varepsilon_i + \varepsilon_j}{2} \quad 1.4.3$$

ε'_{ij} is the effective dielectric constant at each position.

Tanizaki *et al.*[73] used the dielectric constants for different parts of the membrane based on the calculated value from explicit solvent/lipid simulation[72]. A dielectric profile was built by moving a spherical probe ion with pre-defined charge and radius along membrane normal and

solving the PB equation for different positions. The calculated solvation free energy was then used to obtain ϵ' via Eq. 1.3.5 [73].

Similar to the GBMV2 method[51], a linear function of atomic SASA was used to estimate ΔG_{np} . However, based on the insertion energy of an oxygen atom into the bilayer obtained from explicit solvent simulation, the cost of forming cavity is different for various positions along the membrane normal. This feature was considered in the membrane model by introducing a surface tension modulation profile along membrane axis[75]. This profile was further modified to account for the change in the radius of the atom upon insertion[73].

The relative error in ΔG_{GB} was shown to be 0.17% compared to PB for different orientations of a trans-membrane helix [73]. In particular, the insertion energies of neutral amino acid analogs[76] are shown to be in good agreement with explicit solvent simulations[77].

HDGB has been employed in studying several membrane integrated peptides [78, 79] and proteins[76]. In this dissertation, conformational sampling of the influenza fusion peptide and the Ebola fusion peptide in the context of the membrane [78] were investigated using the HDGB model[73, 76]. Also earlier studies of the conformational sampling of phospholamban in different membrane environments using HDGB shows that variation in the thickness of the membrane can alter the secondary structure of the protein [76]. However the overall secondary structures of phospholamban in different membrane environments are in close agreements with the NMR structures obtained in DPC micelle [80].

1.4.2 Elastic implicit membrane model

Elasticity theory, developed in 1973[81], is employed to assess local membrane deformation energy and membrane mediated protein-protein interaction [82-85]. In this model, the bilayer is thought of as self-assembled amphipathic molecules with material properties similar to those of a liquid crystal[81] which describe its resistance towards distortions. Later Huang[86] and Nielsen *et al.*[82, 87] developed a new formalism of membrane deformation based on elasticity theory that was calibrated to reproduce gramicidin A channel life time. This model is often referred to as mattress model since it allows membranes to deform vertically much like a mattress. A cartoon representation of a deformed membrane along with the mattress model representation is shown in Figure 1-1.

According to the mattress model, the membrane deformation energy can be expressed in terms of the compression, bending and stretching as follows

$$\Delta G_{mem} = \iint \frac{1}{2} \left(\frac{K_a}{L_0^2} u^2 + \frac{1}{2} K_c (\nabla^2 u)^2 + \frac{1}{2} \alpha (\nabla u)^2 \right) d\Omega \quad 1.4.4$$

where K_a is the bilayer compression modulus, K_c is the bilayer bending modulus, α is the surface tension and u is the deviation of the leaflet height from its equilibrium value. The double integral is taken over the plane of the membrane [81, 82, 87, 88]. The energy minima in Eq. 1.4.4 correspond to the equilibrium shapes that can be obtained by setting the derivative of Eq. 1.4.4 with respect to u equal to zero. This would lead to the following equation

$$\nabla^4 u - \gamma \nabla^2 u + \beta u = 0 \quad 1.4.5$$

where $\gamma = \alpha/K_c$ and $\beta = 2K_a/(d_0^2 \times K_c)$ and d_0 is the membrane equilibrium thickness. To solve the fourth order differential equation in Eq. 1.4.5, four boundary conditions are required which are normally chosen to be the thickness of the membrane at the point of contact with the inclusion and far away from the inserted protein and the slope of the membrane at these points.

The resulting membrane deformation ($u(x,y)$) can be used in Eq. 1.4.5 to obtain the membrane deformation energy.

In order to include the effect of local membrane deformation in solvation, the solvation free energy is defined as follows

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{np} + \Delta G_{mem} \quad 1.4.6$$

where ΔG_{sol} is the solvation free energy and ΔG_{elec} , ΔG_{np} and ΔG_{mem} are the electrostatic, non-polar and membrane deformation free energy.

Once $u(x,y)$ is known, it can be fed into an electrostatic calculation to determine the solvation free energy [88-92]. In most elastic continuum models, ΔG_{elec} is calculated using the PB equation [88, 89, 92]. Commonly, once the membrane geometry around the inclusion is known, the dielectric constant for each point in the grid is modified based on the membrane boundary i.e. if a point outside of the protein boundary resides below the calculated membrane surface, its dielectric will be equal to the membrane dielectric and if it is above the surface it will have the dielectric of water.

Once the dielectric map is modified for all the points in the grid the PB equation is solved for the protein in the perturbed membrane. Similarly, SASA is obtained based on the position of each atom inside or outside of the deformed membrane boundary and ΔG_{np} is calculated as the sum of atomic surface tensions times SASA. However, the critical step for these models is to obtain the boundary of the deformed membrane and membrane deformation free energy (ΔG_{mem})[88, 89]. (cf. Figure 1-1)

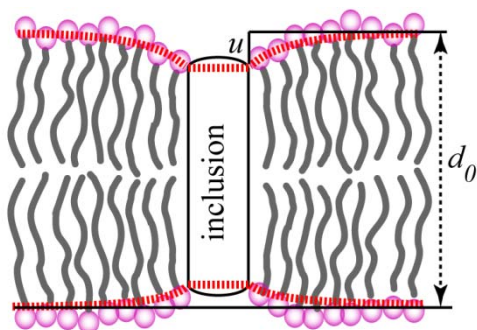


Figure 1-1: Cartoon representation of a deformed membrane. The head groups are shown with purple circles while the acyl chains are represented by gray lines. The unperturbed membrane surface is shown with solid black line and the inserted protein (inclusion) is shown with a cylinder. The dotted red line depicts the position of the membrane boundary obtained from the mattress model. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

The close agreement between charged Arg insertion energy using deformable implicit membrane developed by Choe *et al.* [88] using the technique described above and the explicit solvent simulation[93] reveals that incorporation of membrane deformation in the current

continuum membrane models can lead to a more realistic representation of membrane environment. However in the first model of the deformable membrane presented by Choe *et al.* the initial boundary conditions i.e. the height of the membrane in the vicinity of the inserted protein had to be specified *a priori* and was chosen in a way that it satisfied a predetermined sinusoidal equation. Later Callenberg *et al.* [89] added a new feature to the model in which the membrane thickness could be optimized and no prior knowledge of the local thickness was required.

The ΔG_{elec} in the current deformable membrane models is calculated via the PB equation which can be very costly to be used for MD simulations and also numerical solution of Eq. 1.4.5 would add to the computational cost. In chapter 4 we will introduce a deformable implicit membrane model as an extension to HDGB which avoids the computational cost of solving the PB equation by using GB for calculation of the electrostatic part of the solvation free energy. We also avoid the costly solution of the membrane equation by tabulating pre-calculated deformation energies.

1.5 Conformational sampling of fusion peptides in the context of the membrane

In this work, we have benefitted from the HDGB membrane model which facilitates study different fusion peptides in the context of the membrane. Application of the HDGB allowed us to extend the simulation time to sub microsecond which is an order of magnitude longer than the previous explicit solvent/lipid simulations [94-98] for a fraction of the computational cost [78, 79]. This time scale was available because the kinetic barriers in lipid rearrangements in the response to the conformational change of the peptide were avoided by the use of HDGB. Temperature replica exchange method used as the enhanced sampling technique in these studies

is also known to increase barrier crossing and improve sampling significantly [99, 100]. In this technique, several copies (replicas) of the system in different temperatures are run in parallel and after several cycles the exchange between replicas with neighboring temperature is attempted. The exchange is accepted or rejected based on the Metropolis criteria [99, 100]. As the result, the frequent visits to the higher temperatures make the barrier crossing more probable [99, 100].

In the following, the importance of fusion peptides and their roles in the cell infection process is discussed. The more detailed description on the methodology and the simulation results will be presented in chapters 2 and 3.

1.5.1 Influenza fusion peptide in the context of the membrane bilayer

In many viral diseases, the infection starts with the insertion of viral genetic material into the host cell. Commonly, the glycoproteins that are located on the surface of the viral cell membrane facilitate this process by bringing the viral and host cell membranes closer together and forming a fusion pore that allows the genetic material to be transferred from viral cell in to the host cell. These proteins are normally referred to as the fusion proteins.

One of the most well studied fusion proteins belongs to the influenza virus and is known as hemagglutinin(HA). Prior to infection, the homotrimeric[101] HA is cleaved by one of the human protease[102] resulting in three heterodimers with two subdomains, HA1 and HA2[101], which are attached to each other via a disulfide bond. HA1 is a globular domain that binds to the host cell receptors, and HA2 is a helical bundle that contains the membrane anchoring domains. According to the crystal structure of the precursor and processed HA, the major structural change occurs around the new N terminus of HA2 after cleavage [101, 103, 104].

The first twenty residues of HA2 are known as the influenza fusion peptide (IFP). At high or normal pHs, IFP is buried in the HA2 core. However due to the reduction in pH which occurs in the endosome to facilitate the indigestion, HA2 undergoes a large structural change that leads to the exposure of IFP and its insertion into the host cell membrane [105, 106]. The C terminal of HA2 is attached to the trans-membrane helix, which keeps it anchored to the viral cell membrane. Subsequently, the rearrangement of HA2 brings these two membrane anchored domain together and as a result, the membranes that are associated with these two domains are brought to the proximity of one another [103, 105, 107]. The fusion process will start with the lipid mixing of the two outer leaflets of the membranes which is known as the hemi-fusion [108] and will be completed by the formation of the fusion pore[109-111]. It has been hypothesized that the direct interactions of IFP with the host cell membrane and the perturbations due to the insertion of the peptide and the resulting imposed negative curvature is responsible for the membrane fusion initiation [112-114].

A schematic representation of a model glycoprotein in its fusogenic and non-fusogenic states is shown in Figure 1-2.

To obtain more insight about the role of the fusion peptide in membrane fusion and its structure in the membrane bound state, IFP has been the target of several mutagenesis and structural studies at both fusogenic and non-fusogenic pHs[115-122]. Many of these studies focus on the truncated form of IFP including the first 20 amino acids of HA2 sub-domain and its mutants. The liquid state NMR structure of the short form IFP in dodecylphosphatidyl choline (DPC) micelles reveals two distinct structures at pH=5.0 and pH=7.4. At fusogenic pH, the peptide adopts an invert V shape structure with its charged N[114] terminus and its neutral C terminus inserted into the hydrophobic core of the membrane and its apex exposed to the

solvent[115]. The C terminal helix of the structure seems to be severely disrupted at high pHs and the insertion depth of the peptide is reduced [115].

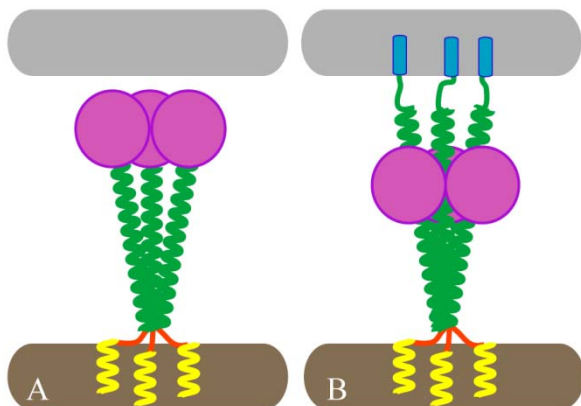


Figure 1-2: Schematic representation of a model fusion protein in (A) non-fusogenic state and (B) fusogenic state. The approximate location of the host cell is shown with a gray cylinder while the brown cylinder shows the viral membrane. The yellow and green helices show the trans-membrane domain and helical bundle respectively. The globular domain is shown with magenta circles and the fusion peptides are shown with blue cylinders.

Solid state NMR study of IFP suggests that pH has little effect on the secondary structure of the peptide and the overall structure is helix-break-helix at both pHs. However, the two distinct sets of ^{13}C chemical shifts observed in membrane bound peptide suggest that Glu11 can have two co-existing protonation states at pH=5.0[123].

Finally, the structure of IFP with additional Trp21-Tyr22-Gly23 from sero sub-type H1 in DPC micelles suggests that the peptide adopts an interfacial helix-break-helix structure with a

tight turn in the middle which is stabilized by the hydrogen bond between the charged N terminus and carbonyl oxygen of Gly20 and either Trp21 or Gly23[124].

In chapter 2, the conformational sampling of IFP in the context of the membrane has been studied as a function of pH and the termini using HDGB model and temperature replica exchange MD to enhance sampling. The details of the simulations and the results will be presented and discussed in chapter 2.

1.5.2 Ebola fusion peptide

In the second study, conformational sampling of Ebola fusion peptide (EFP) in the context of the membrane was investigated. In case of Ebola virus, the fusion protein (GP)[125] is composed of two sub-domains GP1 and GP2 that are linked via a disulfide bond. The first sub-domain, GP1, is responsible for binding to the cell receptors while the second sub-domain, GP2, mediates the membrane fusion[126]. GP shares a lot of common features with HA[127] however the significant difference between HA and GP is the location of their fusion peptides. In HA the IFP is located in the N terminus of the HA2 domain [101] while GP fusion peptide is located 22 residues away from the N terminus[128] which makes it an internal fusion peptide. Based on circular dichroism (CD) and infrared (IR) study the wild type EFP structure can adopt three states[129]. The membrane bound peptide can take either α -helix or β -sheet structure while the unbound peptide adopts a random coil structure[129]. However, according to a more recent NMR study, in the presence of sodium dodecyl sulfate (SDS) micelle, EFP adopts a helical structure in the middle region while it seems more flexible in the N and C terminus. This observation suggests that the peptide is interacting more favorably with the membrane from its middle region while its termini are exposed to the solvent. Loss of secondary structure due to

mutation of Trp 8 to Ala suggests that the aromatic-aromatic interactions between Trp8 and Phe12 play critical role in stabilizing the helical conformation in the middle region of the peptide[130].

The conformational sampling of wild type EFP and W8A mutant were studied in the membrane environment and water using HDGB and GBMV2 implicit models. In chapter 3 the results are described and compared with the experimental findings[130].

2 CONFORMATIONAL SAMPLING OF INFLUENZA FUSION PEPTIDE IN MEMBRANE BILAYERS AS A FUNCTION OF TERMINI AND PROTONATION

Afra Panahi¹, Michael Feig^{1,2*}

¹Department of Chemistry

²Department of Biochemistry & Molecular Biology

Michigan State University, East Lansing, MI, 48824

Adapted with permission from

J. Phys. Chem. B, 2010. **114**(3): p. 1407-1416

Copyright 2010 American Chemical Society

2.1 Abstract

Influenza fusion peptide is critical for mediating the fusion of viral and host cell membranes during viral entry. The interaction of monomeric influenza fusion peptide with membranes is studied with replica exchange molecular dynamics simulations using a new implicit membrane model to reach effectively microsecond to millisecond time scales. The conformational sampling of the fusion peptide was studied as a function of different N- and C-termini, including an experimental construct with an additional C-terminal tag, as well as a function of protonation of acidic residues. It is found that the influenza fusion peptide mostly adopts helical structures with a pronounced kink at residues 11-13 with both N-terminal and C-terminal helices oriented mostly parallel to the membrane surface. A charged C-terminus and the presence of a charge C-terminal tag significantly alters the conformational sampling of the fusion peptide and results in more diverse conformational ensembles that include obliquely inserted N-terminal peptide structures. Protonation of acidic residues also affects the conformational sampling, however, based on pK_a shift estimates the overall effect of pH=5 on the conformational sampling of the influenza fusion peptide appears to be only minor.

2.2 Introduction

Viral infection commonly begins with docking of a virus particle to a host cell and/or endosomal ingestion followed by fusion of the viral and host cell membranes to release the virus capsid into the interior of the host cell. Viral fusion is facilitated by viral glycoproteins. In the case of the influenza virus, the corresponding glycoprotein is hemagglutinin(HA). HA is a homotrimer[101] composed of two subdomains, HA1 and HA2, which are connected to each other via a disulfide bond. HA1 is the globular receptor domain that is attached to the virus surface, while HA2 is the membrane anchoring domain. HA2 forms the helical stem of the homotrimer and its 20 N-terminal residues (GLFGA-IAGFI-ENGWE-GMIDG) are known as the influenza fusion peptide (IFP). At high or normal pH, IFP is buried within the helical stem of HA, but under low pH conditions, as commonly found in endosomal cavities, HA undergoes a large conformational change that results in exposure of the IFP and its insertion into the host cell membrane [105, 106]. The direct interaction of the IFP with the host cell membrane is assumed to be responsible for facilitating membrane fusion. It has been hypothesized that IFP triggers the fusion process by perturbing the membrane and inducing a negative curvature [112, 113], but the detailed mechanism of how IFP (or other fusion peptides) lead to membrane fusion remains largely unclear. A first step in this direction involves studies of how just the fusion peptide interacts with membrane environments.

Low resolution structural data have suggested that the IFP adopts a regular α -helical conformation that is tilted with respect to the membrane. Reported estimates of the tilting angle are 45°[131] and 25°[114, 132]. In contrast, conformations of isolated IFP in the presence of dodecylphosphocholine (DPC) micelles derived from solution NMR at pH=5 (PDB code 1IBN) and pH 7.4 (PDB code 1IBO)[133] suggest kinked conformations that differ as a function of pH.

At pH=5, the peptide has an N-terminal α -helix and a short 3_{10} -helix at the C-terminus [115], but at pH=7.4 the short C-terminus exhibits a random coil conformation in lieu of the helix according to the solution NMR data. The kink at the center of the peptide is more pronounced at low pH where it allows formation of a hydrophobic pocket, presumably to make the insertion more favorable [115]. Furthermore, studies by Lai et al. indicate that the bent structure is essential for the fusion activity of the peptide[134]. Finally, recent solid-state NMR (SSNMR) results of IFP interacting with phospholipid bilayers also suggest that IFP adopts a bent conformation with otherwise predominantly helical secondary structure and little apparent structural variation as a function of pH[123]. According to this study, the IFP is partially inserted into the membrane bilayer with the N-terminal helix being tilted relative to the bilayer normal.

The IFP has also been studied with molecular dynamics (MD) simulations in explicit membranes [94-98] and implicit solvent/membrane environments [55, 65, 135, 136]. According to simulations of IFP embedded in a dimyristoyl-phosphatidylcholine (DMPC) lipid bilayer over 18 ns the peptide is located at the region between the head group and hydrocarbon chains of the lipid with the N-terminus close to the amphipathic interface and a secondary structure similar to the solution-NMR structure at pH=5.0[94]. Other simulations of wild-type IFP and mutant fusogenic and non-fusogenic variants in palmitoyl-oleyl-phosphatidylcholine (POPC) bilayer suggest that the fusogenic peptides preserve a kinked structure and a tilted insertion angle while the non-fusogenic peptides adopt a non-kinked conformation parallel to the membrane [95]. In these simulations, the native IFP and the fusogenic mutants preserve an α -helical structure for the first 11 N-terminal residues. Furthermore, lipid disruption and membrane thinning due to peptide insertion were also observed for the tilted structure[95]. Simulations of IFP in DPC micelles of two different sizes and in a POPC bilayer have also suggested that the peptide lies at the surface

of the micelle while its N-terminal region is inserted deeper into the bilayer sampling conformations that are similar to the solution NMR structures[96]. Another simulation study of a water-soluble analog of IFP suggests that the overall structure of the peptide may be sensitive to the composition of the membrane. This mutant was found to insert its N-terminal region obliquely into a relatively fluid DMPC bilayer while an orientation parallel to the membrane surface was maintained in dipalmitoyl-phosphatidylcholine (DPPC) bilayers [97]. In the most recent simulation study of IFP by Woolf *et al.* , IFP was embedded in a mixed lipid bilayer composed of palmitoyl-oleyl-phosphatidylglycerol (POPG) and POPC (POPC:POPG = 4:1). The main conclusion of this study was that the secondary structure depends on the insertion depth. This dependency is more pronounced in the simulation started with 1IBN (pH=5.0) as initial structure compared to a simulation started with 1IBO (pH=7.4). According to this study, the insertion angle with respect to the membrane normal and the kink angle increase as the N terminal α -helix inserts deeper into the membrane. Furthermore, it was found that the disordered C-terminus of the pH=7.4 structure folds into a helix upon deep bilayer insertion[98]. Although much has been learned about IFP interactions with membranes from past explicit lipid simulations, all of the published simulations are relatively short. They cover at most time scales of tens of nanoseconds while complete conformational sampling of IFP at the membrane-water interface likely involves μ s-ms time scales. It is therefore not clear whether the results from such simulations are sufficiently representative of the true dynamics of membrane-associated IFPs.

In order to reach longer time scales, there have been several efforts to simulate the conformational change of IFP or its mutants using implicit solvent/membrane models. Implicit membrane simulations offer the advantage of lower computational costs so that longer simulations can be run with the compromise that an atomistic description of lipid and solvent is

lost. However, the absence of explicit lipids avoids potentially long lipid relaxation times and transient bilayer imbalances due to unilateral peptide insertion thereby greatly accelerating the effective time scales that can be covered in such simulations. Most implicit membrane studies of IFP have reported a predominantly horizontal orientation with the peptide parallel to the membrane bilayer or only slightly inserted conformations[55, 65, 136], whereas highly oblique conformations were largely unfavorable. Furthermore, Sammalkorpi *et al.* [136] reported that the monomer is predominantly α -helical without a kink at the center of the peptide which contradicts the experimentally reported kinked structures from solution NMR. Due to the relatively simple nature of the implicit membrane models used to date, it is not clear, though, that the results are sufficiently representative of the interactions of IFP with membranes over extended time scales.

Here, we are revisiting IFP dynamics in the context of phospholipid bilayers with a new, more realistic implicit membrane model in combination with replica exchange simulations to enhance sampling and effectively study dynamics that occur in reality over microseconds to milliseconds. Such time scales are several orders of magnitude longer than previous simulations of IFP. The resulting conformational sampling differs significantly from previous studies but is in overall good agreement with the most recent SSNMR studies of IFP in phospholipid bilayers.

In the following, the computational methodology is described before results are presented and discussed.

2.3 Computational methodology

Replica exchange molecular dynamics simulations of the influenza fusion peptide were carried out with the CHARMM program[137] in conjunction with the MMTSB Tool Set[138]. Initial structures were downloaded from the Protein Data Bank[133] (PDB codes 1IBN (pH=5.0) and

1IBO (pH=7.4)). Different combinations of termini and protonation states were simulated for a total of 8 different sets as summarized in Table 2-1.

Table 2-1: Simulated IFP systems with different C- and N-termini and protonation states of acidic residues.

System	Abbreviation	Time
Neutral N- terminus, neutral C –terminus	NH ₂ /CONH ₂	100 ns
Neutral N –terminus, charged C –terminus	NH ₂ /COO ⁻	100 ns
Charged N-terminus, charged C –terminus	NH ₃ ⁺ /COO ⁻	100 ns
Charged N-terminus, neutral C –terminus	NH ₃ ⁺ /CONH ₂	100ns
Charged N-terminus, neutral C –terminus, Glu11 protonated	NH ₃ ⁺ /E11p/ CONH ₂	100 ns
Charged N-terminus, neutral C –terminus, Glu15 protonated	NH ₃ ⁺ /E15p/ CONH ₂	100 ns
Charged N-terminus, neutral C –terminus, Asp19 protonated	NH ₃ ⁺ /D19p/ CONH ₂	100 ns
Charged N-terminus, GCGKKKK tag, neutral C –terminus	NH ₃ ⁺ /tag/CONH ₂	75ns

In all of the simulations, the CHARMM22[139] force field including the CMAP correction[24] was used to describe the peptide interactions. The membrane was considered implicitly using the heterogeneous dielectric generalized Born model (HDGB)[73] that has been developed in our group. In the HDGB model, the dielectric constant and surface tension vary continuously along the membrane normal (z axis). In this paper the thickness of the membrane is defined as the distance between the center of the membrane and water with dielectric constant equal to 80. The thickness of the membrane was chosen to be 25 Å. The dielectric and non-polar profiles described in the initial HDGB paper[73] were recently optimized further based on amino acid side chain analog free energy of insertion profiles from explicit solvent simulation as described elsewhere[76]. In this study, the modified dielectric and non-polar profiles were repeated periodically as depicted in Figure 2-1 in order to achieve a finite peptide concentration in z -direction and ensure that peptides that dissociate and diffuse away from the membrane interface eventually interact with the periodically repeated membrane continue to contribute to the sampling of membrane-bound peptides. Other implicit solvent parameters were set as described previously[74]. The SHAKE algorithm[140] was applied to constrain bonds involving hydrogen atoms. Since the peptide is small, no cutoff was used for the calculation of non-bonded interactions. Each system was minimized for 50 steps of steepest decent method followed by 500 steps with the adapted-bases Newton-Raphson method. The minimized structures were then used as initial structures to run temperature replica exchange molecular dynamics simulations[99, 100]. Eight replicas were run with temperatures spaced exponentially between 300 and 500 K

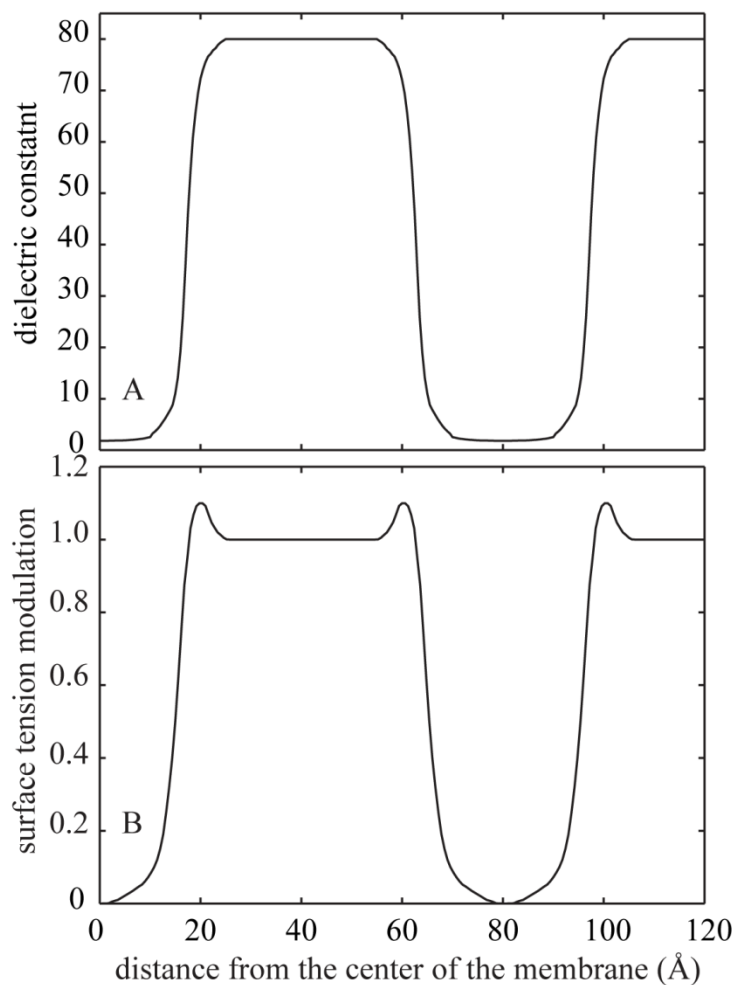


Figure 2-1: Dielectric and non-polar profiles for periodic membrane along z axis. The center of the membrane is located at the center of the origin. The membrane has a thickness of $25\text{\AA}$ and the dielectric constant of water is set to 80.

and controlled using Langevin dynamics[141] with a friction constant of 1ps^{-1} to further enhance sampling[142]. An integration time step of 1.5 fs was used for all simulations to maintain stable simulations with the implicit solvent model[143]. Exchanges between replicas were attempted after every 200 steps. For each simulation, the first 9 ns were discarded. In order to find

representative structure for each system, the structures were clustered based on mutual similarities as measured by root mean square deviations (RMSD). The standard superposition scheme used during the RMSD calculation was modified to preserve peptide orientations with respect to the z axis and distinguish between different membrane insertion depths. This was achieved by allowing only transitions along x and y directions and rotation around the z axis to reach optimal superpositions according to the following scheme:

To align any structure ($C_{mp,old}$) with a reference structure (C_{ref}), c and r were defined as a matrix of the difference of each C_α atom from the average C_α positions for $C_{mp,old}$ and C_{ref} along Cartesian coordinates respectively. R was defined as

$$R = \begin{bmatrix} c_x \times r_x & c_x \times r_y & 0 \\ c_y \times r_x & c_y \times r_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad 2.3.1$$

The rotation matrix (U) was calculated as the matrix of the normal eigenvalues of R . The coordinates for rotated structure are

$$C_{mp,rot} = U \times C_{mp,old} \quad 2.3.2$$

This rotation matrix allows only rotation around z axis. Finally the new coordinates for aligned structure after translation were calculated as

$$\begin{aligned} C_{mp,new}(x) &= C_{mp,old}(x) + C_{ref}(x) \\ C_{mp,new}(y) &= C_{mp,old}(y) + C_{ref}(y) \\ C_{mp,new}(z) &= C_{mp,old}(z) \end{aligned} \quad 2.3.3$$

Using the modified superposition scheme, the K-means method was used to cluster conformations based on RMSD with a radius of 3 Å. Only structures with a center of mass distance from the membrane center (COFM) between 15 Å and 25 Å were considered in order to focus on peptides interacting with the membrane in an interfacial manner. The most representative structures were determined as the structures closest to the center of each cluster.

Potentials of mean force (PMF) at 300K were calculated from the replica exchange simulations using weighted histogram analysis (WHAM)[144, 145] to include sampling from higher temperature replicas in the overall free energy surface. The peptide insertion depth and angle were used as the primary reaction coordinates in the construction of the PMFs. The insertion depth was calculated as the distance of the COFM of each peptide from the center of the membrane ($z=0$). Peptides interacting with different implicit membranes due to the periodic membrane profiles or with opposite leaflets of the same membrane after spontaneous crossing of the bilayer were mapped onto a common scale by calculating the distance to the closest membrane for all structures. The insertion angle was calculated as the angle between membrane normal (z axis) and the vector that connects the N ($i+4$) and O (i) atoms of the backbone for i from 3 to 6. The focus on the N-terminal part of the peptide was chosen to match experimental measurements based on SSNMR[123]. The conformation of the peptides was further analyzed by calculating helicity as a function of residue number. Two criteria were used to define the presence of helical structure: 1) An O(i) – H-N($i+4$) distance of less than 2.6 Å[146] and 2) backbone torsions in the range $-70 \leq \psi \leq -50$ and $-35 \leq \phi \leq -55$.

In order to investigate the pH effect on the configuration of the peptide, alternate protonation states of E11, E15 and D19 were considered for the peptide terminated with NH_3^+ and CONH_2 .

pK_a values for protonation of each of these acidic residues were calculated as a function of z to determine the fraction of protonated vs. charged amino acids and to average the conformational ensembles accordingly. In order to carry out this analysis, the membrane was divided into slabs each with a thickness of 0.5 Å. For each slab i average energies were then obtained from the structures located within that slab from simulations of the protonated ($\bar{E}_{P,i}$) and charged ($\bar{E}_{C,i}$) states. The average of the energies of the slabs from $z=38$ Å to $z=40$ Å were chosen as the reference energy for protonated ($\bar{E}_{P,W}$) and charged states ($\bar{E}_{C,W}$) in water. ΔE_i was then defined as

$$\Delta E_i = (\bar{E}_{P,i} - \bar{E}_{C,i}) - (\bar{E}_{P,W} - \bar{E}_{C,W}) \quad 2.3.4$$

and the pK_a shift for the i^{th} slab (ΔpK_a^i) was calculated as

$$\Delta pK_a^i = \frac{\Delta E_i}{kT} \log(e) \quad 2.3.5$$

The ratio of protonated versus charged residue for the i^{th} slab $\left(\frac{[P]}{[C]}\right)_i$ is then obtained according

to:

$$\left(\frac{[P]}{[C]}\right)_i = 10^{\left(\Delta pK_a^i + pK_a^* - \text{pH}\right)} \quad 2.3.6$$

where pK_a^* is the standard pK_a value of Glu and Asp residues. A value of 4.25 was used for E11 and E15 and a value of 3.65 for D19[147]. pH was set to 5.0 to match experimental

conditions[115, 123]. The ratio of protonated vs. unprotonated side chain calculated from Eq. 2.3.6 was then used to weight the ensembles obtained with the fixed charged and protonated ionization states. The weighting was achieved by randomly selecting from the two ensembles according to the fraction of protonated structures.

2.4 Results and discussion

Replica exchange molecular dynamics of wild-type influenza fusion peptide with an implicit membrane model were carried out to study the dynamic sampling of membrane associated IFPs. Because the exact protonation state of IFP at pH=7.4 and pH=5 is not entirely clear, the effect of different combinations of termini and Asp/Glu protonation states on the conformational sampling of IFPs was investigated as listed in Table 2-1.

2.4.1 Conformational sampling as a function of different termini

The conformational sampling of membrane-associated IFP was characterized with two-dimensional PMFs as a function of the N-terminal insertion angle and the position of the peptide center of mass relative to the membrane center (COFM) (see Figure 2-2 A-D). Structures with a COFM between 15 Å and 25 Å were clustered and the resulting clusters are projected onto the same reaction coordinates (see Figure 2-2 E-H). Representative structures are shown in Figure 2-3. Furthermore, average insertion depths and insertion angles for membrane-associated peptides are given in Table 2-2. In all cases, the IFP predominantly samples interfacial configurations with the N-terminal helix either parallel to the membrane surface (around 90° insertion angle) or nearly perpendicular to the surface with the N-terminus pointing towards the membrane (insertion angles near 150°). Different termini primarily affect the balance between parallel and perpendicular conformations

In addition, the peptide dissociates from the membrane for a significant amount of time, especially when both termini are charged. At rare occasions, short-lived transmembrane configurations are seen where the IFP crosses the membrane with the help of elevated temperatures.

Table 2-2: Average COFM, membrane insertion angles, and helicity for IFP peptides with different termini from structures at 300K with COFM values of less than 25 Å. The tag in $\text{NH}_3^+/\text{tag}/\text{CONH}_2$ is excluded in calculating the average values reported here. Statistical errors are given in parentheses.

Termini	COFM [Å]	insertion angle [deg]	helicity (based on H-bonds)	helicity (based on ϕ/ψ -sampling)
$\text{NH}_2/\text{CONH}_2$	18.03(0.004)	98.81(0.034)	0.68(0.0003)	0.41(0.0002)
NH_2/COO^-	19.94(0.005)	117.73(0.062)	0.43(0.0004)	0.27(0.0002)
$\text{NH}_3^+/\text{COO}^-$	21.69(0.008)	124.14(0.155)	0.24(0.0004)	0.18(0.0004)
$\text{NH}_3^+/\text{CONH}_2$	18.34(0.004)	92.29(0.037)	0.70(0.0007)	0.40(0.0003)
$\text{NH}_3^+/\text{tag}/\text{CONH}_2$	18.31(0.014)	91.06(0.15)	0.64(0.0008)	0.24(0.0007)

A more detailed analysis reveals further differences as a result of different termini. With a neutral C- and N-terminus ($\text{NH}_2/\text{CONH}_2$) the PMF (Figure 2-2 A) shows a broad minimum at insertion angles between $80\text{-}100^\circ$ and insertion depths between $15\text{ \AA} - 20\text{ \AA}$. The dominant structure of this region (A1) involves a long N-terminal α -helix parallel to the membrane followed by a sharp turn and a short C-terminal helix that is nearly parallel to the N-terminal helix but points away from the membrane (see Figure 2-3). In this structure, all hydrophobic side chains are oriented towards the membrane interior while the charged side chains are pointing away into the aqueous solvent. Other less-populated similar structures (A2 and A3, A4) have more obliquely inserted N-termini with angles of 111° , 120° and 142° respectively. Because the parallel N-terminal conformations dominate, the average insertion angle for uncharged termini is 98.81° while the average insertion depth is $18.03\text{ \AA}$.

With a neutral N-terminus but charged C-terminus (NH_2/COO^-), two distinct, similarly populated minima appear in the PMF, one with deeper insertion and parallel orientation, the other one with oblique orientation and shallower insertion into the membrane bilayer. As a result, the average COFM distance and insertion angle are increased to about $20\text{ \AA}$ and 120° , respectively. Representative structures shown in Figure 2-3 indicate that the peptide conformation in the parallel orientation (B1, B2) is similar to the dominant conformation with the neutral termini (A1-A3). However, they differ in how the C-terminal helix is oriented towards the membrane. The charged C-terminus points away from the membrane and as a result the hydrophobic face of the C-terminal helix is rotated away from the membrane. The structures in the second minimum are more diverse (see B3 and B4), but all feature a C-terminus that is pointing away from the membrane. The calculated N-terminal insertion angles for these

conformations are relatively large (144° for B3 and 167° for B4), but as can be seen from the conformations these values do not correspond to a single long obliquely inserted N-terminal helix. Rather, they result from a short N-terminal helix oriented parallel to the membrane followed by a kink and structure elements pointing in the opposite direction than the initial helix. This implies that the oblique orientation observed in experiment[123] can also be obtained as an average over a rather diverse set of structures with a negligible population of obliquely inserted rigid N-terminal helices.

When both termini are charged ($\text{NH}_3^+/\text{COO}^-$), the propensity of insertion into the membrane bilayer is reduced further due to the additional charge. There is significant sampling of fully dissociated peptides and while there are also two minima for membrane-associated peptides in the PMF in Figure 2-2 C, the minimum corresponding to parallel orientation and deeper insertion is less populated than the minimum with shallower insertion and large insertion angles. Representative structures C1-C5 are shown in Figure 2-3. All of the structures except for C2 involve the formation of a salt bridge between the termini and contain a short helix in N and C termini, which are bent towards each other due to the attraction between the termini. Again, the large insertion angles are a result of more complex internal geometries rather than being indicative of an obliquely inserted N-terminal helix.

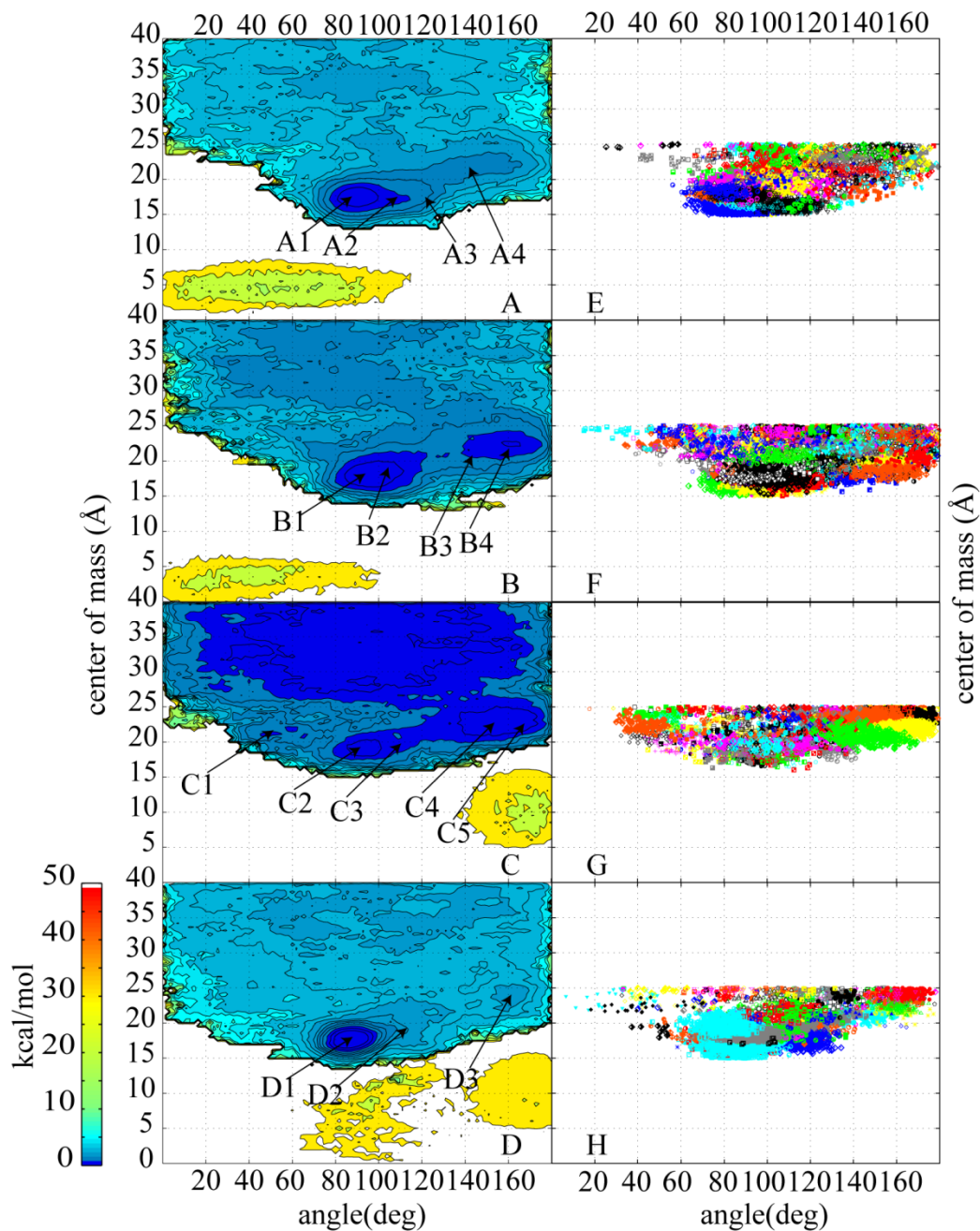


Figure 2-2: Left panels: PMFs from WHAM for simulations with NH₂/CONH₂ (A), NH₂/COO- (B), NH₃⁺/COO- (C) and NH₃⁺/CONH₂ (D) termini as a function of the N-terminal membrane insertion angle (in degrees) and distance of the peptide center of mass from the membrane center (in Å). Different colors reflect relative free energies as indicated by the color bar.

Figure 2-2 (cont'd):

Right panels: Clusters of conformations at 300 K with COFM between 15 Å and 25 Å projected onto insertion angle and center of mass distance for simulations with NH₂/CONH₂ (E), NH₂/COO-(F), NH₃⁺/COO-(G) and NH₃⁺/CONH₂(H) termini. Clustering was performed based on RMSD as described in the computational methodology section with a cluster radius of 3 Å.

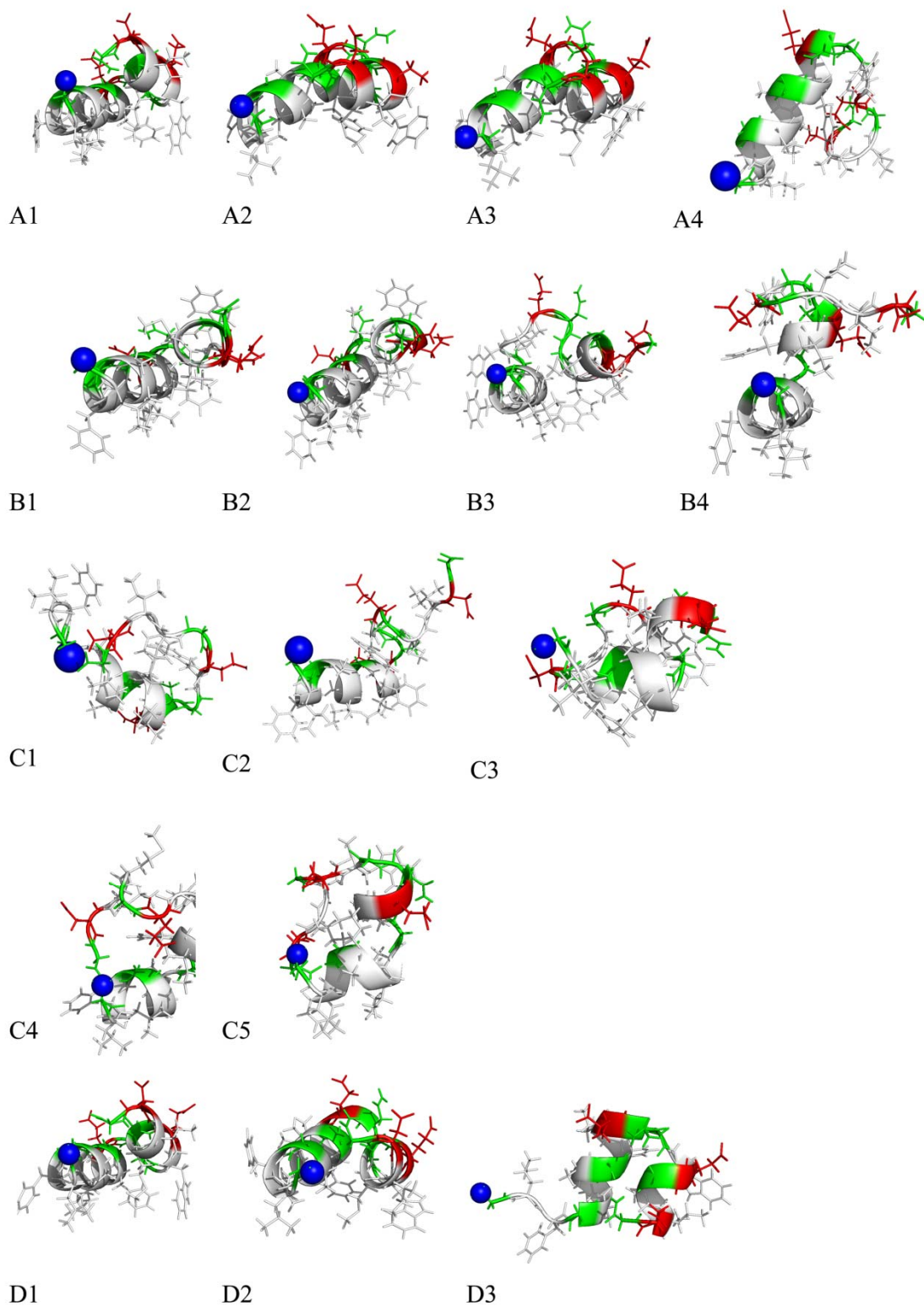


Figure 2-3: Dominant structures for simulations with $\text{NH}_2/\text{CONH}_2$ termini (A1-A4),

Figure 2-3 (cont'd):

NH₂/COO⁻ termini (B1-B4), NH₃⁺/COO⁻ (C1-C5) termini, and NH₂/CONH₂ termini (D1-D3).

The backbone is shown as a ribbon and side chain in stick representation. Charged residues are colored in red, glycines in green and hydrophobic residues in white. The z-axis is pointing up in all figures. All figures were generated with pyMol[148].

Finally, with a charged N-terminus and neutral C-terminus ($\text{NH}_3^+/\text{CONH}_2$) the IFP strongly associates with the membrane in a parallel orientation similar to the dominant conformation with neutral N and C-termini (see D1 and D2 in Figure 2-3) which allows the charged N-terminus to be pointed away from the membrane interface. There is also a much less populated perpendicular configuration (D3) that was not observed in the simulations with the other termini.

Average helicities were calculated based on the distance of $\text{O}(i)$ and $\text{H-N}(i+4)$ and based on dihedral angles as explained in the computational methodology section for structures within 25 Å from the center of the membrane. Averages for the entire peptide are given in Table 2-2 and helicities as a function of residue are shown in Figure 2-4. It is found that while the structures with a neutral C-terminus are mostly helical, the average helicity for the peptide with a charged C-terminus ($\text{NH}_3^+/\text{COO}^-$ or NH_2/COO^-) is significantly lower, especially in the C-terminal region. This appears to be due to the repulsion between the two charged residues (E15 and D19) and the charged C-terminus so that more extended structures are favored to maximize the distance between the charged C terminus and the charged side chains. In contrast, the protonation state of the N-terminus does not seem to have a strong effect on average helicity. With all termini, there is a break in helicity at residues 11-13 since hydrogen bonds 10-14 and 11-15 are disrupted and dihedral angles at residues 12 and 13 lie outside the α -helical region. The break in helicity is the consequence of a consistently observed kink at that part of the sequence. While this finding is not consistent with an early low-resolution structure of IFP[132] and with some simulation results[136], the presence of a kink agrees with solution NMR[115] and SSNMR[123] results that indicate a break of the helical structure near E11/N12. It also agrees with the most recent simulations of IFP in explicit membranes[98].

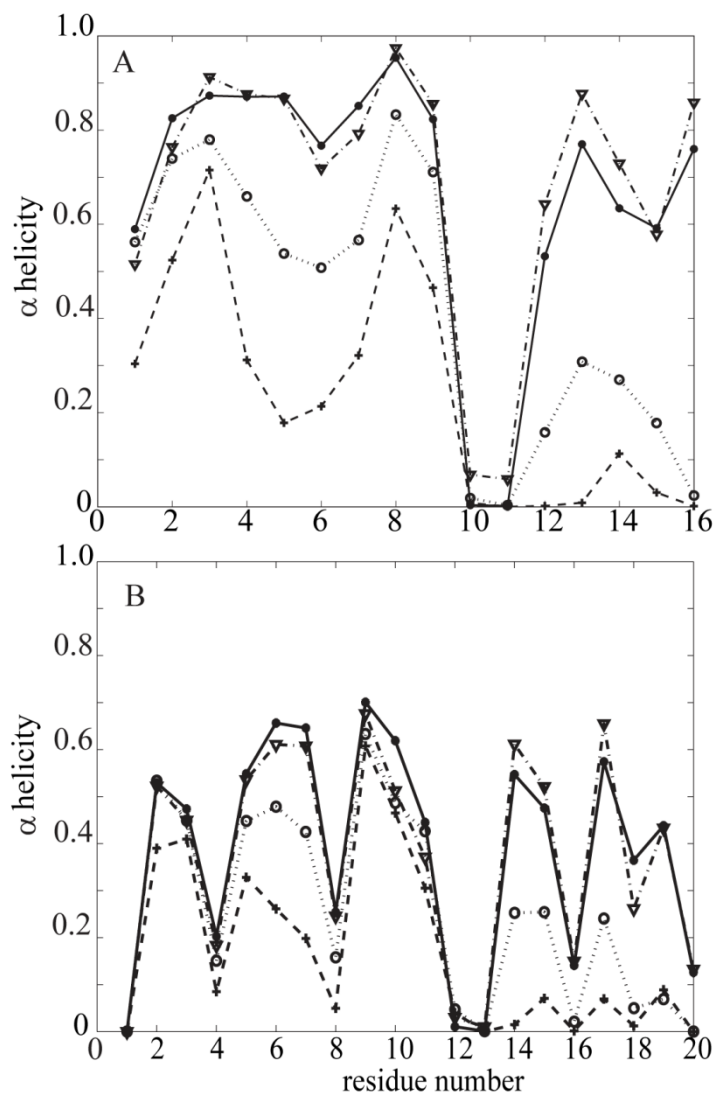


Figure 2-4: Average fraction of helical conformations as a function of residue for IFP with different termini. A: Helicity is defined by $O(i) - H-N(i+4)$ hydrogen bonding ($< 2.6 \text{ \AA}$ distance); B: Helicity is defined by $-70 \leq \psi \leq -50$ and $-35 \leq \phi \leq -55$. NH_2CONH_2 —★—, NH_2COO^- ...○..., $NH_3^+COO^-$ -+·, $NH_3^+CONH_2$ -▽-

¹⁵N NMR measurements of the pK_a of the α amino acid of Gly1 indicate that the N-terminus is in fact protonated[114]. The C-terminus, on the other hand, would be attached to the remainder of the full-length HA2 domain in the virus while most IFP experiments use a charged C-terminal tag composed of seven different residues (GCGKKKK) to prevent aggregation[115, 123, 134]. Therefore, the $\text{NH}_3^+/\text{CONH}_2$ combination would be most appropriate to describe the behavior of IFP *in vivo*. A charged C-terminus may be relevant for comparison with the experimental studies, however, the salt bridges formed between the $\text{NH}_3^+/\text{COO}^-$ termini are not likely to be representative of the conformational sampling of the experimental IFP construct.

Based on the results from the simulations presented here, it appears that the dominant structure of the IFP involves a mostly helical structure, oriented mostly parallel to the membrane with hydrophobic residues oriented towards the membrane and a pronounced kink near residues 11-13. The parallel membrane orientation matches results from previous simulations[136], but in contrast to those studies we do not observe a significant population of completely straight helices.

In order to test directly how the charged C-terminal tag GCGKKKK that is commonly used in the experimental studies may affect the conformational sampling of IFP, we also simulated IFP with such a tag with a charged N-terminus and a neutral C-terminus. The PMF and clusters as a function of N-terminal insertion angle and center of mass position are shown in Figure 2-5. There are two distinct minima in the PMF. Representative structures for these two minima are shown in Figure 2-6, labeled E1 and E2. The structure with an insertion angle of 84° (E1) is very similar to D1 in Figure 2-3, the dominant structure of $\text{NH}_3^+/\text{CONH}_2$. It contains a long N-

terminal helix followed by a break in the middle, a short helix (W14-M17), and another kink followed by a C-terminal helix for the tag sequence that points away from the membrane surface. The second conformation, E2, has an N-terminal insertion angle of 128° with a slight loss of helicity for the first three amino acids but otherwise similar topology with kinks near residues 11-13 and after residue 19. In this conformation, the two lysine residues from the tag form a salt bridge with E11. This interaction stabilizes the oblique orientation by bringing E11 further out of the membrane. These results indicate that the presence of the charged tag at least in part affects the conformational sampling of IFP compared to peptides without such a tag.

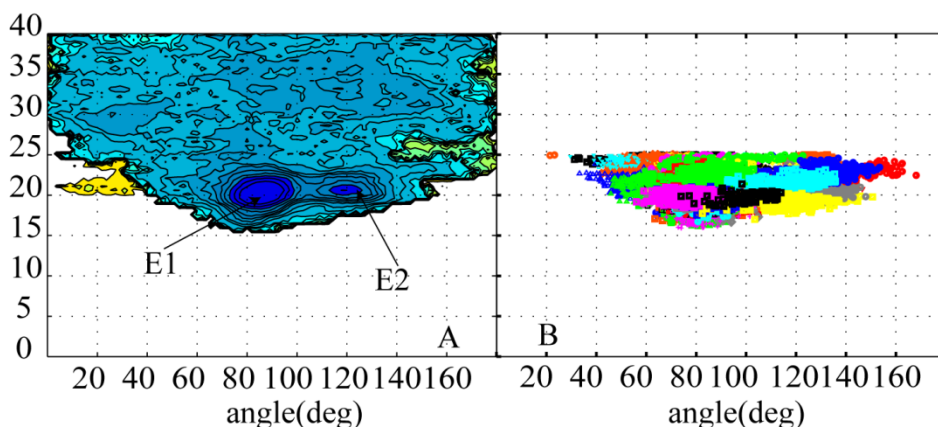


Figure 2-5: PMFs from WHAM for simulations with $\text{NH}_3^+/\text{tag}/\text{CONH}_2$ as a function of the N-terminal membrane insertion angle (in degrees) and distance of the peptide center of mass from the membrane center (in Å). Different colors reflect relative free energies as indicated by the color bar in Figure 2-2 . Right pannel Clusters of conformations at 300 K with COFM between 15 Å and 25 Å projected onto insertion angle and center of mass distance.

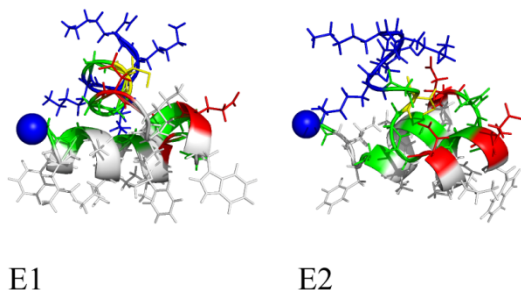


Figure 2-6: Dominant structures for simulations with $\text{NH}_3^+/\text{tag}/\text{CONH}_2$. The backbone is shown as a ribbon and side chain in stick representation. Charged residues are colored in red, glycines in green hydrophobic residues in white cysteine in yellow and lysine in blue. The z-axis is pointing up in all figures. All figures were generated with pyMol[148].

2.4.2 Conformational sampling as a result of different protonation states

The results discussed so far assumed that the acidic residues E11, E15, and D19 are charged which is appropriate for pH=7 but not necessarily for lower pH values around 5 where one or more acidic residues may be protonated due to shifted pK_a values. Low pH is physiologically relevant for fusion activity and it has been reported previously that the IFP structure depends highly on the pH of the environment [115]. There is experimental evidence[123] that E11 may exist in two different protonation states at pH=5. Another study indicates that pK_a values may lie between 5.2 and 6.0 for E11, E15 and D19 in a glutamic acid rich analog of IFP[149]. We therefore investigated the effect of pH by repeating simulations with the $\text{NH}_3^+/\text{CONH}_2$ for singly protonated E11, E15 or D19 residues in combination with pK_a shift calculations to estimate the population of protonated peptides as a function of insertion depth. New PMFs were

then generated by combining the sampling from the acidic and protonated structure according to the calculated pK_a shifts.

The PMFs for the protonated IFP variants are shown in Figure 2-7 and representative structures from clustering are shown in Figure 2-8. The conformational sampling of the $NH_3^+/E11p/CONH_2$ features a broad minimum below 25 Å that encompasses structures that are at first sight similar to the parallel-oriented configurations found for the fully charged $NH_3^+/CONH_2$ peptide. However, the N-terminal helix extends further and the kink is shifted to residues 13 and 14 while slightly oblique orientations are observed with insertion angles ranging between 60° and 120°. Furthermore, the insertion depth is reduced for the E11 protonated peptide, presumably as a result of the orientation of the protonated E11 residue towards the membrane and exposure of W14 to the high dielectric solvent (see F1-F4 in Figure 2-8). An interesting observation is that the $NH_3^+/E11p/CONH_2$ peptide seems to be more likely to cross the membrane via a transmembrane configuration compared to all the other IFP variants simulated here. According to our simulations the transmembrane orientation is only about 5 kcal/mol higher in energy than an interfacial membrane orientation. This may point to a specific role of $NH_3^+/E11p/CONH_2$ in disrupting the membrane in a pH-dependent fashion during the initiation of fusion.

Protonation of E15 results in a conformation with a long C-terminal helix, a kink near residue 6 and a short N-terminal helix (see G1-G3 in Figure 2-8). In these structures, both N- and C-termini are deeply inserted at the membrane interface. The N-terminal insertion varies from 70° to about 140° but because of the very short N-terminal helix, the insertion angle is affected by the

local structural variations and in particular the kink region and does not actually correspond to an obliquely inserted N-terminal helix.

Finally, the $\text{NH}_3^+/\text{D19p}/\text{CONH}_2$ variant has a central helix but largely disordered C- and N-terminal regions (see Figure 2-8 H1-H4). This peptide is not inserted deeply and features N-terminal insertion angle of around 120° .

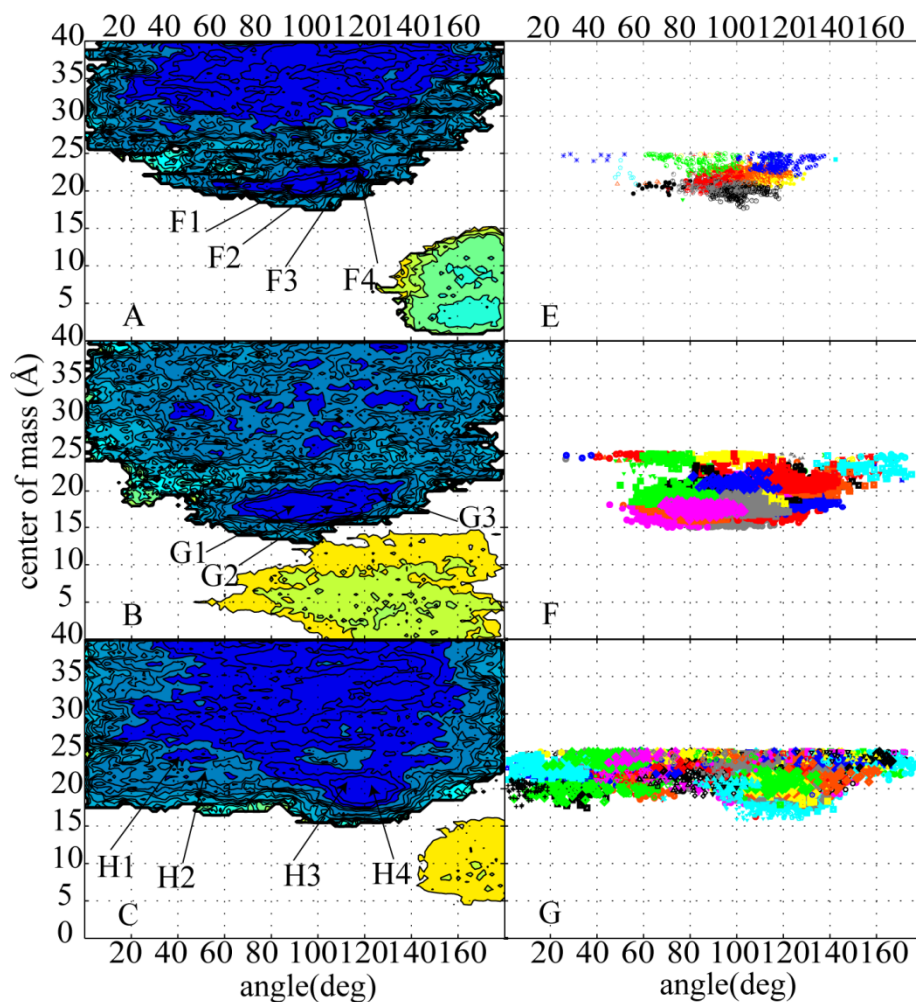


Figure 2-7: Left panels: PMFs from WHAM for simulations with $\text{NH}_3^+/\text{E11p}/\text{CONH}_2$ (A), $\text{NH}_3^+/\text{E15p}/\text{CONH}_2$ (B), $\text{NH}_3^+/\text{D19p}/\text{CONH}_2$ (C), as a function of the N-terminal membrane

Figure 2-7 (cont'd)

insertation angle (in degrees) and distance of the peptide center of mass from the membrane center (in Å). Different colors reflect relative free energies as indicated by the color bar in Figure 2-2. Right panels: Clusters of conformations at 300 K with COFM between 15 Å and 25 Å projected onto insertion angle and center of mass distance for simulations with $\text{NH}_3^+/\text{E11p}/\text{CONH}_2$ (E), $\text{NH}_3^+/\text{E15p}/\text{CONH}_2$ (F), $\text{NH}_3^+/\text{D19p}/\text{CONH}_2$ (G). Clustering was performed based on RMSD as described in the computational methodology section with a cluster radius of 3 Å.

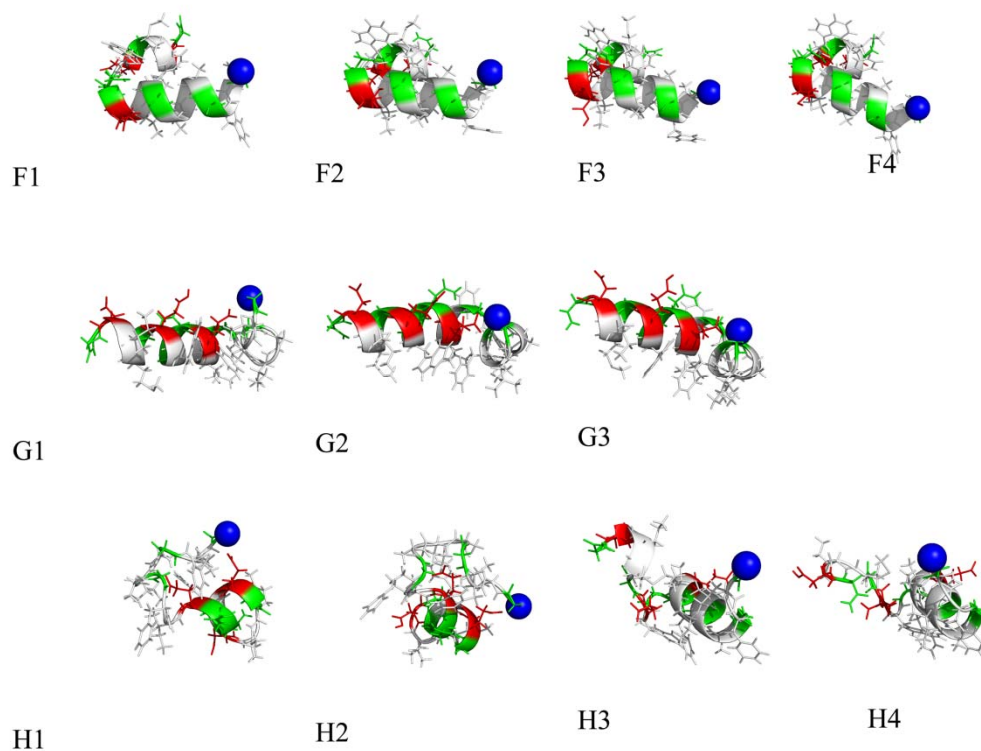


Figure 2-8: Dominant structures for simulations with $\text{NH}_3^+/\text{E11p}/\text{CONH}_2$ (F1-F4) , $\text{NH}_3^+/\text{E15p}/\text{CONH}_2$ (G1-G3) and $\text{NH}_3^+/\text{D19p}/\text{CONH}_2$ (H1-H4). The backbone is shown as a ribbon and side chain in stick representation. Charged residues are colored in red, glycines in green and hydrophobic residues in white. The z-axis is pointing up in all figures. All figures were generated with pyMol[148].

After establishing the structural ensembles of the protonated peptides, we calculated pK_a values at different insertion depths for single protonation of E11, E15, and D19 based on the sampled peptide conformations using a continuum electrostatics scheme with the HDGB implicit membrane model. Figure 2-9 shows the resulting pK_a shifts. A significantly shifted pK_a value is found for E11 protonation (up to 2 pK_a units at insertion depths between 20 and 25 Å). A slight shift is also present for D19 protonation (about 1 pK_a unit at insertion depths of 25 to 32 Å). It should be pointed out that large positive pK_a shifts are expected for fully inserted peptides because the insertion of charged residues into the membrane interior is not favorable, however, this is not reflected in the pK_a curves in Figure 2-9 because there is only little conformational sampling of fully inserted peptides at 300K.

Based on the calculated pK_a shifts, we then generated mixed ensembles at pH=5 by calculating the probability of protonated vs. charged residues at different insertion depths and selecting conformations from the ensembles of the charged and protonated peptides accordingly. The resulting, artificially generated ensemble was then used to generate the PMFs depicted in Figure 2-10. The resulting PMFs are similar to the PMFs of the charged $\text{NH}_3^+/\text{CONH}_2$ peptide. However, for E11 and D19 protonation, the sampling at large insertion angles of 160° is slightly enhanced. Furthermore, the minimum at 90° is extended slightly towards insertion angles of 110°. These results indicate that low pH does not appear to significantly alter the conformational sampling of IFP. This conclusion is in agreement with recent results from SSNMR measurements[123].

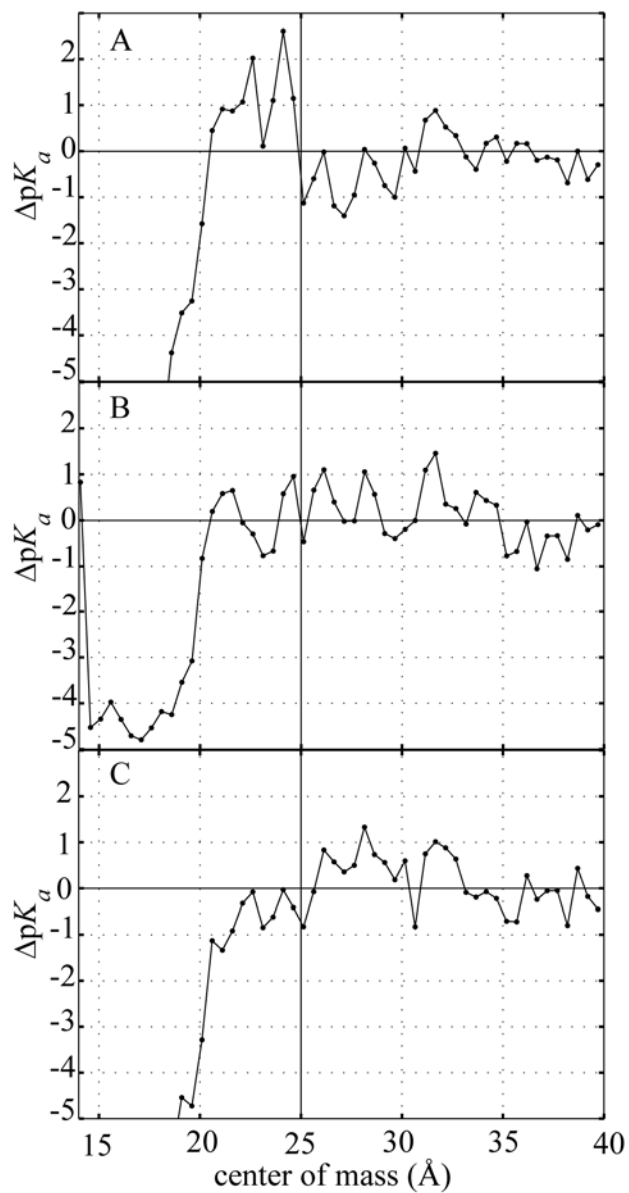


Figure 2-9: pK_a shifts of E11 (A), E15(B) and D19 (C) as a function of insertion depth. The

lines mark the membrane surface at 25 Å and no pK_a shift at 0.

2.5 Conclusions

In this study we simulated the conformational sampling of the influenza fusion peptide as function of termini and protonation of acidic residues. An implicit description of water and membrane was employed to reach estimated μ s-ms time scales. All of the structures are mostly helical with a pronounced kink near the middle of the IFP sequence. Most conformations involve a more or less parallel orientation of the N- and C-terminal helices where the hydrophobic side chains are inserted deeply into the membrane and the charged/polar side chains are oriented towards water. With charged C-termini, helicity is reduced and more diverse structures, some with average N-terminal insertion angles closer to 160° are observed. Furthermore, the addition of a charged C-terminal tag commonly used in experimental studies of fusion peptides to prevent aggregation appears to affect the conformational sampling by stabilizing a conformation with an obliquely N-terminal helix that is not observed in any of the other simulations without the tag.

The possible effect of pH was studied by first analyzing the conformational sampling of fusion peptides with protonated amino acids and then mixing conformational ensembles according to calculated pK_a shifts. Although differences in the conformational sampling of the protonated peptides were found, the pK_a shifts do not appear to be enough to significantly affect the overall ensembles at pH=5. As a result, we conclude that low pH only has a very minor effect on the conformational sampling of monomeric fusion peptide.

In this study the enhanced sampling replica exchange and the use of implicit solvent model has enabled us simulated the influenza peptide for an order of magnitude longer than previous studies done with explicit solvent. Although the results appear to be consistent with available

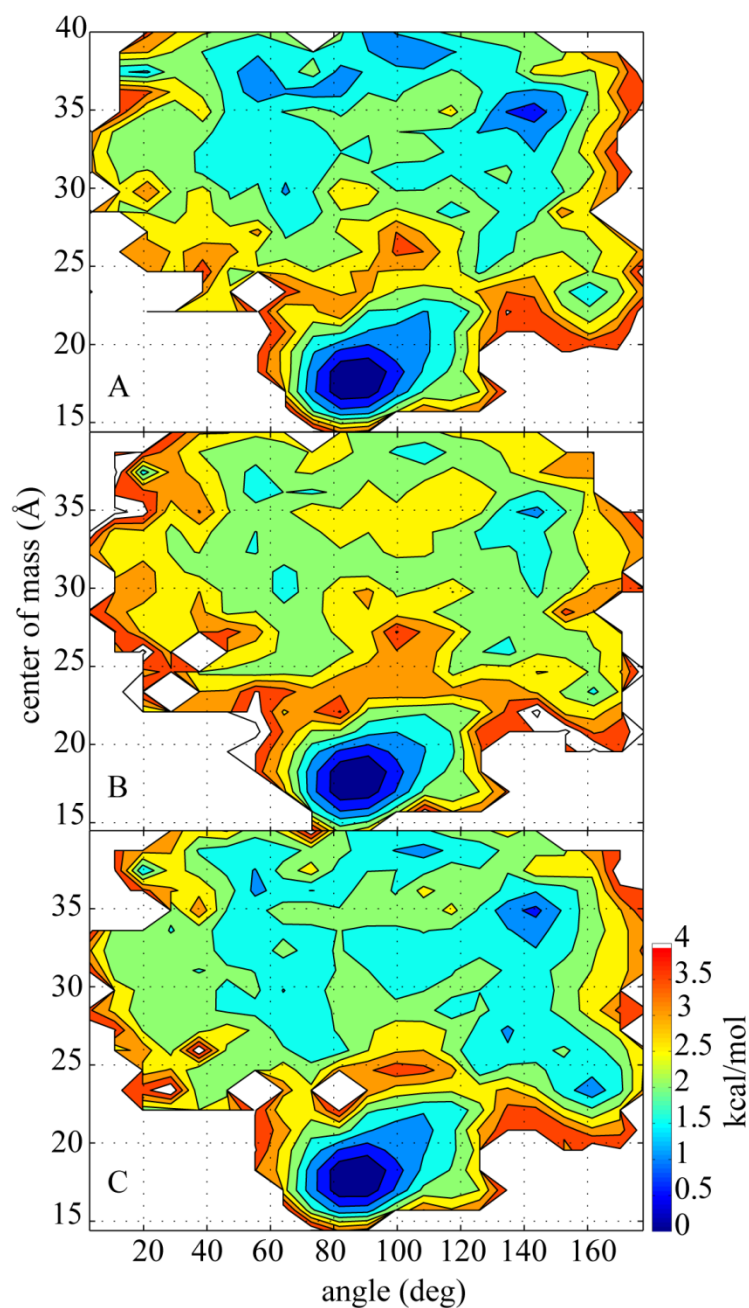


Figure 2-10: PMFs for mixed ensembles of NH₃⁺/CONH₂ and NH₃⁺/E11p/CONH₂ (A), NH₃⁺/E15p/CONH₂ (B) and NH₃⁺/D19p/CONH₂ (C) as a function of the N-terminal membrane insertion angle (in degrees) and distance of the peptide center of mass from the membrane center (in Å). The fraction of the protonated side chains is calculated based on pK_a shifts that are depicted in Figure 2-9 (see computational methodology section)

Figure 2-10 (cont'd)

experimental data, the inherent limitations of the implicit membrane model require further validation through additional computational and experimental studies. Furthermore, the sensitivity to the C-terminal tag raises the question of how the sequence context of the full-length fusion protein may affect the conformational sampling of the fusion peptide. A related question is how oligomerization affects the conformational sampling of fusion peptides. Both of these questions will be addressed in future studies.

2.6 Acknowledgements

We are grateful to Dr. David Weliky and Dr. Yan Sun for providing experimental data and Maryam Sayadi for helpful discussions. This work was financially supported from NSF grant MCB 0447799 and an Alfred P. Sloan fellowship (to MF). Access to computer resources at the High Performance Computer Center at Michigan State University is acknowledged.

3 EFFECT OF FLANKING RESIDUES ON THE CONFORMATIONAL SAMPLING OF THE INTERNAL FUSION PEPTIDE FROM EBOLA VIRUS

Adam J. Jaskierny¹, Afra Panahi², Michael Feig^{1,2*}

¹Department of Biochemistry & Molecular Biology

²Department of Chemistry

Michigan State University, East Lansing, MI, 48824

Adapted from

Proteins, 2011. **79**(4): p. 1109-1117

Afra Panahi contributed to running the simulations and analysis of the results and also produced several figures for the publication.

3.1 Abstract

Fusion peptides mediate viral and host cell membrane fusion during viral entry. The monomeric form of the internal fusion peptide from Ebola virus was studied in membrane bilayer and water environments with computer simulations using replica exchange sampling and an implicit solvent description of the environment. Wild type Ebola fusion peptide, the W8A mutant form, and an extended construct with flanking residues were examined. It was found that the monomeric form of wild type Ebola fusion peptide adopts coil-helix-coil structure with a short helix from residue 8 to 11 mostly sampling orientations parallel to the membrane surface. W8A mutation disrupts the helicity in the N-terminal region of the peptide, and leads to a preference for slightly oblique orientation relative to the membrane surface. The addition of flanking residues also alters the fusion peptide conformation with either a helix-break-helix structure or extended N and C-termini and reduced membrane insertion. In water, the fusion peptide is found to adopt structures with low helicity.

3.2 Introduction

Ebola virus belongs to the *Filoviridae* family and causes severe hemorrhagic fever in primates[150]. Viral infection requires the fusion between viral and host cell membranes. The membrane fusion process is mediated by fusion proteins that extrude from the viral membrane[101, 151-153]. Fusion peptides that are part of the fusion proteins are the key components that are in contact with the host cell membrane. In case of Ebola virus, Ebola glycoprotein (GP) is responsible for both receptor binding and membrane fusion[125]. This protein is composed of two sub-domains, GP1 and GP2, which are connected via a disulfide bond. Ebola GP shares many common features with other membrane fusion proteins such as HA1 and HA2 in influenza and gp120 and gp41 in HIV type 1. In all of these fusion proteins, the first subunit binds to the cell receptor while the second subunit mediates membrane fusion[126]. The role of the fusion peptide during the fusion process is further illustrated in Figure 3-1. The crystal structure of soluble GP2 reveals a trimer where a long coiled coil structure is surrounded by C-terminal helices. A key part of GP2 is the fusion peptide which interacts with the host cell membrane to induce fusion. In the soluble GP2 construct, the fusion peptide is missing[127]. Based on structures of influenza virus fusion protein, it is assumed that the fusion peptide is buried within the hydrophobic core of fusion protein in its inactive form, but becomes solvent exposed and able to insert into the host cell membrane once the fusion protein undergoes a structural change to its fusogenic form. Ebola fusion peptide (EFP) (G₅₂₄AAIGLAWIPYFGPAA₅₃₉) is thought to be in direct contact with the host cell membrane and is conserved within the virus family[128]. Unlike the N-terminal influenza and HIV fusion peptides[101] EFP is an internal fusion peptide that is located 22 residues from the N-terminus of

GP2. In the following, EFP residues will be referred to with a more convenient numbering scheme that starts at 1 for the first residue of the fusion peptide (G524 in GP2).

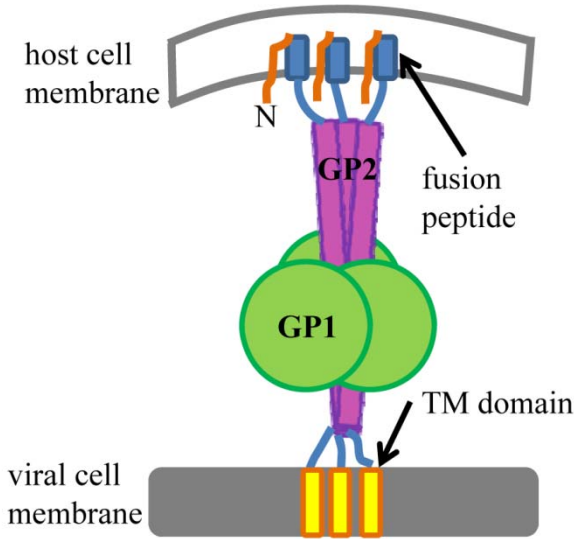


Figure 3-1: Schematic representation of Ebola fusion protein in fusogenic state. The globular domain GP1 is initially responsible for binding to the host cell receptor. The GP2 domain contains a helical bundle with the fusion peptide near the N-terminus.

Circular dichroism (CD) and infrared (IR) spectroscopy studies show that EFP_E (wild type EFP with one additional glutamic acid residue at the C terminus) has three states[129]: random coil in solution and either an α -helix or a β -sheet when bound to the membrane[129]. The secondary structure of the membrane-bound peptide depends on the presence of Ca^{2+} [154]. In the presence of Ca^{2+} a β -sheet structure is preferred while in the absence of Ca^{2+} helical structures are dominant[154]. In a different nuclear magnetic resonance (NMR) study of EFP it was also observed that the peptide adopts a random coil structure in aqueous buffers and more

defined structure in the presence of sodium dodecyl sulfate (SDS) micelles[130]. Furthermore, tryptophan fluorescent emission data suggests that W8 enters the hydrophobic core of SDS micelles and according to chemical shifts and distance constraints according to nuclear Overhauser effect (NOE) measurements obtained from ^1H NMR there is a short 3_{10} helix from I9 to F12 in the middle of the peptide while N- and C-termini appear to be less structured[130]. The presence of a short helix in the middle of the peptide suggests that this region is in contact with the membrane core whereas the presumably flexible N and C termini may interact more favorably with the solvent. The flexibility of the EFP structure is presumably related to the presence of glycines[155], but the glycines are also proposed to facilitate favorable insertion of EFP at the membrane head group-tail interface[130]. The secondary structure of EFP is apparently stabilized by an aromatic-aromatic interaction between W8 and F12 since the W8A mutation leads to a loss of helicity around I9 and a tendency to form helical structures between I4 and A8[130]. Another key residue is P14 which is believed to be essential for destabilizing the membrane and promoting fusion since the P14R mutant disrupts membrane destabilization but does not affect peptide membrane association[156]. Since proline residues are known to promote kinked helical structures [157] it was hypothesized that the mutation would disrupt a bent structure of EFP and result in a more linear, less destabilizing helical structure in the membrane[156]. The sequence context of the N- and C-termini is also essential for fusion activity. Addition of glutamic acid residues to N-and C-termini was found to reduce fusion activity[129]. In particular, addition of glutamate to the N-terminus of the peptide severely impairs peptide-vesicle interaction while addition of glutamate to the C-terminus of the peptide appears to only affect peptide association to membranes that lack phosphatidyl-inositol[129].

Computational studies of EFP have offered a more detailed view of EFP structure and dynamics. Molecular dynamics (MD) simulations of EFP in SDS micelles and water-hexane using explicit solvent show that during 10 ns of simulation, the NMR structure (Protein data bank (PDB) code 2RLJ[130]) is largely maintained in SDS micelles, especially the short central helix, while larger structural deviations were found at the water-hexane interface[130]. Another MD simulation of EFP with a relatively simple implicit membrane model [62] suggests that EFP adopts an α -helical structure with an interfacial membrane orientation at the center of the peptide while both termini are exposed to high- dielectric solvent[158].

Here, we report results from extensive replica exchange simulations of wild-type EFP and the W8A mutant with a more realistic implicit membrane model. Wild-type EFP is studied both as the 16-residue construct and with three additional residues from the GP2 sequence at either terminus in order to examine the effect of the sequence context on EFP structure and dynamics. The resulting conformational sampling is in overall good agreement with the recent MD studies[130, 158] and data from NMR[130] but also offers new insights into the effect of the flanking sequence on EFP structure and dynamics.

In the following, the computational methodology is described before results are presented and discussed.

3.3 Methods

Replica exchange MD simulations of EFP were carried out. Three systems were studied: wild type EFP with the sequence G₁AAIGLAWIPYFGPAA (G₁ corresponds to G524 in the GP2 fusion protein), the W8A mutant with the sequence GAAIGLAAIPYFGPAA, and a longer construct (EFP-X) with three extra residues from Ebola fusion protein on either side: QDE-

G₁AAIGLAWIPYFGPAA-EGI. Wild-type EFP was studied in an implicit membrane environment as well as a homogeneous dielectric to represent aqueous solvent. The W8A mutant and EFP-X were only simulated in the membrane environment. All simulations were carried out with the CHARMM program [159] version 34a2 in conjunction with the MMTSB Tool Set[138]. The lowest energy structure from the NMR ensemble of EFP in SDS micelles (PDB code 2RLJ[130]. was chosen as the initial structure for the EFP simulations. The initial structure for EFP-X was built by adding the extra residues with a random coil secondary structure using pyMol[148]. Both N- and C-termini were assumed to be neutral and modeled as NH₂ and CONH₂, respectively. To describe the peptide interactions, the CHARMM22[139] force field was used including the CMAP correction[24] To implicitly model the membrane, we used the heterogeneous dielectric generalized Born model (HDGB)[73] which has been developed in our group and applied previously in the simulation of membrane-bound proteins and peptides [74, 76, 78, 160]. The HDGB model describes the phospholipid bilayer with a continuously varying dielectric constant and surface tension along the membrane normal (z-axis). The thickness of the membrane measured as the distance between the center of the membrane and water with dielectric constant equal to 80 was set to 25 Å. The dielectric and non-polar profiles used in this paper were modified from the initial HDGB model[73] as described elsewhere[76] Here, we also used a modified periodic dielectric and non-polar profiles to achieve a finite peptide concentration in z-direction as described earlier[79]. Other implicit solvent parameters were set as described previously[74]. The simulation of EFP in implicit water was performed with GBMV [50, 51] using a solvent dielectric constant of 80 and the following parameters to obtain numerically stable simulations[161] $\beta=-12$, $S_0=0.65$, $C_0=0.3225$, $C_I=1.076$ and $D=-0.1418$. The

surface tension coefficient in the implicit water simulations was set to $0.015 \text{ kcal}/\text{\AA}^2$. Since the peptide is small, no cutoff was used for the calculation of non-bonded interactions. In all simulations, SHAKE[140] was applied to constrain bonds involving hydrogen atoms and an integration time step of 1.5 fs was used for all simulations to maintain stable simulations with the implicit solvent model[143]. Simulations were run in the canonical ensemble with Langevin dynamics[141] using a friction constant of 1 ps^{-1} to enhance sampling[142]. Each system was minimized for 50 steps of steepest decent method followed by 500 steps with the adapted-basis Newton-Raphson method[159]. The minimized structures were then used as initial structures to run temperature replica exchange molecular dynamics simulations [99, 100]. Eight replicas were run with temperatures spaced exponentially between 300 and 500 K. Exchanges between neighboring replicas were attempted after every 200 steps (0.3 ps). The acceptance ratio for successful exchanges was 23-37%. Total simulation times for different peptides ranged from 23-45 ns/replica (see Table 3-1) but the first 9 ns of each simulation were considered equilibration and discarded during subsequent analysis.

Table 3-1: Overview of EFP simulations

System	Environment	Abbreviation	Simulation time/replica
GAAIGLAWIPYFGPAA	Membrane	EFP	37ns
GAAIGLA ^A IPYFGPAA	Membrane	W8A	36ns
GAAIGLAWIPYFGPAA	Water	EFP-AQ	23ns
^{QDE} GAAIGLAWIPYFGPAA ^{EGI}	Membrane	EFP-X	45ns

Potentials of mean force (PMF) at 300K were calculated from the replica exchange simulations using weighted histogram analysis (WHAM)[144, 145] to include weighted sampling from higher temperature replicas in the overall free energy surface. For membrane-interacting peptides, the peptide insertion depth and angle were used as the primary reaction coordinates in the construction of the PMFs. For the wild-type EFP and W8A mutant, the insertion depth was calculated as the distance of the center of mass (COM) of the entire peptide from the center of the membrane ($z=0$) along the z -direction. For EFP-X, the center of mass was calculated as above but without considering the three additional residues at each terminus. Peptides interacting with a different implicit membrane due to periodic replication or with opposite leaflets of the same membrane after spontaneous crossing of the bilayer were mapped onto a common scale by calculating the distance to the closest membrane for all structures. The insertion angle was calculated as the angle between membrane normal (z -axis) and the vector that connects the $N(i+4)$ and $O(i)$ atoms of the backbone for residues i from 3 to 6 for wild type and W8A mutant and residues 7 to 10 for EFP-X corresponding to the same sequence as in EFP. The conformation of the peptides was further analyzed by calculating α -helicity as a function of residue number. Two criteria were used to define the presence of α -helical structure: 1) An $O(i) - H-N(i+4)$ distance of less than 2.6 Å [146] and 2) backbone torsion values in the range of $-70 \leq \phi \leq -50$ and $-35 \leq \psi \leq -55$. An $O(i) - H-N(i+3)$ distance of less than 2.6 Å was also considered as the criterion for defining 3_{10} helicity.

The PMF for EFP in water was projected onto two reaction coordinates: helicity based on hydrogen bond formation between $O(i) - H-N(i+4)$ as described above and radius of gyration. Representative structures for each system were obtained through clustering based on mutual similarities as measured by root mean square deviations (RMSDs).

For analysis of EFP structures at the membrane interface, the superposition scheme used during the RMSD calculation was modified to preserve peptide orientations with respect to the z -axis and distinguish between different membrane insertion depths. This was achieved by allowing only transitions along x and y directions and rotation around the z -axis[78]. Using the modified superposition scheme, the K-means method was then used to cluster conformations based on RMSD with a radius of 3 Å and representative structures were extracted based on proximity to the center of each cluster. The standard superposition procedure was applied for clustering conformations from the simulation in implicit aqueous solvent.

3.4 Results

Replica exchange simulations were used to study the dynamics of EFP in implicit membrane/solvent environments. The wild-type EFP, a W8A mutant, and the extended sequence EFP-X were simulated in a membrane environment. Wild-type EFP was also simulated in aqueous solvent. The simulations are summarized in Table 3-1.

3.4.1 Conformational sampling of EFP, W8A, and EFP-X in membrane bilayer

All of the systems exhibited significant dynamics involving different peptide conformations and different interactions with the model membrane but most of the sampled structures displayed at least some degree of helicity. Conformational sampling was characterized in more detail based on a 2D-PMF as a function of the N-terminal insertion angle and the distance of the center of mass from the membrane center (see Figure 3-2 A-C). Conformations with a center of mass distance of less than 25 Å were considered to be associated with the membrane. The corresponding structures were clustered and representative structures near the centers of the most

populated clusters are shown in Figure 3-2. Furthermore, average properties for membrane-associated peptides are reported in Table 3-2.

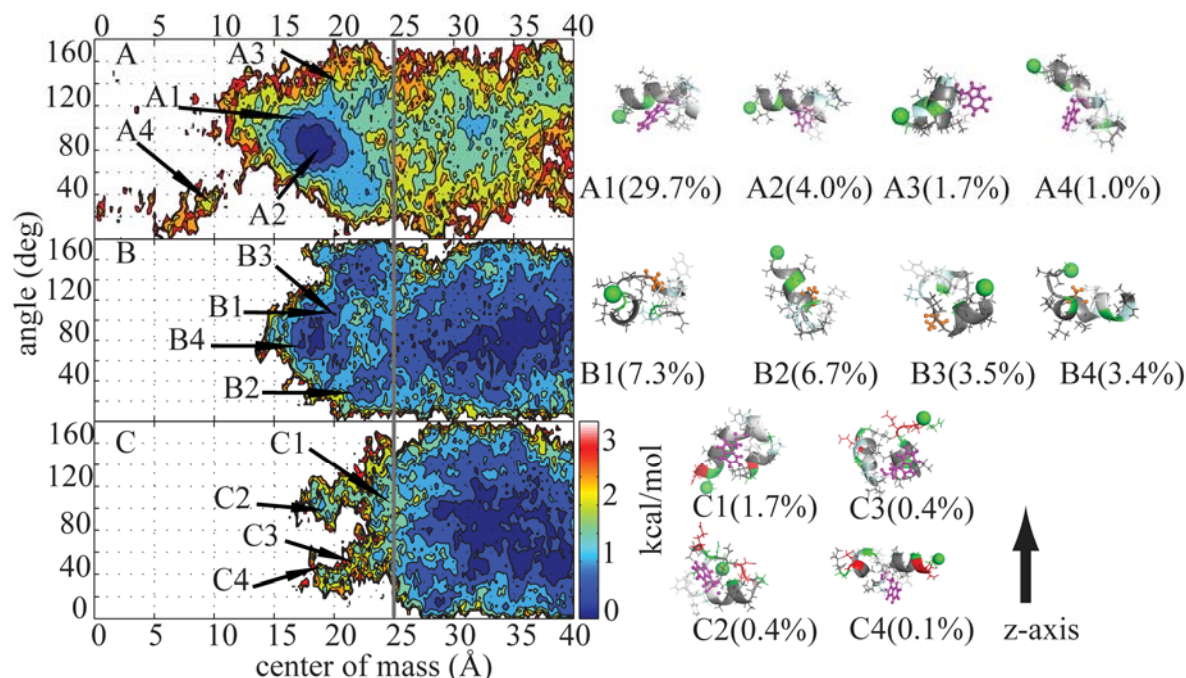


Figure 3-2: Left panel: Potentials of mean force (PMFs) for EFP (A), W8A (B), and EFP-X (C) from weighted-histogram analysis (WHAM), projected onto insertion angle (in degrees) and distance of the peptide center of mass from the membrane center (in Å). Different colors reflect relative free energies as indicated by the color bar (in kcal/mol). Right panel: Dominant structures for simulations with EFP (A1-A4), W8A (B1-B4) and EFP-X(C1-C4) from clustering. The backbone is shown as a ribbon and side chains are shown as stick representation. Charged residues are colored in red, glycines in green, prolines pale cyan, aromatic residues in white and hydrophobic residues in gray. The N-terminus is marked by a green sphere. Residue 8 is shown in ball and stick representation and colored magenta (W8) and orange (A8 in mutant). The z-axis is pointing up in all figures. All figures were generated with pyMol[148].

Table 3-2: Population of inserted peptides, average center of mass (COM) insertion depth, insertion angles, and helicity fractions from simulated EFP structures at 300 K for membrane-associated conformations (defined as having COM insertion depths of less than 25 Å). Analysis of EFP-X was only applied to the core EFP sequence. Standard deviations are reported in parentheses. The standard errors based on block averaging are about 0.3 Å for COM, 10.5 degrees for insertion angles and 0.03 for helical fractions.

Simulation	Population of inserted peptides	COM insertion[Å]	Insertion Angle [deg]	Helicity fraction (from H-bonds)	Helicity fraction (from ϕ/ψ sampling)
EFP	72.5%	19.05 (2.87)	85.64 (25.74)	0.345 (0.14)	0.214(0.10)
W8A	25.6%	20.84 (2.45)	91.80 (41.09)	0.229 (0.12)	0.161(0.09)
EFP-AQ	-	-	-	0.154(0.14)	0.047(0.07)
EFP-X	2.2%	22.49 (2.18)	92.49 (31.27)	0.362 (0.13)	0.262(0.11)

For wild-type EFP, there is a deep minimum at insertion depths of 15-22 Å and insertion angles of 60-100° (Figure 3-2 A) indicating that the peptide tends to lie parallel to membrane surface and strongly prefers interfacial association. The most dominant structure is A1, which consists of a bent helix with a kink at P10 and a less-ordered C-terminus. A second highly populated cluster (A2) adopts a straight helix for residues 3 to 12 also followed by a disordered C-terminus. A V-shaped orientation of EFP with the N- and C-termini pointing towards the membrane was observed as well (A3). In this structure, there is a short helix from residues 3 to 7

followed by a kink with a helix from residues 9 to 15. Interestingly, there was also a small population of trans-membrane (TM) orientations with an insertion depth of 0-10 Å and insertion angles of 0-40° (A4 in Figure 3-2). The TM structures are mostly helical with an oblique orientation and with the N-terminus facing away from the membrane. The sampling of predominantly helical structures with a kink near P10 is consistent with experiments[156] and a previous simulation[158]. The observation of spontaneous TM insertion may suggest a mechanism by which the host cell membrane is disrupted and membrane fusion is facilitated but this point will be discussed in more detail below.

The conformational sampling of W8A was significantly different from wild-type EFP. First, W8A did not associate as strongly with the membrane as the wild-type peptide. Secondly, membrane-associated structures were overall less helical (see B1-B3). However, helical structures with parallel membrane orientation and insertion depths of 15 to 20 Å similar to the dominant wild-type conformation were also observed in the mutant (B4) but with a lower probability. The loss of helicity in the middle of the dominant structures (B1-B3) as a result of the W8A mutation is in agreement with earlier experimental studies[130] and is presumed to be a consequence of the lost interactions between W8 and F12. The loss of helicity in turn results in a more flexible peptide and broader sampling of different conformations. The relative populations indicated in Figure 3-2 suggest that the mutated peptide samples a broader range of different structures with similar frequency.

The sampling of EFP-X with three additional residues at the N and C-termini also differs significantly from the shorter EFP construct (see Figure 3-2 C). Membrane association is even less favorable than in the W8A mutant, presumably as a result of the added negatively charged aspartate and glutamate residues at the N-terminus and another glutamate at the C-terminus. The

structures that are associated with the membrane appear to cover a broader ensemble consisting of oblique and parallel orientations relative to the membrane. The representative structure from the most populated cluster has an insertion angle of 110° and an insertion depth of 24 Å (see C1 in Figure 3-2). It is mostly helical with kinks at residues P10 and P14. The second most populated membrane-associated conformation (C2) is more deeply inserted and consists of a short helix at residues 6 to 12 but has disordered N- and C-termini. The short helix lies parallel to the membrane surface as in A1 but the helix does not extend as far in C2 as in A1. C3 is similar to C1 but it is turned around with respect to the membrane allowing the turn region to interact with the membrane and the termini to be exposed towards the aqueous solvent. Finally, C4 has a reduced helical structure but lies parallel to the membrane surface similar to the A1 conformation for EFP.

3.4.2 EFP in aqueous solvent

Conformational sampling of EFP in aqueous solvent was also investigated to better understand the effect of the membrane. A PMF based on the average α -helicity and radius of gyration is shown in Figure 3-3. There is a broad minimum region that reaches from helical fractions of 0 to 0.7 with two preferred states: compact structures with partial helicity (D2-D4 in Figure 3-3) and fully extended conformations with no helical structure (D1). The helical structures have only short helices either from A3 to A7 (D2) from I9 to Y11 (D3) (a short 310 helix) or from I4 to Y11 (D4). A loss of secondary structure in aqueous buffer was previously observed[130] and our simulation results confirm that the membrane stabilizes the secondary structure of EFP.

Average insertion depths, insertion angles, and helicity based on hydrogen bonding and dihedral angles are reported in Table 3-2. For membrane simulations, only membrane associated

structures were considered. Again, it can be seen that EFP is inserted most and EFP-X is inserted least according to the average insertion depths of the peptide centers of mass. The average N-terminal insertion angle is near 90° for all peptides. However, while the average reflects the dominant structures for EFP and W8A it is a result of an average over populations with different insertion angles for EFP-X (see also Figure 3-2)

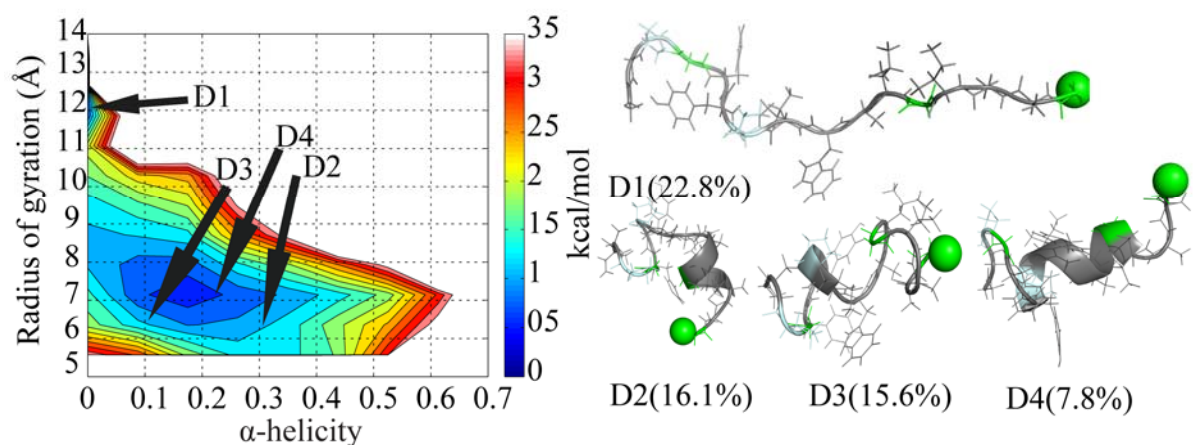


Figure 3-3: Left panel: Potential of mean force (PMF) of EFP-AQ from weighted-histogram analysis (WHAM), projected onto average helicity based on hydrogen bond criterion and radius of gyration (in Å). Different colors reflect relative free energies as indicated by the color bar (in kcal/mol). Right panel: Dominant structures for simulations with EFP-AQ from clustering as in Figure 3-2.

3.4.3 Helical propensity in EFP

Experimental data suggests that EFP has a tendency to form helical structures in the presence of the membrane[130]. From our simulation data, helicity was calculated according to two criteria: based on hydrogen bonding and backbone ϕ/ψ -sampling (see Methods section). Overall, average

helicity is largest for EFP-X and slightly lower for EFP (see Table 3-2). Helicity is significantly reduced in the W8A mutant and even further for EFP in aqueous solvent. In order to examine helical propensity in more detail, we have calculated helicity as a function of residue along the EFP sequence for all of the peptides studied here. The results shown in Figure 3-4 indicate that all peptides have a strong tendency to form helical structures in the central region of the peptide (residues 8-10) consistent with NMR chemical shift measurements[130]. Furthermore, there is significant helical propensity at the N-terminus. Mutating W8 to alanine reduces the helicity in the N-terminal domain but does not cause any significant change in helicity in the C-terminal part of the peptide. This observation is also consistent with experiment[130]. There are two breaks in helicity close to the proline residues (P10 and P14). Since there is no amide hydrogen on P10 and P14 these two residues are not able to form hydrogen bonds with L6 and P10. The role of proline in inducing helical breaks was also observed experimentally and is hypothesized to be important for fusion activity in Ebola[155] as well as influenza fusion peptide[134]. Including extra residues at the N- and C-termini of EFP changes the overall structure and orientation of EFP as described above but does not appear to significantly change the α -helicity profiles.

Formation of 3_{10} -helices was also investigated based on an $i, i+3$ hydrogen bond criterion (see Methods section). In all four simulations, the peptide has a high propensity to form 3_{10} -helix at W8. It should be noted that a residue may be in both α - and 3_{10} -helical form if bifurcated hydrogen bonds are present. Formation of a 3_{10} -helix in the central region of the peptide was also observed experimentally[130]. In addition to W8, G13 also shows a high propensity to form 3_{10} -helices. According to Figure 3-4 C, EFP appears to prefer α -helical

structures from G1 to I4 3_{10} -helices. This preference for α -helices is even more pronounced for EPF-X whereas W8A and EFP-AQ have similar propensities towards either helical form.

3.5 Discussion and conclusion

In this study we describe results from simulation of EFP in the presence of membrane and water environments. An implicit description of the environment allowed us to reach much longer time scales than what would be possible with explicit lipid/water simulations. In particular, the implicit membrane model avoids the need for relatively slow lipid relaxation in response to different peptide conformations. Further sampling enhancement is gained from using replica exchange sampling and a reduced friction coefficient in Langevin dynamics. While the exact time scales covered in our simulations are difficult to determine, we estimate that the μ s-ms range is reached despite a nominal length of the simulations of only tens of nanoseconds.

On the other hand, the use of an implicit membrane model introduces certain approximations and in particular focuses on the mean-field effect of the membrane environment while neglecting specific molecular details of peptide-membrane interactions. While this may appear to be a serious limitation, we have used the same approach previously to study membrane-bound influenza virus fusion peptide[78] and phospholamban[76] In both cases, we found very good agreement with experimental data. In fact, the agreement appears to be better than with previous explicit lipid simulations presumably because of the much longer time scales that could be covered with the implicit solvent model. In particular, the dominant conformation predicted for membrane-bound influenza virus fusion peptide predicted by our simulations was recently confirmed experimentally[124]. Furthermore, similar but simpler implicit membrane models than the HDGB model used here, have also been very successful in describing the

conformational ensemble of other membrane-bound peptides in good agreement with experiment [55, 62, 65, 136]. This suggests that the approach taken here is reasonable for obtaining realistic conformational ensembles of membrane-bound EFP. However, it is clear that a full understanding of fusion peptide-membrane interactions will eventually also have to include simulations with explicit lipids.

In our simulations, EFP samples a variety of conformations with different degrees of partial helicity. Sampling of helical structures is reduced in the W8A mutant and in water while extending EFP by three residues at the N- and C-termini slightly enhances helicity. EFP by itself mostly prefers a central helix that is oriented parallel to the membrane. In contrast, the addition of extra residues from the GP2 sequence context leads to significant sampling of helix-break-helix motifs where N- and C-terminal helices are oriented more obliquely with respect to the membrane normal. A conformation with a central helix parallel to the membrane is also observed, but the helix is shorter in the EFP-X construct than in the EFP sequence (see C2 in Figure 3-2). The extra residues also significantly reduce the affinity of the peptide for the peptide as a result of charged residues at both ends. These results indicate that the sequence context is important in determining the conformation of membrane-bound EFP. Similar conclusions were found for influenza virus fusion peptide (IFP) where the addition of an experimentally used C-terminal tag also appeared to affect conformational sampling and orientation relative to the membrane[78]. Our findings therefore strongly suggest that implications based on structural studies of just the fusion peptide alone should be treated with caution and that it may be necessary to study fusion peptides in the context of the fusion protein to fully understand their interactions with membranes and ultimately the mechanism by which they exert their fusogenic activity.

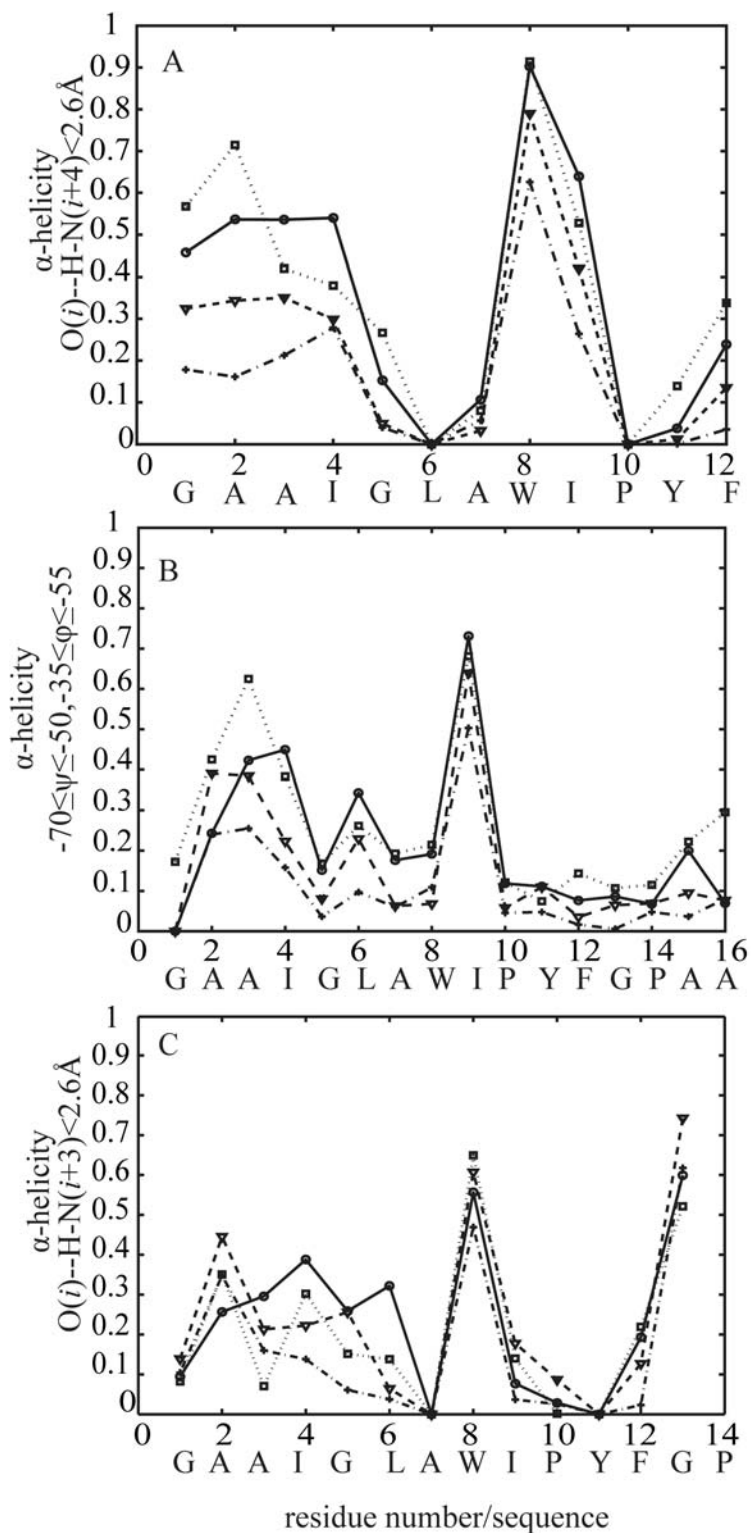


Figure 3-4: Average fraction of helical conformations as a function of residue for EFP (circle), W8A (triangle), EFP-X (square) and EFP-AQ (cross). A: Helicity is defined by $O(i) - H-N(i+4)$

Figure 3-4 (cont'd)

hydrogen bonding ($< 2.6 \text{ \AA}$ distance); B: Helicity is defined by $-70 \leq \psi \leq -50$ and $-35 \leq \phi \leq -55$; C: 3_{10} - helicity is defined by $O(i) - H-N(i+3)$ hydrogen bonding ($< 2.6 \text{ \AA}$ distance)

EFP is an internal fusion protein with extra residues from the GP2 fusion protein on both the N- and C-termini. Under the assumption that neither the N- or C-terminal part of GP2 inserts deeply into the host cell membrane the type of EFP conformations that are possible in the context of GP2 are constrained. Conformations where both N- and C-termini are inserted into the membrane (like A1 or C1) appear to be unlikely. Furthermore, the trans-membrane (TM) structures observed for the short EFP construct would be ruled out because it would require the 22 N-terminal residues of GP2 to either cross the membrane or be inserted into the membrane as well. Membrane insertion or crossing is unlikely due to the presence of charged and polar residues at the N-terminus of GP2. More likely conformations of EFP in the context of GP2 appear to be C2, C3, or C4 all of which would allow N- and C-terminal residues to connect to the peptide outside the membrane or just at the membrane-water interface.

It is also interesting to compare the structures observed for EFP with earlier results for IFP. In our simulations, IFP samples largely helix-break-helix structures as confirmed by recent new experimental data [78, 123, 124].

Those structures resemble the structures C1 and C3 seen for EFP-X but in the IFP simulations, those conformations were found to lie mostly parallel to the membrane while the EFP helix-break-helix structures appear to be oriented perpendicular to the membrane. It

therefore seems from this study and our previous work that there are common structural elements shared between viral fusion peptides. This would suggest a common fusogenic mechanism, but more work is clearly needed to come to definite conclusions in this regard.

Another issue that has been so far neglected is that viral fusion proteins and in particular their fusion peptides are known to oligomerize. It will therefore be critical to expand future studies to fusion peptide oligomers to better understand the membrane fusion process. The implicit membrane methodology applied here could also be used to study fusion peptide oligomers albeit with more extended conformational sampling because of the additional degrees of freedom.

It is our hope that the present study motivates not just further computational studies of viral fusion studies but also stimulates a closer experimental look at the effect of the fusion protein context on the conformational sampling of the fusion peptide in Ebola and other viruses.

3.6 Acknowledgements

This work was supported in part through NSF grant MCB 0447799 and an Alfred P. Sloan fellowship (to MF). Computer resources at the High Performance Computer Center at Michigan State University as well as an NSF TeraGrid computer allocation (TG-MCB090003) are acknowledged.

4 DYNAMIC HETEROGENEOUS DIELECTRIC GENERALIZED BORN (DHDGB): AN IMPLICIT MEMBRANE MODEL WITH A DYNAMICALLY VARYING BILAYER THICKNESS

Afra Panahi¹, Michael Feig^{*1,2}

¹Department of Chemistry

²Department of Biochemistry & Molecular Biology

Michigan State University, East Lansing, MI, 48824

Adapted with permission from

J. Chem. Theory. Comput., DOI: 10.1021/ct300975k

Copyright 2013 American Chemical Society

4.1 Abstract

An extension to the heterogeneous dielectric generalized Born (HDGB) implicit membrane formalism is presented to allow dynamic membrane deformations in response to membrane-inserted biomolecules during molecular dynamic simulations. The flexible membrane is implemented through additional degrees of freedom that represent the membrane deformation at the contact points of a membrane-inserted solute with the membrane. The extra degrees of freedom determine the dielectric and non-polar solvation free energy profiles that are used to obtain the solvation free energy in the presence of the membrane and are used to calculate membrane deformation free energies according to an elastic membrane model. With the dynamic HDGB (DHDGB) model the membrane is able to deform in response to the insertion of charged molecules thereby avoiding the overestimation of insertion free energies with static membrane models. The DHDGB model also allows the membrane to respond to the insertion of membrane-spanning solutes with hydrophobic mismatch. The model is tested with the membrane insertion of amino acid side chain analogs, arginine-containing helices, the WALP23 peptide, and the gramicidin A channel.

4.2 Introduction

Membrane proteins account for about 30% of the proteins in a cell and are responsible for a large number of biological activities[162-164]. Therefore, understanding their structure and dynamics is essential for gaining complete knowledge of life processes occurring in a cell. While experimental techniques have provided much of the available structural information, computer simulations of biological systems and of membrane proteins in particular have been able to provide critical information about the dynamics and energetics of these macromolecules[165-167].

Molecular dynamics (MD) simulations of membrane-interacting proteins have remained a challenging task due to the complex nature of the membrane bilayer[69-71]. Slow relaxation times of the lipid molecules require long simulations at significant computational cost to achieve statistical convergence when lipids and solvent molecules are considered explicitly[168]. In order to address this issue, alternative models have been considered that do not involve an all-atom representation of the water and lipid molecules [33, 35, 38-40, 48, 50, 169]. In these models, the solvent degrees of freedom are approximated by a mean-field term that only considers the average effect of the environment rather than detailed specific interactions between the solvent and the solute [33, 35, 38-40, 48, 50, 169]. In most of these implicit solvent models, the environment is considered as a dielectric continuum where the electrostatic solvation free energy is obtained by numerical solution of the Poisson-Boltzmann (PB) equation [38-40, 169]. The PB model is easily extended to heterogeneous environments and can describe membranes as systems consisting of layers with different dielectric constants [38-40, 170]. The PB-based electrostatic solvation free energy can then be combined with an implicit non-polar term to obtain the total solvation free energy. Non-polar terms are commonly calculated using the

solvent-accessible surface area (SASA)[39, 53, 171] but may also include an additional implicit solute-solvent van der Waals term [172, 173].

Because direct solution of the PB equation is computationally expensive and numerically problematic for use in molecular dynamics simulations [41, 174-176], the generalized Born (GB) model is often used as an analytical approximation [45]. In the GB formalism, the electrostatic contribution to the total solvation free energy is approximated as a sum of screened pairwise interactions between the charges of a given molecule. Many formulations of GB have been proposed with different levels of accuracy and speed compared to the PB equation [35]. The GB formalism has also been extended to heterogeneous environments such as membranes [55, 62, 73]. In early implementations of implicit membrane GB models, the membrane dielectric was assumed to be equal to the solute dielectric so that a simple two dielectric model could be retained [55, 62]. A more detailed model involving a variable dielectric constant was proposed by us [73]. This so-called heterogeneous dielectric generalized Born (HDGB) model was motivated by dielectric profiles obtained from explicit solvent simulation of lipid bilayers[72] and implemented as an extension of the GBMV (GB with molecular volume) model [51]. The HDGB model does not require the protein to have the same dielectric as the membrane interior and therefore allows for a polarization response of the lipid bilayer. HDGB was tested for a variety of protein and peptide systems and was found to closely match PB results [73], to reproduce energetics from explicit solvent/lipid simulations [73, 76], and to generate conformational sampling in good agreement with experimental data [76, 78, 79, 177].

All of the implicit membrane models proposed so far, including the HDGB model, assume that the membrane bilayer does not vary over time. In contrast, experiments and explicit solvent simulations suggest that membrane bilayers are dynamic and can deform in response to inserted

biomolecules [178]. In the presence of a charged particle, for instance, water molecules and lipid head groups may accompany the charge in the membrane interior and keep it partially solvated [77, 93, 179]. As a result, the penalty for inserting charged groups into the hydrophobic bilayer is greatly reduced compared to what is predicted by implicit membrane models that assume a fixed bilayer width. For instance, the penalty for inserting a positive arginine side chain into an explicit dipalmitoylphosphatidylcholine (DPPC) bilayer was reported to be 17.8 kcal/mol [93] while implicit membrane models overestimated the insertion penalty to be as high as 44 kcal/mol [88].

Within the context of an implicit membrane formalism, membrane deformations can be included by allowing the dielectric profile to deviate from standard slab geometries and by adding the intrinsic cost of such membrane deformations to the free energy of solvation. The energy required to deform a model membrane can be obtained from elasticity theory [81, 82, 86, 87, 180, 181]. In this theory, the membrane is considered as an elastic sheet with material properties that describe its resistance towards different deformations such as bending, stretching, and compression [86, 182, 183]. Choe *et al.* developed such an approach to determine the minimum-energy membrane geometry, deformation energy, and solvation free energy for a membrane-embedded protein [88]. This formalism has been tested for a poly-leucine trans-membrane (TM) helix with a positive arginine residue at its center [88]. The calculated insertion energies were found to be in good agreement with previously calculated explicit solvent free energies [93]. An issue with this approach is that the optimal membrane geometry in the vicinity of the inclusion is not known *a priori* and has to be obtained by a relatively costly minimization of the overall energy [88]. The performance of this approach can be improved by using a search algorithm to choose a reasonable initial geometry [89]. However, the membrane geometry minimization followed by the use of PB for calculating the electrostatic contribution to the

solvation free energy essentially limits its application to those cases where protein structure and membrane dynamics are decoupled [88, 89]. In other words, the protein structure is not allowed to vary in response to membrane deformations in this scheme.

In order to obtain a fully dynamic model where both, the membrane shape and solute conformations, can vary in a concerted fashion during MD simulations, we have extended the HDGB model[73], to include membrane deformations. In the following, a description of the theory and simulation methods is given before first applications to the membrane insertion of amino acid side chain analogs and the dynamics of membrane-embedded peptides are discussed.

4.3 Theory

An extended Hamiltonian approach was followed to introduce the membrane deformation through extra degrees of freedom (DOF) that are propagated along with the rest of the system. The modified Hamiltonian (H) is written in terms of the atomic (x^{3N}) and the membrane DOF (S^{Ns}) as follows

$$H(x^{3N}, S^{Ns}) = E_{pot}(x^{3N}, S^{Ns}) + E_{kin} \quad 4.3.1$$

with Ns being the number of DOF that represent the membrane. E_{pot} and E_{kin} are the potential and kinetic energies, respectively. The kinetic energy of the extended system is simply written as the sum of the atomic kinetic energies and the new DOFs (Eq. 4.3.2) which we are going to refer to from hereon as the membrane deformation parameters (MDP).

$$E_{kin} = \frac{1}{2} \left(\sum_{i=1}^{N_{atom}} m_i v_i^2 + \sum_{s=1}^{N_s} m_s v_s^2 \right) \quad 4.3.2$$

v_i and v_s are the velocities of atom i and the membrane associated DOFs, and m_i and m_s are the corresponding masses.

The potential energy of the extended system involves a modified implicit solvation term. Following the procedure by Choe *et al.* [88], the solvation energy of such a system is written in terms of the electrostatic (ΔG_{GB}), the non-polar (ΔG_{np}), and the membrane deformation energy (ΔG_{mem}) as shown in Eq. 4.3.3. The first two terms depend on the MDPs in addition to the atomic coordinates while ΔG_{mem} is only a function of the MDPs.

$$\Delta G_{sol}(x^{3N}, S^{N_s}) = \Delta G_{GB}(x^{3N}, S^{N_s}) + \Delta G_{np}(x^{3N}, S^{N_s}) + \Delta G_{mem}(S^{N_s}) \quad 4.3.3$$

In the following, the calculation of each of these three terms is described in more details:

4.3.1 Membrane deformation energy (ΔG_{mem})

Following Choe *et al.* [88, 89], we used elasticity theory to estimate the membrane deformation energy. In this model, ΔG_{mem} for one leaflet is written in terms of three main deformation moduli as follows[81, 82, 86-88]

$$\Delta G_{mem} = \iint \frac{1}{2} \left(\frac{K_a}{d_0} u^2 + \frac{1}{2} K_c (\nabla^2 u)^2 + \frac{1}{2} \alpha (\nabla u)^2 \right) d\Omega \quad 4.3.4$$

In Eq. 4.3.4, Ka , Kc , and α are the compression, bending and surface tension moduli, respectively. d_0 represents the width of the unperturbed membrane while u denotes the deviation of one leaflet from its unperturbed width, which we are going to refer to as leaflet deformation. In our definition, positive and negative deformation values stand for expansion and compression along the membrane normal, respectively. The double integral is taken over the entire plane of the membrane. The values for each modulus is taken from previous work [88].

We make the assumption here that the membrane deformation energies in the top and bottom leaflets are independent of each other and that both terms can therefore be simply added to yield the total deformation energy for a bilayer. Furthermore, the overall geometry of the molecules inserted in to the membrane is approximated by a cylindrical inclusion. For the purpose of deformation energy calculation we followed the assumption made by Choe *et al.*[88] and considered the axis of the cylinder to be parallel to the membrane normal. The curve along which the membrane meets the inclusion is going to be referred to as the contact curve. The projection of the contact curve on the xy plane, or any planes parallel to the xy plane, forms a circle called contact circle.(cf. Figure 4-1A)

In order to obtain an optimal membrane configuration, we minimize the membrane deformation energy in Eq. 4.3.4 with respect to u . This leads to the following equation[82, 86-89]

$$\nabla^4 u - \frac{\alpha}{K_c} \nabla^2 u + \frac{2K_a}{d_0^2 K_c} u = 0 \quad 4.3.5$$

To solve Eq. 4.3.5, four boundary conditions are required. Many membrane-bound peptides such as TM helices for example, are approximated well with a simple cylindrical model. In this

case, we can rewrite Eq. 4.3.5 in terms of r (the distance from the center of the cylinder) and θ (the angle around the circle where the membrane meets the cylinder). Following Choe *et al.*, [88] we assume that the membrane will be flat with its unperturbed thickness far from the solute.

Therefore,

$$\begin{aligned} u &= 0 \quad r \rightarrow \infty \\ \frac{\partial u}{\partial r} &= 0 \quad r \rightarrow \infty \end{aligned} \tag{4.3.6}$$

The other two boundary conditions are the leaflet deformation and the contact slope around the contact curve.

To estimate the boundary conditions in the vicinity of the solute, we chose five evenly distributed points around the cylinder model for each leaflet. These points will be referred to as $S = \{S_1, S_2, \dots, S_5\}$ for the top leaflet and $S' = \{S_6, S_7, \dots, S_{10}\}$ for the bottom leaflet. They represent the leaflet deformation at $\theta = \{0, 2\pi/5, \dots, 8\pi/5\}$ (cf. Figure 4-1 A) i.e. $S_I = u(r_0, \theta=0)$ where r_0 is the radius of the cylinder. In our model we chose $d_0=50 \text{ \AA}$ as the unperturbed width of a DPPC model membrane as originally parameterized for the static HDGB model [73]. This width approximately corresponds to the width of the entire membrane including its head groups.

Membrane deformation occurs when the contact points vary along z away from the equilibrium value for a planar bilayer. All points are allowed to vary independently, thereby allowing angular-dependent complex deformation geometries. The leaflet deformation for any arbitrary point is generated through cubic spline interpolation from the discrete points, $f(S, \theta)$.

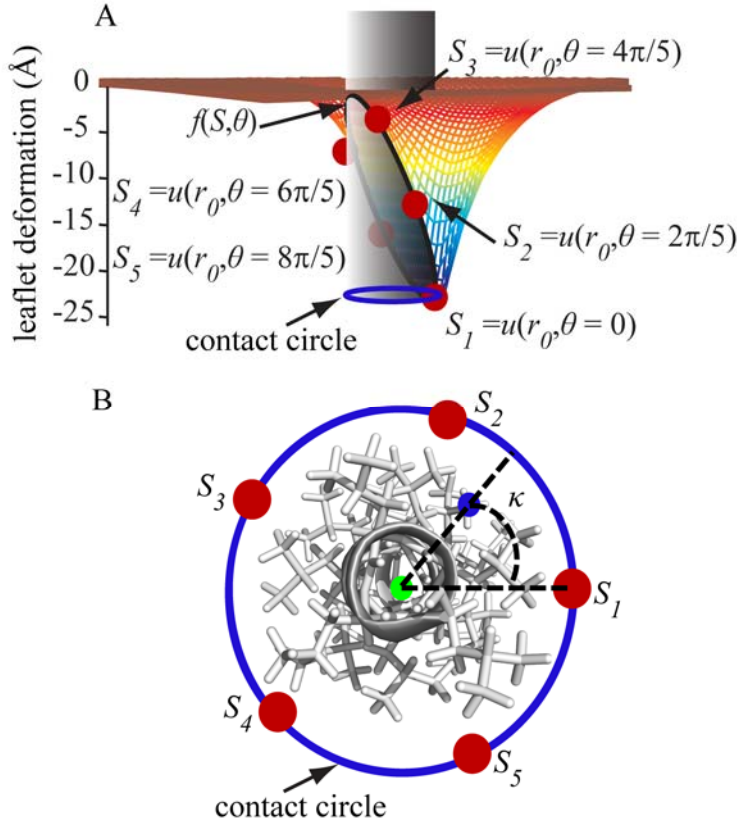


Figure 4-1: (A) A schematic representation of an arbitrary deformation around a cylindrical solute inclusion in the top leaflet of the membrane bilayer. Fixed points at chosen angles (θ) around the contact curve where the membrane meets the solute cylinder (shown in gray) are indicated as red points. A cubic spline interpolation ($f(S, \theta)$) of the leaflet deformation vs. θ is defined to calculate the deformation at any arbitrary point around the contact curve. The projection of the contact curve on an arbitrary plane parallel to the xy plane is shown with a blue circle. (B) Top view of the helix modeled with a cylinder in panel A. For an arbitrary atom marked with a blue point on the helix, the center of the contact circle is calculated as described in Eqs. 4.3.8 and 4.3.9 and shown with a green point. Then κ is calculated as explained in Eq. 4.3.10.

For any given point on the contact curve, assuming that the radius of the cylinder is fixed, the position of the point on the contact circle on the xy plane can be specified with an angle φ . This angle is defined as the angle between the vector that connects the point to the origin and the x axis. Following Choe *et al.*[88] the contact slope at each point on the contact curve with angle φ is calculated as

$$\left. \frac{\partial u(r_0, \theta)}{\partial r} \right|_{\varphi} = - \left. \frac{f(S, \theta)}{r_0} \right|_{\varphi} \quad 4.3.7$$

Given the boundary conditions from Eqs. 4.3.6 and 4.3.7, Eq. 4.3.5 was solved numerically using finite difference method with 20 angular and 3000 radial grid points. From the calculated $u(r, \theta)$ the corresponding values for ΔG_{mem} were then obtained via Eq. 4.3.4. The leaflet deformations at different angles θ were changed from -25 Å to 10 Å with 5 Å increments (total of 8 values) and all possible combinations were sampled exhaustively. The corresponding deformation energy for each combination was calculated, which gave rise to a total of 8^5 calculations. The resulting energies were tabulated so that ΔG_{mem} and smooth derivatives could be obtained for any given deformation via cubic spline interpolation. Because the test systems considered here only involved TM helices and small amino acid side chain analogs, we tabulated deformation energies only for fixed radii r_0 of 7.5 Å (for the helical systems) and 4.0 Å (for the small molecules). In order to simulate other systems, the table would have to be recalculated for the appropriate radius of an approximate cylinder. Alternatively, one could also pre-calculate the deformation energies for discrete values of r_0 and interpolate in the r_0 dimension as well. However, because the additional dimension greatly increases the memory demands for the

deformation energy lookup table beyond what may be available on typical workstations, we did not pursue this option here.

To consider the effect of local membrane deformation on the solvation free energy of each atom, the local deformation at the position of each atom has to be obtained. To calculate the deformation of top and bottom leaflets at the position of any given atom i , $u_{u,i}$ and $u_{l,i}$, we first found the projection of the contact curve on a plane parallel to the xy plane that passes through this atom (shown with solid blue circle in Figure 4-1 B). The center of the circle was chosen to be the weighted average of atomic coordinates of all atoms whose z coordinates are within 15 Å from atom i . The weighting function given in Eq. 4.3.8 was determined empirically:

$$w_{ij} = \frac{1}{\sqrt{0.159 \times 2\pi}} \exp\left(-\frac{(z_i - z_j)^2}{(15 \text{ Å})^2 \times 0.159}\right) \quad 4.3.8$$

The x and y coordinates of the center of the hypothetical circle for atom i (CX_i and CY_i) were then obtained from Eq. 4.3.9

$$CX_i = \frac{\sum_j w_{ij} x_j}{\sum_j w_{ij}}; \quad CY_i = \frac{\sum_j w_{ij} y_j}{\sum_j w_{ij}} \quad 4.3.9$$

Next, the angle (κ_i) of atom i on this circle was defined as follows with respect to the x axis (cf. Figure 4-1 B)

$$\kappa_i = \tan^{-1}\left(\frac{y_i - CY_i}{x_i - CX_i}\right) \quad 4.3.10$$

In Eq. 4.3.10, $\kappa_i = \pi/2$ if $\begin{cases} x_i - CX_i = 0 \\ y_i - CY_i > 0 \end{cases}$ and $\kappa_i = 3\pi/2$ if $\begin{cases} x_i - CX_i = 0 \\ y_i - CY_i < 0 \end{cases}$

However, Eq. 4.3.10 is not defined for atoms where $x_i - CX_i = 0$ and $y_i - CY_i = 0$. In addition, we also observed that for atoms whose distance from their center is smaller than 1.5 Å, the derivatives of κ_i with respect to the atomic coordinates is not continuous. To address this issue, we considered r_i as the distance of each atom from its corresponding circle. Based on the magnitude of r_i , two possible states were considered. For $r_i \geq 1.5$ Å, the calculated angle κ_i was used in both leaflets spline interpolations $f(S, \theta)$ and $f(S', \theta)$ to obtain the corresponding leaflet deformation using the following equations

$$u_{u,i,1} = f(S, \theta)|_{\kappa_i} \quad 4.3.11$$

$$u_{l,i,1} = f(S', \theta)|_{\kappa_i} \quad 4.3.12$$

In the second case, the leaflet deformation was assumed be the average of the top and bottom MDPs for $r_i < 1.5$ Å

$$u_{u,i,2} = \frac{\sum_{j=1}^5 S_j}{5} \quad 4.3.13$$

$$u_{l,i,2} = \frac{\sum_{j=6}^{10} S_j}{5} \quad 4.3.14$$

This distance cutoff was chosen to ensure the continuity of the derivatives of $u_{u,i}$ and $u_{l,i}$ with respect to x_i and y_i . Finally, we defined the deformation in top and bottom leaflets as the weighted averages of the two possible states mentioned above. We chose a simple switching function as the weighting function (Eq. 4.3.17) and we determined its parameters empirically.

$$u_{u,i} = (1 - \lambda_i)u_{u,i,1} + \lambda_i u_{u,i,2} \quad 4.3.15$$

$$u_{l,i} = (1 - \lambda_i)u_{l,i,1} + \lambda_i u_{l,i,2} \quad 4.3.16$$

$$\lambda_i = \frac{1}{1 + \exp(8(r_i - 1.5))} \quad 4.3.17$$

4.3.2 Electrostatic solvation free energy (ΔG_{GB})

In the HDGB formalism, ΔG_{GB} is written in terms of the atomic charges, q_i , q_j and the effective local dielectric constant $\varepsilon'_{i,j}$ as follows

$$\begin{aligned} \Delta G_{GB} = & -166 \sum_{i,j} \left(1 - \frac{1}{\varepsilon'_{i,j}(\varepsilon'_i, \varepsilon'_j)} \right) \\ & \times \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i(\varepsilon'_i) \alpha_j(\varepsilon'_j) \exp\left(-\frac{r_{ij}^2}{8\alpha_i(\varepsilon'_i) \alpha_j(\varepsilon'_j)}\right)}} \end{aligned} \quad 4.3.18$$

$$\varepsilon'_{i,j} = \frac{1}{2}(\varepsilon'_i + \varepsilon'_j) \quad 4.3.19$$

where α_i is the atomic Born radius and the dielectric constants are only a function of the position of each atom along the membrane normal [73, 76].

Here, we are extending this formalism to include the effect of leaflets deformation at the position of any atom i via the $u_{u,i}$ and $u_{l,i}$ parameters.

$$\varepsilon'_i = \varepsilon'_i(z_i, u_{u,i}, u_{l,i}) \quad 4.3.20$$

This formalism and the following analysis consider only the deformation in one leaflet (i.e. $u_{u,i}=u_i$, $u_{l,i}=0$), but later in this chapter we will address how to consider deformations in both leaflets.

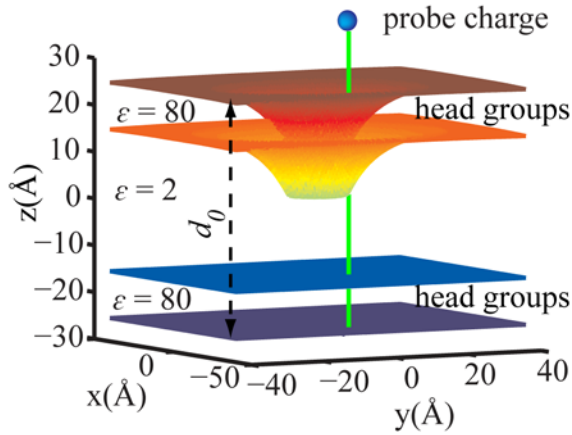


Figure 4-2: Schematic illustration of the dielectric profile calculation for a symmetric deformation of -15 Å in the top leaflet. The hydrophobic core is assumed to be 30 Å thick with the dielectric constant of 2. The z axis is chosen to be the membrane normal. The probe charge is shown as a blue sphere. It is translated along the green line in 1 Å increments. The unperturbed membrane width (d_0) is shown with a black dashed line.

In order to calculate the effective dielectric constant for an atom i as a function of both, its position along the membrane normal and the local leaflet deformation, we revised the original dielectric profile calculation [73]. First, we considered a simple two-dielectric membrane model

where the thickness of the hydrophobic core was chosen to be 30 Å with a dielectric constant of 2. The head group region and water dielectric constant was set to 80. The top leaflet was symmetrically deformed from $u=+10$ Å (expansion) to $u=-25$ Å (compression towards the center of the bilayer) around a cylinder with the radius of 7.5 Å. For each deformation value, a monovalent probe charge with a radius of 2 Å was displaced along the edge of the cylinder from $z=-25$ Å to $z=+25$ Å with 1 Å increments. For each position of the probe charge along the inclusion edge, the solvation free energy of the charge was calculated according to the PB equation using the APBS package [184]. 161 grid points were used in each direction with grid spacing of 0.5 Å. Figure 4-2 illustrates the setup of dielectric profile calculation for $u=-15$ Å.

Dielectric profiles $\epsilon'(z,u,0)$ were then obtained from the inverted Born equation using the solvation energies calculated from PB with $q^2=1$ and $a=2$ Å for the probe charge:

$$\epsilon'(z,u,0) = \frac{166q^2}{166q^2 + \Delta G_{PB}a} \quad 4.3.21$$

The dielectric profiles calculated from this method were subjected to further optimizations to improve the agreement between the calculated insertion energies of amino acid side chain analogs and the previously reported values by MacCallum *et al.*[77] For each amino acid side chain analog, the center of mass of the analog was first positioned at $z=30$ Å and the top leaflet was deformed around the hypothetical cylinder from $u=0$ Å to $u=-25$ Å with 2 Å increments. The minimum total energy was saved as the reference. The same procedure was repeated for $z=0$ Å and the insertion energy (from bulk solvent to the center of the membrane) was then calculated as the difference between the two minimum energies. The optimization was continued until a

good qualitative agreement between the calculated insertion energies and those obtained from explicit solvent simulation[77] was achieved.

As a result of this optimization, a general scheme for calculating dielectric constant for an arbitrary position along the inclusion edge was devised as follows:

$$\begin{aligned}\varepsilon'(z_i, u_i, 0) = & \varepsilon'(z_i, 0, 0) + w(u_i)P(u_i, z_i, a_\varepsilon, e_\varepsilon, c, I) \\ & + (1 - w(u_i))P(u_i = 1.5, z_i, a_\varepsilon, e_\varepsilon, c, I)\end{aligned}\quad 4.3.22$$

$$w(u_i) = \frac{1}{1 + \exp(2(u_i + 1.5))} \quad 4.3.23$$

$$P(u_i, z_i, a_\varepsilon, e_\varepsilon, c, I) = \frac{a_\varepsilon(z_i)}{1 + \exp\{-0.9145(-u_i - c(z_i)) + I\}} - e_\varepsilon(z_i) \quad 4.3.24$$

In Eq. 4.3.22, $\varepsilon'(z_i, 0, 0)$ is the original dielectric profile for a flat membrane[73, 76], a_ε , e_ε and c are adjustable cubic spline functions of z and $I=0$. The final functions for a_ε , c and e_ε are shown in Figure 4-3 A and the calculated dielectric profiles for some deformation values are shown in Figure 4-3 B as examples.

The dielectric profiles calculated for the distorted membrane show that, compared to the flat membrane [73, 76], the polarization effect extends to longer distances along the membrane normal from the surface of the membrane. In other words, the diffusion of high dielectric solvent to the low dielectric region of the membrane hydrophobic core mimics the water defect that was previously observed in explicit solvent simulations[77, 93] (see Figure 4-3).

4.3.3 Non-polar solvation free energy (ΔG_{np})

In the original HDGB model [73], ΔG_{np} was calculated from the atomic solvent accessible surface areas ($SASA_i$) multiplied by the cost of cavity formation (γ') at the position of atom i along the membrane normal.

$$\Delta G_{np} = \sum_i \gamma'(z_i) SASA_i \quad 4.3.25$$

In the new formulation of ΔG_{np} , γ' is assumed to be a function of the deformations in both top ($u_{u,i}$) and bottom ($u_{l,i}$) leaflets at the position of each atom i and the position of the atom along the membrane normal (z_i):

$$\gamma' = \gamma'(z_i, u_{u,i}, u_{l,i}) \quad 4.3.26$$

Following the same procedure for obtaining the dielectric constant profiles, here we only consider the deformation in one leaflet (i.e. $u_{u,i}=u_i$, $u_{l,i}=0$). However, later we will describe how simultaneous deformation in both leaflets are considered.

To include the effect of a leaflet deformation in γ' , first $a_{\gamma'}$ and then $e_{\gamma'}$ were obtained by multiplying a_ϵ and e_ϵ (Figure 4-3 A) with 0.015 to match the range of non-polar values from 0 to 1.2. The resulting scaled coefficients were then used to obtain the primary non-polar profiles using the following equation:

$$\begin{aligned}\gamma'(z_i, u_i, 0) = & \gamma'(z_i, 0, 0) + w(u_i)P(u_i, z_i, a_{\gamma'}, e_{\gamma'}, c, I) \\ & + (1 - w(u_i))P(u_i = 1.5, z_i, a_{\gamma'}, e_{\gamma'}, c, I)\end{aligned}\quad 4.3.27$$

where $\gamma'(z_i, 0, 0)$ is the original non-polar profile of the unperturbed membrane[76] and $w(u_i)$ is the same as described in Eq. 4.3.23. The scaled coefficients were then subjected to optimization to match insertion energies for amino acid analogs from explicit solvent simulations[77] as explained above for the dielectric constant profile. The best value for I was found to be 1.5 and c was chosen to be the same for both dielectric and non-polar profile. The optimized functions along with selected non-polar profiles at different deformations are shown in Figure 4-3 C and D.

In the derivation presented so far, one leaflet was kept flat while the other leaflet was allowed to deform. If both leaflets deform simultaneously, the deformation energy is assumed to be additive but the dielectric and surface tension modulation profiles are obtained as weighted averages of the individual values calculated separately for the top and bottom leaflets according to Eqs. 4.3.22-4.3.24 and Eq. 4.3.27. To determine the weighting function, 21 membrane geometries with different deformations in top and bottom leaflets were generated. The dielectric profiles were calculated from PB following the same procedure as described above. A weighting factor (η) based on a simple switching function was then fitted to match profiles obtained individually according to Eqs. 4.3.28 and 4.3.29 for the dielectric and surface tension calculations:

$$\varepsilon'_i(z_i, u_{u,i}, u_{l,i}) = \eta_i \varepsilon'_i(z_i, u_{u,i}, 0) + (1 - \eta_i) \varepsilon'_i(z_i, 0, u_{l,i}) \quad 4.3.28$$

$$\gamma'_i(z_i, u_{u,i}, u_{l,i}) = \eta_i \gamma'_i(z_i, u_{u,i}, 0) + (1 - \eta_i) \gamma'_i(z_i, 0, u_{l,i}) \quad 4.3.29$$

for an atom i that resides in a bilayer whose top leaflet is deformed by $u_{u,i}$ and bottom leaflet is deformed by $u_{l,i}$ at the position of atom i . As a result of the optimization, the weighting function was determined to be:

$$\eta_i = \frac{1}{1 + \exp\left(-5\left(z_i + \frac{u_{u,i} - u_{l,i}}{2}\right)\right)} \quad 4.3.30$$

4.4 Computational Methodology

The dynamic heterogeneous dielectric generalized Born model (DHDGB) was implemented as an extension of the HDGB model[73] in the macromolecular simulation package CHARMM, [159] version c36a1. Simulations of the WALP23 peptide were performed using the CHARMM22 force field[139] with the CMAP correction[24] while the gramicidin A channel was modeled using the CHARMM36 force-field[185] with the improved CMAP potential for D-amino acids [185]. No cutoffs were used for the non-bonded interactions. The MMTSB Tool Set[138] was used for the minimizations, equilibration steps, and for the analysis.

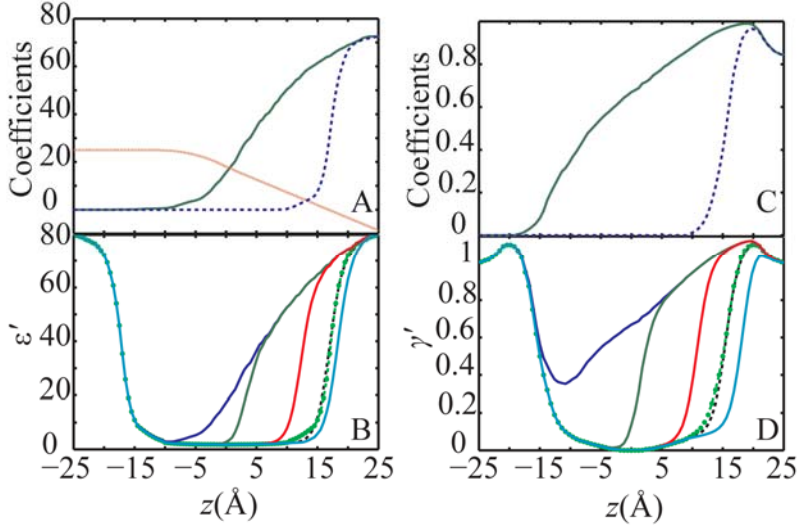


Figure 4-3: (A) Optimized coefficients in Eq. 24 for the dielectric profile calculations a_ϵ : solid green line, c : dotted red line and e_ϵ : dashed blue line; (B) dielectric profiles calculated using coefficients in panel A and Eqs. 22-24 for different symmetric deformations in one leaflet: $u=-25$ Å: blue, $u=-15$ Å: green, $u=-5$ Å: red, $u=+3$ Å: cyan, and $u=0$ Å dashed black line. The original dielectric profile for a flat membrane is shown with green circles[73, 76]; (C) optimized coefficients for surface tension modulation profile calculations $a_{\gamma'}$: solid green line and $e_{\gamma'}$: dash blue line (D) calculated non-polar profiles for different deformations in one leaflet with colors as in panel B.

4.4.1 Membrane Insertion of Small Molecules

The performance of the DHDGB model was first tested by comparing amino acid side chain analog insertion free energy profiles with previously calculated data from explicit solvent/lipid simulations [77]. For each amino acid analog, the center of mass of the corresponding molecule was translated from 30 Å to 0 Å along the membrane normal (z axis) using 1 Å increments. For each insertion depth, all MDPs corresponding to the top leaflet were varied from $u=0$ Å (flat leaflet) to $u=-25$ Å with 2.5 Å increments while the bottom leaflet was kept flat. In calculating deformation energies, we assumed that each side chain analog can be approximated as a cylindrical inclusion with a fixed radius of 4 Å. For each insertion depth of the side chain analogs, the top leaflet was deformed from flat state to -25 Å while the center of mass of the side chain analog was kept fixed. For each resulting membrane deformation, average energies were obtained by exhaustive sampling of different rotation angles of the molecule around its x and y axes at 18° increments and averaging over all possible states. Finally, the insertion energy was obtained as the difference of the minimum average energy at each insertion depth relative to $z=30$ Å where the molecule is outside the membrane and solvated only with water. The deformation energy for each membrane configuration was obtained using tabulated values for $r_0=4$ Å as described above.

4.4.2 Arg-Containing TM Helix

To compare the insertion free energy of charged arginine (Arg) side chains embedded in long TM helices, a poly-leucine helix with 91 residues was constructed and the middle residue was mutated to Arg, following previous work by Dorairaj *et al.*[93] The helix was oriented in such a way that its long axis is parallel to the z -axis and its center of mass is located at the center of the

bilayer. The helix was rotated around the z -axis so that the projection of the Arg center of mass onto the contact curve is located at $\theta=0$. Thus, the MDP closest to the Arg is S_I . The insertion depth of Arg was defined as the position of the $C\alpha$ of Arg side chain along the z -axis. To obtain the insertion profile, the TM helix was translated along z with the Arg location changing from 0 Å to 30 Å. For each Arg insertion depth, the helix and the MDPs were subjected to 50 steps of steepest descent (SD) minimization followed by a long adopted-basis Newton-Raphson (ABNR) minimization run. Minimization was carried out for at least 5000 steps or until the energy changed less than a tolerance of 10^{-7} kcal/mol. The final energy at each insertion depth was chosen as the minimum energy and the insertion energy was calculated relative to $z=30$ Å.

4.4.3 WALP23 Peptide

The helical WALP23 peptide (sequence: GWWLALALALALALALALWWA) was studied to test the effect of hydrophobic mismatch. The peptide was capped with an acetyl group on the C-terminus and an N-methyl amide group on the N-terminus. The initial structure was subjected to 50 steps of SD minimization followed by 1000 steps of ABNR minimization in an implicit membrane represented by the HDGB model[73] with the modified dielectric and non-polar profiles [76]. Other implicit solvent parameters are described elsewhere [74]. The temperature of the minimized structure was gradually increased to 300K during six rounds of equilibration. In each equilibration stage, the temperature of the peptide was increased by 50K and 10,000 MD steps were carried out. The time step was chosen as 1.5 fs to maintain stable simulations with the implicit solvent model [143]. The SHAKE algorithm was applied to constrain bonds involving hydrogen atoms [140]. The temperature was maintained by using a Nosé-Hoover thermostat[186, 187] with a temperature coupling constant (q_{ref}) of 50 K.cal.s². The equilibrated structure was

used as the initial structure for 32 independent MD simulations using the DHDGB model, each run over 20 ns length for a total simulation time of 640 ns. In each simulation, the temperature of the peptide was set to 303 K and the temperature of the MDPs was set to 0.5 K with a Nosé-Hoover coupling constant of $0.1 \text{ K}\cdot\text{cal}\cdot\text{s}^2$. The mass of the MDPs was set to 50,000 amu. The MDPs parameters were chosen in a way that the fluctuations in membrane deformation in a 20 ns simulation of MDPs mimicked the fluctuations observed in the reference explicit solvent simulation of a DPPC bilayer with the same length. To compare with the DHDGB model, 32 independent simulation using the HDGB[73, 76] model (with a fixed membrane) were also run for 20 ns each. The last 16 ns of each simulation were used for further analysis.

In order to identify the dominant structures of WALP23 in DPPC using DHDGB model, 5% of the total sampling was used to generate 2 ensembles: one with the structures with the insertion angle smaller than 15° and the second ensemble with the structures with insertion angle of 20° to 45° . The dominant structure of each ensemble was obtained using kclust method and the radius of $2 \text{ \AA}$ using heavy atom RMSD. The RMSD for clustering was obtained after aligning the structures in such a way that the insertion angle and insertion depth are preserved[78].

4.4.4 Gramicidin A

As another test case for hydrophobic mismatch, the gramicidin A (gA) dimer was simulated with the DHDGB model. The initial structure of the dimer was obtained from the Protein Data Bank (PDB code 1JNO[188]). The minimization and equilibration steps followed the same protocol as for the WALP23 peptide described above. The equilibrated structure was then used as the starting structure for 16 independent simulations using the DHDGB model, each over 12.3

ns simulation time. The first 3.5 ns of each simulation were discarded and the remaining 8.8 ns were used for analysis.

4.5 Results and Discussion

The DHDGB model was implemented in CHARMM and applied to a number of test cases as detailed in the following:

4.5.1 Amino acid side chain analogs

The first test case involves membrane insertion free energy profiles for amino acid side chain analogs. Figure 4-4 compares profiles from HDGB and DHDGB with explicit lipid simulation results [77]. As discussed previously [73, 76], HDGB is in good agreement with the explicit lipid results for most residues except for the charged amino acids (top row). DHDGB closely matches HDGB for non-charged amino acids but greatly improves the profiles for charged amino acids, avoiding the overestimation of the insertion free energies for Arg(+), Lys(+), Asp(-) and Glu(-) residues. This is a consequence of an effective decrease in the membrane thickness as the charged residues enter the membrane (see Figure 4-4, bottom panel). In case of the charged analogs, the top leaflet deforms by ~ -19 Å when the charged residues are at $z=0$. This deformation models the effect of water penetration and leads to a significant reduction of the insertion energy. For Arg(+), for instance, the application of DHDGB reduces the insertion free energy by ~ 15 kcal/mol. Minor membrane deformations are also found for polar residues at intermediate insertion depths, but the effect on the insertion free energy profiles is small with changes of less than 0.3 kcal/mol.

To calculate deformation energies, the side chain analogs were assumed to be encompassed with a cylinder with a fixed radius of 4 Å. The use of a cylindrical model instead of considering the

actual shape of the inserted small molecules is necessitated by practical constraints - our model relies on pre-calculated deformation energies that are possible only for limited sets of geometries – but it can also be justified on physical grounds. Because of the size and shape of the lipids, we expect that the length scale over which the membrane deforms, and over which deformation energies vary, in response to inserted solutes is larger than the scale of shape variations of small molecules. Hence, a simple geometric approximation like a cylinder appears to be a reasonable approximation. The close agreement between insertion energies calculated for different side chain analogs using DHDGB and the explicit solvent model provides partial justification for this assumption, but this point certainly deserves further attention in future efforts to improve our model.

4.5.2 TM Helix with a Central Arg(+)

The second test case is a long poly-leucine TM helix with a central Arg(+) that has been previously studied to understand the insertion of charged residues into membranes. The insertion free energy profile was calculated with both, HDGB and DHDGB, and was compared with potentials of mean force from long explicit water/lipid umbrella sampling simulations in a DPPC membrane [93]. The results are shown in Figure 4-5. As for the charged amino acid side chain analogs, the use of DHDGB matches the explicit lipid insertion profiles much better than the original HDGB model. The explicit lipid and DHDGB curves are in near perfect agreement from 9 Å to 30 Å while for deeper insertions the DHDGB model slightly underestimates the explicit lipid insertion free energies. When Arg(+) is at the center of the membrane, the insertion free energy is reduced from 35 kcal/mol with the HDGB model to 13.5 kcal/mol, compared to the explicit lipid free energy result of 17.8 kcal/mol [93].

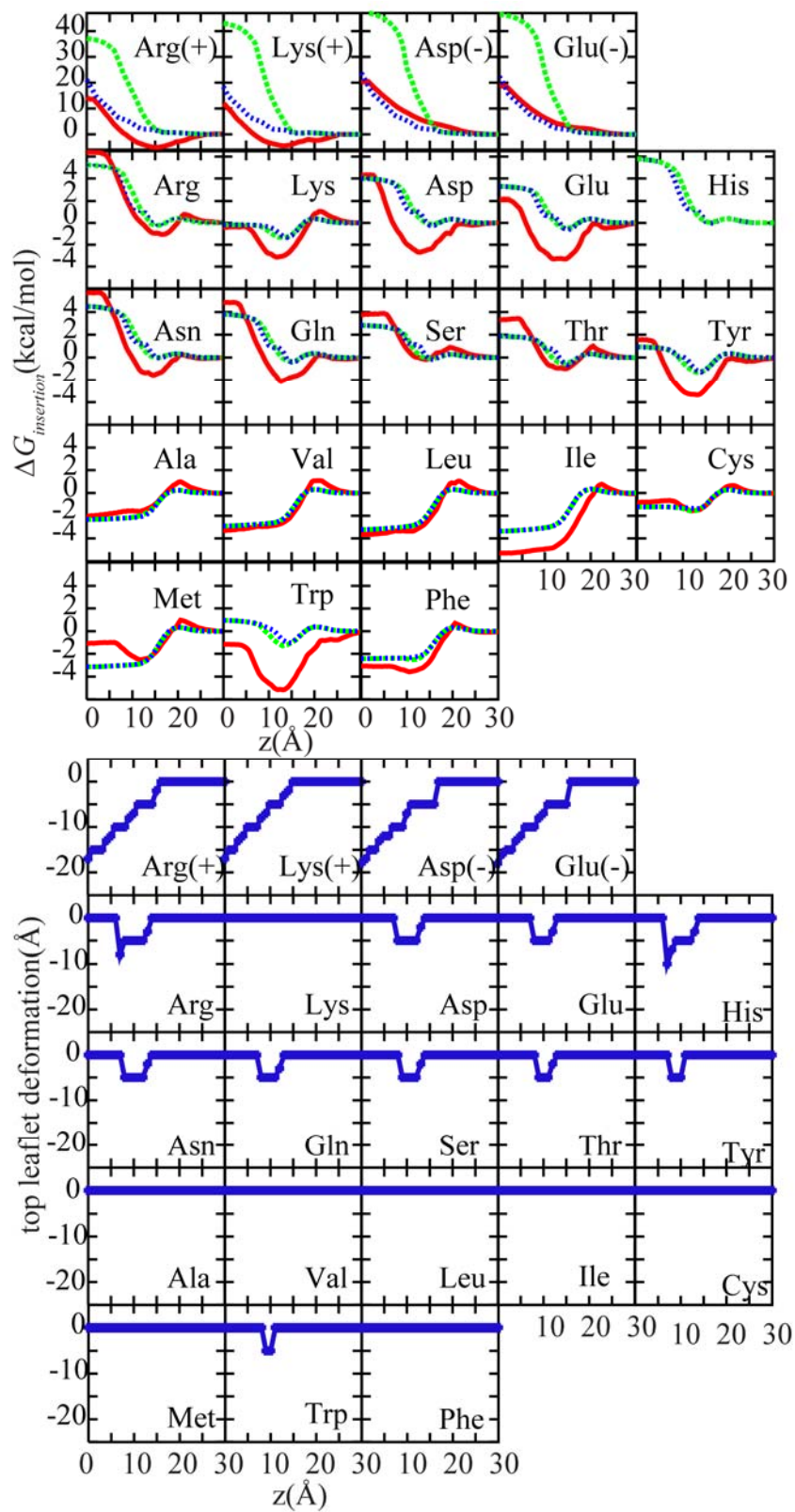


Figure 4-4: Top panel: Amino acid analogs insertion free energy profiles

Figure 4-4 (cont'd)

with HDGB[73, 76] (dashed green) and DHDGB (dotted blue) compared to results from explicit solvent/lipid simulations[77] (solid red). Bottom panel: Minimized leaflet thickness upon amino acid side chain analog insertion in the DHDGB model.

The discrepancy for deep insertions is likely due to idealizations of the deformation model that may become more problematic for very deep insertions [92]. However, differences in how the insertion profiles were obtained (minimization of an MMGB/SA function vs. PMF from umbrella sampling) or sampling inadequacies in the explicit lipid simulations that are notoriously difficult to converge [189], may also play a role.

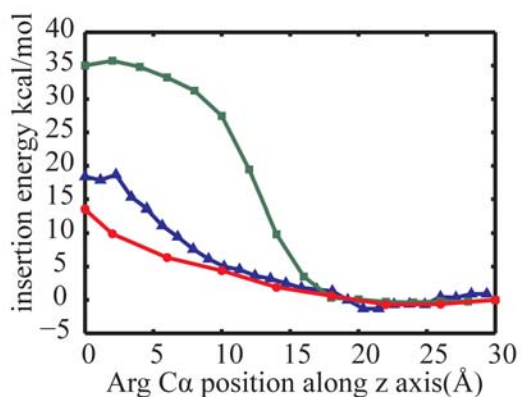


Figure 4-5: Membrane insertion energies for a poly-leucine TM helix with a central Arg(+). Energy calculated with the HDGB[73, 76] model (green squares) are compared with DHDGB energies (red circles), and explicit solvent results[93] (blue triangles).

In the free energy calculations for the same TM helix by Choe *et al.*[88] using the same elasticity model in combination with the PB equation the membrane was compressed systematically around the helix so that the contact curve followed the predetermined sinusoidal equation. As a result, the degree of deformation at a given Arg insertion depth was presumed without considering the coupling between peptide conformation and membrane deformation. In the DHDGB method, the MDPs were allowed to vary freely along with the Arg side chain orientation to find the optimal configuration. Therefore, no constraint was applied to the membrane configuration and no prior knowledge of a contact curve is required.

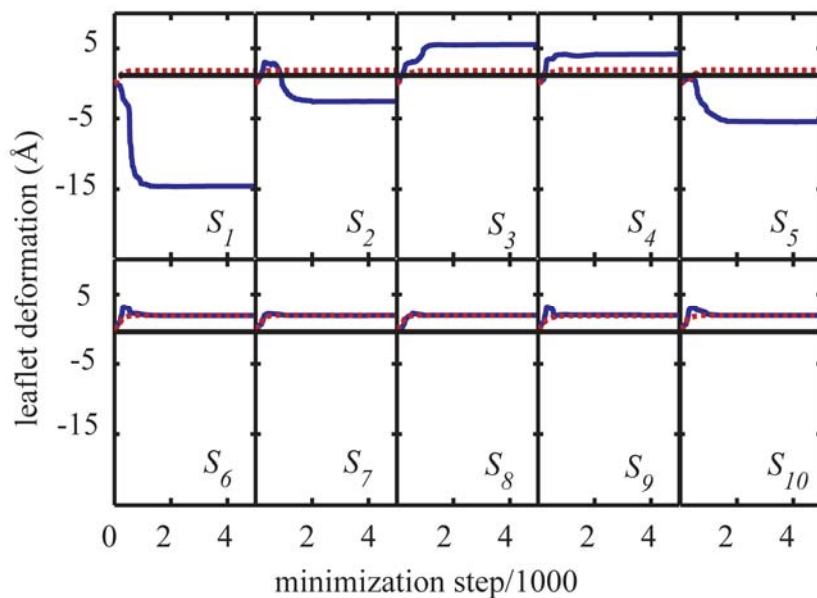


Figure 4-6: Variation of deformations of top $\{S_1, \dots, S_5\}$ and bottom $\{S_6, \dots, S_{10}\}$ leaflets. The dashed red line represents the variation of MDPs for $z(\text{Arg:C}\alpha)=30\text{\AA}$ and the solid blue line shows the variation for $z(\text{Arg:C}\alpha)=6\text{\AA}$. The thin solid black line represents the flat leaflets.

To follow the membrane deformation during the course of minimization, the MDPs were monitored and plotted against the minimization step in Figure 4-6 for two different insertion depths of Arg(+). When Arg(+) is outside of the membrane at $z=30\text{\AA}$ (dashed red line), there is a slight increase in the membrane thickness in both top($\{S_1, \dots, S_5\}$) and bottom ($\{S_6, \dots, S_{10}\}$) leaflets, presumably due to membrane adaptation to the hydrophobic poly-leucine helix that extends out of the membrane. However when the charge is inserted into the hydrophobic core, i.e. $z=6\text{\AA}$, the top leaflet in the vicinity of the helix (S_1) is significantly perturbed with a deformation of as much as $-15\text{\AA}$.

The insertion energy calculated here is $\sim 5.3\text{ kcal/mol}$, smaller than the value obtained by Choe *et al.*[88] This difference probably results from using two different approaches for obtaining solvation free energies (GB vs. PB which was employed in Choe *et al.*[88] approach).

4.5.3 MD simulation of WALP23 in a membrane bilayer

The orientation of TM helices in membrane proteins is closely related to their structure and function [190]. In mechano-sensitive channels, for example, tilting and reorientation of the TM helices can lead to channel opening and closing [191, 192]. It is generally believed that the tilt angle of a TM helix is affected by the hydrophobic mismatch between the helix and the membrane bilayer. When the peptide hydrophobic length is greater than the membrane hydrophobic thickness, tilting provides better peptide-membrane interaction and less exposure of hydrophobic residues to the solvent. However, for short peptides whose lengths are comparable to the hydrophobic thickness of the membrane, the tilt angle is reported to be smaller [193].

WALP peptides are synthetic short peptides that can model the hydrophobic mismatch in different membrane environments very well and, therefore, they have been the target of both experimental and theoretical studies [194-206].

As a further test, the effect of membrane deformations on the tilt angle of WALP23 was examined using MD simulations with both the DHDGB and the HDGB models. The tilt angle of the peptide was defined as the angle of the peptide axis with the membrane normal (z axis) and the membrane deformation was calculated as sum of the average deformations of top and bottom leaflets. The probability distribution of the tilt angle shows two states in both deformable and non-deformable models (see Figure 4-7 A). For the static membrane with a fixed thickness modeled by HDGB, the most probable insertion angles are located at 10.8° and 16.3° (Figure 4-7 A, blue curve) with the smaller angle being slightly more favorable. However, the inclusion of the membrane deformation with the DHDGB model clearly affects the tilt angle distribution (Figure 4-7 A the red curve). The first peak of the distribution occurs at 9.2°, similar to the first peak with HDGB, but the second peak is shifted to much larger tilt angles of 26.4°. In Figure 4-7 B the population distribution of membrane variation from DHDGB model is depicted. Closer look at the membrane variation reveals that conformations in the first peak do not coincide with significant variations in membrane thickness (-0.8 Å), while more significantly tilted conformations are concomitant with more significant membrane deformations (-2.4 Å). The dominant structures with different insertion angles are also depicted in Figure 4-7 C with green (insertion angle of 8°) and cyan (insertion angle of 26°). Aligning of these two dominant structures (Fig. 7C right pannel) shows that the orientation of the Trp side chain (shown in the black box) varies in the structures with different insertion angles. This observation suggests that the orientation in the green structure is favored in the perpendicular insertion while the

orientation in the cyan structure favors the tilted orientation which results in the membrane deformation.

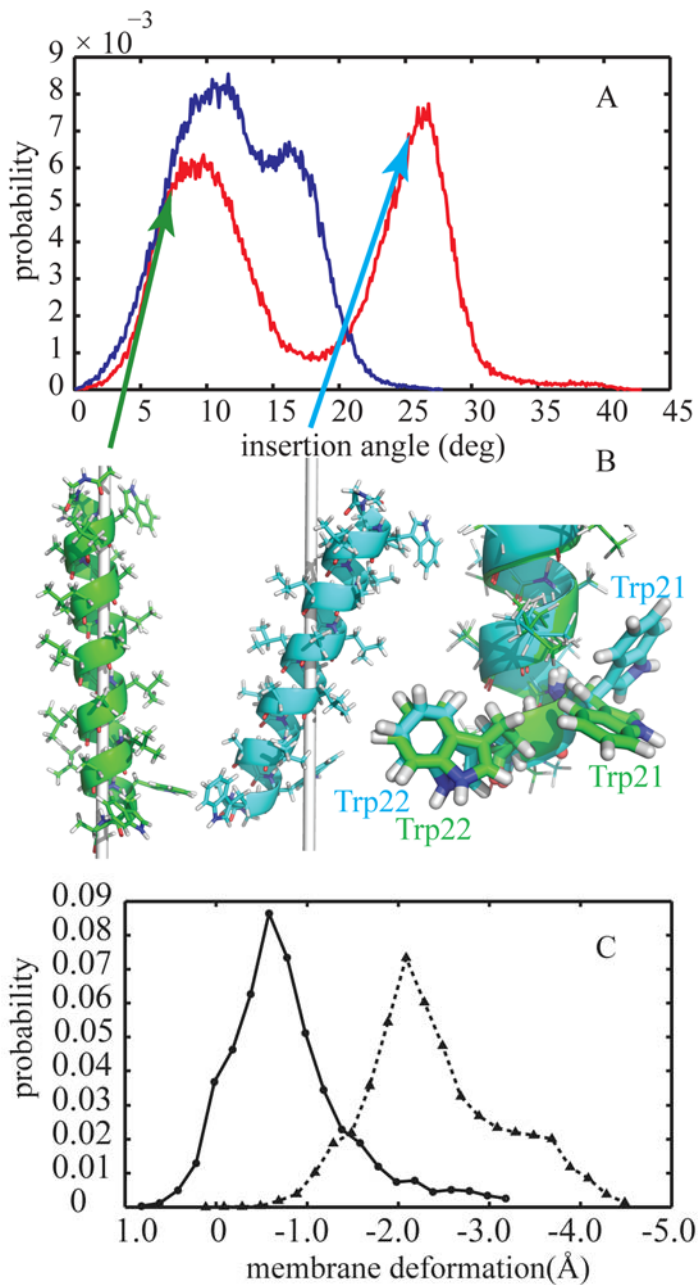


Figure 4-7: (A) Probability distribution of the membrane insertion angle of WALP23 in implicit membrane environment using HDGB (blue line) and DHDGB (red line) (B) Probability distribution of membrane deformation for insertion angles smaller than 15° (solid line with circle

Figure 4-7(cont'd)

marks) and insertion angles between 20° and 45° (dotted line with triangle marks) (C) The dominant structures of WALP23 with different insertion angles: green, insertion angle of 8° and cyan, insertion angle of 26°. The gray line shows the z axis (membrane normal). The right panel shows the aligned structures Trp with different side chain orientations are shown in a box. The structures are rendered in pymol[148].

The overall average insertion angle obtained for WALP23 using the DHDGB model is $17.6^\circ \pm 1.2^\circ$. This angle is larger than the value of 14.9° observed in explicit solvent simulations of the peptide in a palmitoyloleoylphosphatidylcholine (POPC) bilayer with a comparable thickness to DPPC, the membrane modeled here with DHDGB [206]. The larger value with DHDGB is a result of sampling both smaller and much larger insertion angles. Interestingly, much larger insertion angles have also been observed in MD simulation of WALP23 in DMPC where angles between 28° and 33° were reported [203, 206]. The sampling of small angles for an undeformed membrane while larger angles are preferred for a deformed membrane with DHDGB is therefore in good qualitative agreement with the explicit solvent simulation results. Interestingly, experimental studies also report values that appear to cluster either around 15° or 30° depending on the membrane and experimental technique that has been used [207-209]. The comparison with the experimental data is hindered, though, by apparent difficulties in unambiguously deriving insertion angles from the experimental measurements [203].

The suggestion of a two-state equilibrium between undeformed membrane, small peptide insertion angle and deformed membrane, large insertion angle states by our simulation results is

not supported by other simulations. However, this may be due to significant kinetic barriers due to the presence of explicit lipids that prevent a transition from an undeformed membrane to a deformed membrane in the limited simulation time scales. Alternatively, it is possible that the elastic membrane model used here slightly underestimates the cost for deforming the membrane, hence allowing such a state to be sampled with a significant population. Finally, it should be noted that the elasticity model used in this work is parameterized for cylindrical inclusions whose main axis is parallel to the membrane normal. Such a model is assumed to work well for small insertion angles but may become problematic as insertion angles become larger.

Further tests should be carried out in the future to examine this system more closely and possibly fine-tune the DHDGB model if necessary.

4.5.4 MD simulation of gramicidin A in a membrane bilayer

Finally, we applied the DHDGB model to gramicidin A (gA), a short dimeric peptide with 15 amino acids per monomer and an alternating L-D sequence. GA forms a small channel that is shorter than most membrane thicknesses and it is, therefore, a good model for studying protein-induced membrane deformations [82, 86, 87, 180-183, 210]. In recent MD simulations of gA in different membrane bilayers with various thicknesses, it has been observed that the tilt angle of the dimer is a function of the thickness, apparently to maximize protein-membrane interactions; in particular, the peptide adopts larger insertion angles in membranes with thinner hydrophobic core. Furthermore, it was also reported that the peptide can induce an overall 4 to 5 Å compression of the membrane in different bilayers [211]. A comparison between WALP23 peptides and gA channels suggests that gA responds to hydrophobic mismatch by inducing membrane deformations rather than tilting of the peptide, while the WALP23 peptide

predominantly uses tilting to adapt to membrane environments with different thickness [206, 211-213].

From simulations of the gA channel with the DHDGB model we obtained the tilt angle and membrane deformation shown in Figure 4-8. The insertion angle and membrane deformation were calculated using the same definitions as were applied for WALP23.

We found an average insertion angle of $7.4 \pm 0.2^\circ$ which is in good agreement with the insertion angles that were previously obtained for gA in DOPC ($\sim 9.1^\circ$) and POPC[211] ($\sim 8.9^\circ$). Furthermore, we observe a significant membrane deformation of $-13.9 \text{ \AA} \pm 0.1 \text{ \AA}$. While these values are larger than the 4 to 5 $\text{\AA}$ that have been reported from explicit solvent simulations[211] they support at least qualitatively the hypothesis that gA tilts less than WALP23 but deforms the membrane more significantly.

The discrepancy between the membrane deformation observed in our simulation and the values in explicit solvent[211] may be explained in part by different references for measuring deformation. In the analysis of the explicit solvent simulations, the membrane thickness was measured as the average distance between the C2 atoms of acyl chains in both leaflets which amounts to a unperturbed thickness of $\sim 29 \text{ \AA}$ [211]. However our definition of membrane thickness refers to the top of the head-group[73] region which is $\sim 10 \text{ \AA}$ away from the hydrophobic core [73]. Considering the difference between the thickness definitions and assuming that the membrane deformation would be distributed along the entire membrane the deformation observed in DHDGB may be scaled by 29/50 to estimate a value of $\sim 8.2 \text{ \AA}$ for the deformation of just the hydrophobic layer. This value compares more favorably with the 5 $\text{\AA}$ deformation observed in the explicit simulations [211]. An additional factor may be that

membrane deformation is kinetically inhibited in explicit lipid simulations and may not have been fully realized due to limited simulation lengths. In the DHDGB model, the response of the environment is much more rapid and exhibits no kinetic barriers. Hence, the optimal deformation, within the limits of the model used here, is readily achieved within relatively short simulation times.

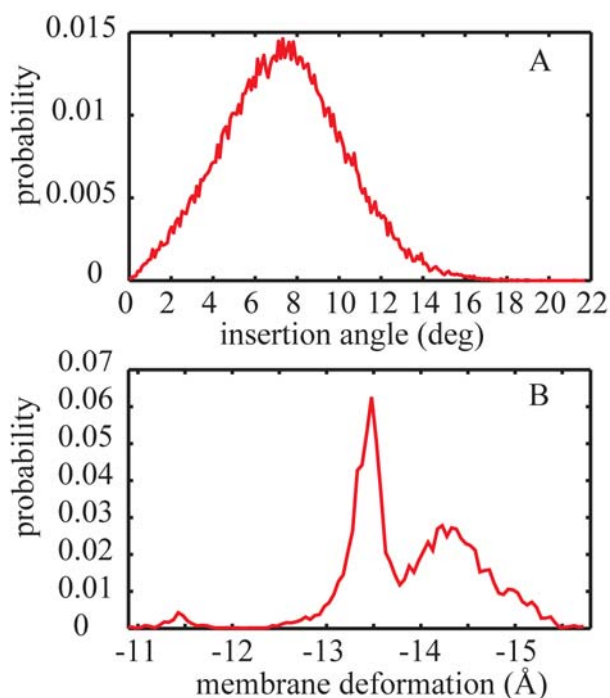


Figure 4-8: (A) Probability distribution of the membrane insertion angle for the gA dimer in the membrane bilayer. (B) Probability distribution of the average membrane deformation.

4.6 Conclusions

We describe the DHDGB implicit membrane model that incorporates change of the local membrane thickness in response to macromolecules embedded into the membrane. DHDGB couples the previously developed HDGB implicit membrane model to elasticity theory in a

framework that is suitable for molecular dynamics simulations. DHDGB correctly predicts membrane deformations in response to the insertion of charged residues and as a result of hydrophobic mismatch. DHDGB significantly improves the agreement for the insertion energy of amino acid side chain analogs between the implicit membrane model and results from all-atom explicit solvent simulations [77]. The same is found for the insertion of Arg(+) as part of a long TM helix where the inclusion of membrane deformation decreases the insertion free energy by ~ 21.5 kcal/mol, close to the value obtained by explicit lipid simulations.

We have also tested our model with TM peptides such as WALP23 and gA dimer using MD simulation. In the case of WALP23, we observed significant tilting of the peptide in the response to the hydrophobic mismatch with tilt angles that are in close agreement with the corresponding values from explicit solvent simulations and experiments. For gA, tilting is less pronounced and instead significant membrane deformations are found. Therefore, DHDGB captures the well-known tendency of WALP23 to tilt in favor of significant membrane deformations vs. gA, which favors significant membrane deformations over tilting.

While the DHDGB method represents significant progress over fixed membrane implicit membrane models, there are still a number of limitations in the current implementation. Our method of calculating the dielectric and surface tension modulation is based on a cylindrical approximation of a solute inserted into the membrane. This model fails for peptides that have very large tilt angles, systems that interact only interfacially with the membrane, or peptides that induce pore formation. In principle, such cases can be addressed by improving the calculation of the profiles and possibly along the MDPs to vary in x and y as well as in the z direction. Drastic changes of the membrane, such as pore formation, would also require a better estimate of membrane deformation energies than what is provided by the mattress model used here.

Despite its remaining limitations, the DHDGB is a significant advance over the HDGB model and leads the way to wider applications of implicit membrane models in the study of membrane-bound systems.

4.7 Acknowledgements

The authors thank Michael Grabe and Keith Callenberg for providing a numerical solver for the partial differential equation for the elastic membrane model and for insightful comments on elasticity theory. We also greatly appreciate comments by Toby Allen and Wonpil Im with respect to membrane parameterization and gA modeling. Funding from NIH GM084953, NIH GM092949 and NSF CBET 0941055 is acknowledged. Computer resources were used at XSEDE facilities (TG-MCB090003) and at the High-Performance Computing Center at Michigan State University.

5 CONCLUSIONS AND PRESPECTIVES

5.1 Study of the membrane fusion process and the role of fusion peptides

A large body of the work discussed in this dissertation is dedicated to the study of fusion peptides in the context of membrane bilayers and their conformational sampling under different conditions such as protonation state of the titratable residues, termini and flanking residues. The ultimate purpose of investigating the structure of the fusion peptides is to obtain deeper understanding of the fusion process, which is a critical step in initiating viral infection. The structure of many fusion proteins are composed of two sub-domains: the globular domain that attaches to the cell receptors and the helical bundle, which is a homotrimer that hides the fusion peptides until the proper conditions for their activity are provided [101, 103, 105]. For the influenza fusion protein, hemagglutinin (HA), the decrease in the pH of the endosome triggers the fusion process and causes the helical bundle to undergo a large structural reorientation that exposes the fusion peptides to the bundle exterior and allows them to interact with the host cell membrane[103]. Unfortunately, the structure of the fusion peptide in the fusogenic state of the protein is not known because the peptide had to be cleaved to increase the fusion protein solubility[103].

In chapter 2, the conformational sampling of the influenza fusion peptide (IFP) in a membrane bilayer has been investigated. Application of the temperature replica exchange (TRMD)[99, 100] as an enhanced sampling technique along with the implicit membrane model enabled us to achieve sub-microsecond timescale with minimal computational cost. We have observed that in our simulations the peptide tends to form a tight break in the middle region. This structure was later confirmed by NMR spectroscopy[124]. We also observed that the effect of the pH on the secondary structure of the peptide is minimal which was found to be in good agreement with ^{13}C chemical shifts measured in solid state NMR spectroscopy studies[123]. The

parallel configuration of IFP relative to the membrane bilayer observed in our simulation also matched the orientation and the interfacial insertion that was obtained from peptide-lipids NOE measurements in the liquid state NMR spectroscopy of IFP in micelles[124]. Similar insertion angle was later observed in MD simulations of IFP in explicit POPC bilayers[214].

The same technique was applied to investigate the conformational sampling of the Ebola fusion peptide in the membrane bilayer and water[79]. In chapter 3 the details of the simulations and the results are presented and discussed. We observed that in the membrane bilayer, the peptide adopts a short helical structure in the middle region while the N and C termini are mostly extended. The structure remained stable during the course of the simulation while it adopted a more extended configuration in water environment. The same behavior was previously observed in the NMR structure of EFP in SDS micelles and water[130].

Although recent experimental and theoretical studies have broadened our knowledge about the structure and dynamics of the fusion peptides and their role in membrane fusion [96, 98, 115, 119, 123, 124, 130, 214], several aspects of the fusion process have remained unexplored.

Molecular dynamic simulation of the Ebola fusion peptide(EFP) [79] and NMR structures of the influenza fusion peptide (IFP)[124] clearly show that the residues in the vicinity of the fusion peptide can affect its secondary structure. Therefore, the structural findings about the isolated fusion peptide without including the nearby residues, although valuable, may not provide a complete model to study the overall fusion mechanism. To address this issue it might be necessary to include the neighboring residues in the fusion peptides constructs.

Another possible phenomenon that has been considered important in fusion process is the oligomerization of the peptide. The conformational sampling of the dimeric form of the IFP was

studied using the HDGB model. The simulation details along with the results are discussed in the appendix. The dominant structure of the inserted dimeric form is helix-break-helix with a tight bend in the middle which is structurally close to the recent NMR findings of the monomeric form[124]. However, the non-inserted dimers adopt more extended structure in the C terminal region. Although the inserted structures obtained from these simulations are in agreement with the monomeric form[124], the diversity of the sampled configurations in the simulations that have different initial structures, suggests that the simulations are far from convergence and the results are not conclusive. Perhaps using umbrella sampling molecular dynamic simulations which start from a non-inserted dimer and gradually dragging the dimer into the membrane bilayer can lead to a more converged map and avoid trapping in local minima.

5.2 Study of membrane deformation

In the current implicit membrane models[40, 42, 62, 73] pioneered by Parsegian[38], the bilayer is described as a static slab of low dielectric surrounded by high dielectric solvent. In these simplified models, the thickness fluctuations and local and global bending and curvature variations of the membrane are ignored. However, membrane fluctuations have been observed in many biologically relevant phenomena such as cellular motility[215] as well as explicit solvent MD simulations of membrane bilayers using all atom or coarse grain models [178, 216-218]. It has also been noticed that the membrane undergoes substantial deformation as a response to the insertion of charges [77, 93, 178, 179, 219-222]. For instance, when a single charged particle such as a cation or an anion is inserted into the bilayer, several water molecules and lipid head groups are allowed to penetrate the bilayer along with the ion to keep it partially solvated in the membrane hydrophobic core [77, 93, 178, 179, 219-222]. However, these favorable interactions are obtained at the expense of imposing strains on the membrane. In the current implicit

membranes the total loss of the hydration shell of the ion, which is the result of ignoring membrane perturbations mentioned above, leads to overestimation of the insertion energy[93].

In order to allow membrane perturbations around proteins and couple the effect of variations in membrane geometry and proteins structure, it is critical to estimate the membrane fluctuation energy and the changes in the local environment of the protein due to these perturbations[88].

In chapter 4 of this dissertation, a new technique was introduced to estimate the membrane perturbation energy and couple its effect on the solvation free energy via variation in local apparent dielectric constants and surface tension modulations. In this model the heterogeneous dielectric generalized Born (HDGB) theory, which estimates the solvation free energy [73, 76], was merged with the mattress model that calculates the membrane deformation term[82, 86, 87]. This is the first application of the mattress model in the GB theory. The modified membrane model, called dynamic heterogeneous generalized Born (DHDGB), improves the agreement between the calculated insertion free energy of amino acid side chain analogs using the implicit membrane model and explicit lipid/solvent. The performance is especially enhanced for charged analogs such as Arg(+), Lys(+), Glu(-) and Asp(-). This is due to the formation of water defect that shields the charge and prevents its exposure to the hydrophobic core of the membrane. The new model can also predict hydrophobic mismatch very well. For WALP23 for instance, the membrane deformation and the insertion angle of the peptide are coupled i.e. the peptide adopts more tilted conformation when the bilayer thickness is decreased. The DHDGB model, outperforms the current implicit membrane model in terms of charge insertion and hydrophobic mismatch but it cannot be employed to study complicated systems such as ion channels where the geometry of the channel requires water penetration throughout the whole membrane spanning region[223]. In order to include the effect of water spanning region perhaps the presence of a

high dielectric region of water that spans through the membrane has to be considered in the dielectric and surface tension modulation profiles. In other words the current profiles have to be generalized for the cases where water spans through the whole range of the membrane interior.

Another important feature that is missing in the DHDGB model is the dependency of the membrane deformation energy and membrane profiles on the variation of the inclusion radius. For relatively dynamic systems such as fusion peptides, the radius of the part of the peptide that is in contact with the membrane is constantly changing. Therefore, although considering a fix inclusion radius for trans-membrane proteins might not be far from reality it may not work for flexible systems such as fusion peptides. At this point addition of another dimension to the tabulated deformation energies and interpolation is not computationally favorable. But more efficient interpolation algorithm can provide the opportunity of expanding the tabulated values.

Another approach that has been considered in our group to improve the current implicit membrane model [73, 76] is to modify the current non-polar energy contribution. In this approach the non-polar energy is divided into two parts: the cost of cavity formation in the membrane bilayer and water, and the proteins interactions with the membrane and water as the new additional feature. The results show that including a more realistic van der Waals interactions improve the amino acid insertion profiles and the helix-helix association in the membrane bilayer. However, this method is developed for a flat static membrane without considering the membrane fluctuation[224]. Ultimately, merging the two current methods, i.e. the improved non-polar calculation and membrane deformation, would lead to a model that can very closely represent the explicit solvent and lipid medium.

The DHDGB model can be applied to study single pass trans-membrane peptides and proteins. The close agreement between the DHDGB model and explicit solvent data suggests that this model can provide a computationally affordable alternative for explicit solvent and lipid simulations while offering a more realistic representation of the membrane environment.

6 SUPPLEMENTAL CHAPTER: CONFORMATIONAL SAMPLING OF THE DIMERIC FORM OF THE INFLUENZA FUSION PEPTIDE IN THE CONTEXT OF THE MEMBRANE BILAYER

6.1 Introduction

Conformational sampling of the influenza fusion peptide (IFP) has been extensively studied with several experimental [115, 123, 124] and theoretical methods [78, 98, 136, 214]. These studies always have targeted the monomeric form of the fusion peptide. However, the trimeric structure of the helical stem of the fusion proteins and some fluorescence quenching experiments on the IFP support the formation of peptide complexes in membrane bilayers [101, 103, 105, 225]. Circular dichroism, and infrared spectroscopy studies of the IFPs in lipid bilayers also suggest that fusion peptides associate with one another at the surface of the membrane [226]. Fourier transform infrared spectroscopy also reveals that the secondary structure of the inserted peptide is severely affected by the association process. In low pHs, IFPs tend to stay as monomers in the membrane-water interface while increasing the ionic strength of the solution shifts the population towards oligomers with mostly anti-parallel β sheet structures [226]. While several experiments have been conducted on the secondary structure of the monomeric form of the fusion peptides [115, 123, 124, 130], little is known about its oligomeric structure. Theoretical studies on the trimers of the IFP suggest that the oligomerization does not alter the secondary structure of the peptide. However, it can change its orientation in membrane bilayers [136]. But perhaps the size of the system and its flexibility require longer simulation time to draw any conclusion about the globally stable conformation of the trimer in membrane bilayers[136]. In addition, the simple representation of the membrane bilayer used in this study can be a source of error as well[136].

In order to study the effect of oligomerization on the structure and dynamics of the IFP, we conducted temperature replica exchange [99, 100] simulations of dimeric form of the peptide using HDGB implicit membrane model [73, 76]. In the following, the computational methodology is presented before the results are discussed.

6.2 Computational methodology

CHARMM program [159] version 36a1 along with the MMTSB tool set[138] were used to perform TRMD simulation of the dimeric forms of the IFP. In all of the simulations, the CHARMM22[139] force field including the CMAP correction[24] was used to describe the peptide interactions.

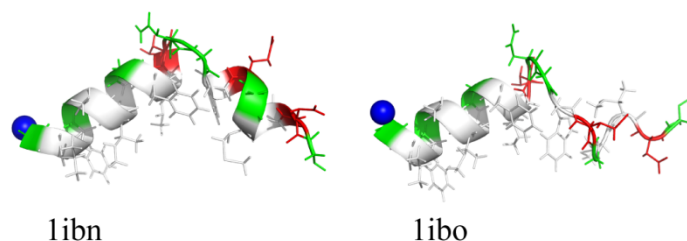


Figure 6-1: initial monomeric structures of IFP at pH=5.0(1bn) and pH=7.4(1bo)[115]. The acidic residues (Glu 11, Glu 15 and Asp 19) are shown in red while the hydrophobic residues and Gly are shown in gray and green respectively. The N terminus is marked with a blue sphere. The picture is rendered using pymol[148].

The dimers were constructed from the monomeric structures of the IFPs at different pHs [115]. For each simulation, the initial monomer structure was downloaded from Protein Data Bank (PDB code: 1bn at pH=5.0 and 1bo at pH=7.4[115]). The initial structures of the monomers are shown in Figure 6-1. Both N termini were positively charged and C termini were amidated (see ref. [78]). The second monomer which was the exact replica of the first chain was positioned in the xy plane of the first monomer with 30Å distance from its center of mass and the center of mass of the system was positioned at $z=17\text{\AA}$ relative to the center of the bilayer. Each monomer of the dimer was rotated around its z axis with 45° increment which resulted in an

ensemble of 16 structures. All structures were minimized using 50 steps of the steepest descent and 500 steps of the adapted-basis Newton-Raphson methods. The minimized ensemble was used as the initial structures for temperature replica exchange (TRMD) [99, 100] using 16 replicas spaced between 300 and 500K. The time step of the simulation was chosen to be 1.5 fs[143] and the exchange was attempted every 200 steps. The other simulation parameters can be found in ref [78]. However, the periodic membrane was not applied in this study to avoid unrealistic interactions of the peptide chains with one another. The two monomers were considered dimers if the distance between at least two C α s of the two chains was less than or equal to 7Å. Each simulation was 22ns long from which the first 9ns was discarded for equilibration. For each simulation an ensemble of dimeric fusion peptides at 300K was extracted from the overall sampling and used for analysis. For each dimer ensemble, the inter-chain angle was calculated as the angle between the first 10 residues of each chain and the center of mass was obtained as the projection of the center of mass of the dimer on the z axis. The two-dimensional potential of mean force (PMF) was obtained at 300 K along the center of mass of the dimer vs. inter-chain angle. The most populated structures of each ensemble were obtained using a modified clustering method (the details can be found in ref [78]).

6.3 Results and discussion

TRMD was performed on systems composed of two monomers of the IFP with different orientation with respect to each other and with starting structures obtained from NMR study of monomeric form of the IFP at different pHs in DPC micelles [115].

In Figure 6-2 (top panel) the 2D PMF for the dimer ensemble obtained from TRMD of monomers of 1ibn is plotted.

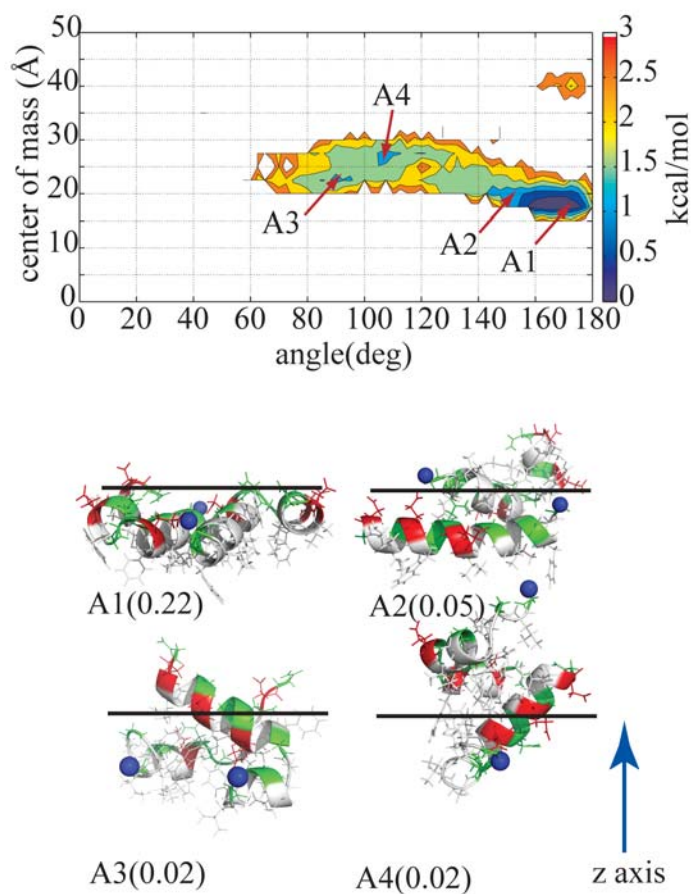


Figure 6-2: Top panel: 2D PMF of dimers obtained from TRMD of 1bn monomers. Bottom panel: the most visited structures in the ensemble of dimers of 1bn. Relative population of each structure is written in parentheses. The color coding is the same as Figure 6-1 and the membrane position relative to the dimer for the inserted structures is shown with a black line while the membrane normal (z axis) is shown with a blue arrow. The structures are rendered using pymol[148].

The PMF clearly reveals that the dimers formed from 1ibn monomers tend to insert in to the membrane and the most dominant structure (A1) does not show large structural deformations due to the dimerization (compare with 1ibn in Figure 6-1). The tight break that was previously observed in the monomeric form[78] and confirmed by NMR data[124]. The final orientation of the dominant structure of the dimers in the membrane bilayer is also the same as monomer, which was earlier observed to adopt a parallel interfacial orientation [78]. In addition to the parallel structure, some oblique orientations are also observed with more severe structural deformation with respect to the monomer but their populations are comparatively small(A2-A4). A2 shows a long helical structure that lays flat on the membrane parallel to the membrane surface while the other chain is partly inserted and its secondary structure is almost intact. In A3 and A4 the N terminal helical structure is severely disrupted in both chains while the C terminal helix is occasionally elongated.

The TRMD of dimers formed form 1ibo monomers show different structures and orientations (see Figure 6-3). The minimum of the PMF shifts towards higher center of mass values which implies that the insertion affinity of the dimer is significantly reduced. This might be due to the formation of stable dimer structures in which the two monomers conceal each other's hydrophobic side chains from water by forming a hydrophobic network (Figure 6-3 B1-B3). This stable conformation allows the peptides to expose their charged amino acids (Figure 6-3 bottom panel) while keeping the hydrophobic side chains covered. The inserted structures are mostly helical in N terminal region while the C terminal is extended with no specific secondary structure (Figure 6-3 B4).

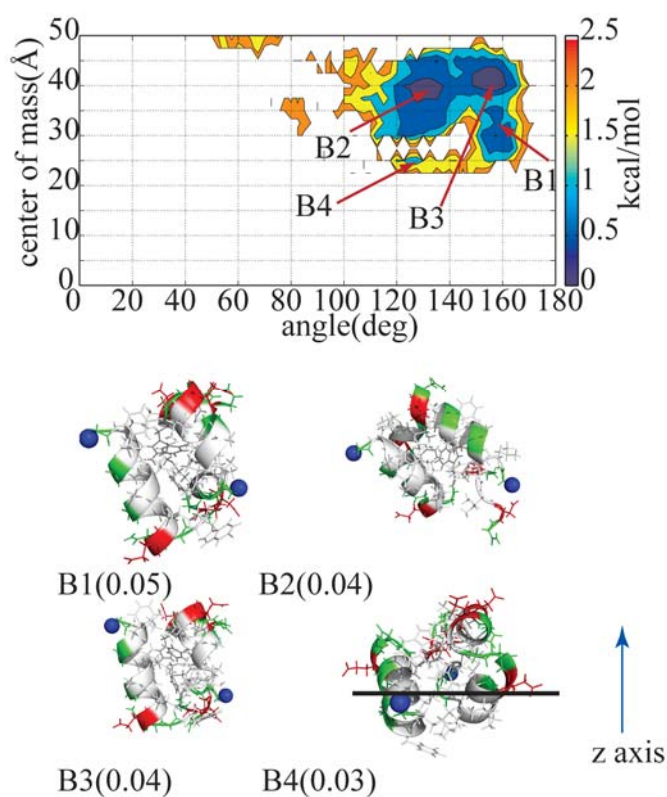


Figure 6-3: Top panel: 2D PMF of dimers obtained from TRMD of 1ibo monomers. Bottom panel: the most visited structures in the ensemble of dimers of 1ibo. Relative population of each structure is written in parantheses. The color coding is the same as Figure 6-1 and the membrane position relative to the dimer for the inserted structures is shown with a black line while the membrane normal (z axis) is shown with a blue arrow. The structures are rendered using pymol[148].

6.4 Conclusions

Our preliminary results suggest that insertion is more favorable for the dimeric form of 1ibn compared to 1ibo. However, none of the structures sampled in one simulation is visited by the other one and the large difference between the two PMFs implies that the simulations are trapped in their local minima and far from convergence.

One alternative approach to tackle this problem could be to use periodic boundary condition where the interaction of the peptides with their images and the image of the membrane may prevent the diffusion of the dimers in to water and formation of the stable hydrophobic network. The periodic boundary is recently implemented in HDGB by our group and can be applied to study the oligomerization problem. Another alternative path could be employing umbrella sampling technique to drag the dimer into the membrane and study its structural change. This approach will guarantee that the simulations will not be trapped in local minima and different insertion depths for various initial structures will be visited.

Ultimately, although our simulations are the first attempt of studying IFP dimers, they are far from convergence and longer simulation time or different sampling techniques might be required to converge.

REFERENCES

REFERENCES

1. Kendrew, J.C., et al., *A three-dimensional model of the myoglobin molecule obtained by x-ray analysis*. Nature, 1958. **181**(4610): p. 662-6.
2. Brooks, C.L., 3rd, M. Karplus, and B.M. Pettitt, *Proteins: A Theoretical Prespective of Dynamics, Structure and Thermodynamics*. 1988, New York: J. Wiley.
3. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules*. Nat Struct Biol, 2002. **9**(9): p. 646-52.
4. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins*. Nature, 1977. **267**(5612): p. 585-90.
5. Karplus, M. and G.A. Petsko, *Molecular dynamics simulations in biology*. Nature, 1990. **347**(6294): p. 631-9.
6. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins*. Chem Rev, 2006. **106**(5): p. 1589-615.
7. Schlick, T., *Molecular Modeling and Simulation, an Interdisciplinary Guid*, ed. 2. 2010, New York: Springer.
8. Cheatham, T.E., 3rd and P.A. Kollman, *Molecular dynamics simulation of nucleic acids*. Annu Rev Phys Chem, 2000. **51**: p. 435-71.
9. Norberg, J. and L. Nilsson, *Molecular dynamics applied to nucleic acids*. Acc Chem Res, 2002. **35**(6): p. 465-72.
10. Moraitakis, G. and J.M. Goodfellow, *Simulations of human lysozyme: probing the conformations triggering amyloidosis*. Biophys J, 2003. **84**(4): p. 2149-58.
11. Alder, B.J. and T.E. Wainwright, J Chem Phys, 1957. **27**: p. 1208.
12. Perez, A., F.J. Luque, and M. Orozco, *Dynamics of B-DNA on the microsecond time scale*. J Am Chem Soc, 2007. **129**(47): p. 14739-45.
13. Liwo, A., et al., *Implementation of molecular dynamics and its extensions with the coarse-grained UNRES force field on massively parallel systems; towards millisecond-scale simulations of protein structure, dynamics, and thermodynamics*. J Chem Theory Comput, 2010. **6**(3): p. 890-909.
14. Feig, M. and Z.F. Burton, *RNA polymerase II with open and closed trigger loops: active site dynamics and nucleic acid translocation*. Biophys J, 2010. **99**(8): p. 2577-86.

15. Ramachandran, A., et al., *Coarse-grained molecular dynamics simulation of DNA translocation in chemically modified nanopores*. J Phys Chem B, 2011. **115**(19): p. 6138-48.
16. Pieniazek, S.N., M.M. Hingorani, and D.L. Beveridge, *Dynamical allostherism in the mechanism of action of DNA mismatch repair protein MutS*. Biophys J. **101**(7): p. 1730-9.
17. Mei, C., et al. *Enabling and scaling biomolecular simulations of 100 million atoms on petascale machines with a multicore-optimized message-driven runtime*. in *Proceedings of the 2011 ACM/IEEE Conference on Supercomputing*. 2011.
18. MacKerell, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.
19. Mackerell, A.D., J. Wiorkiewicz-kuczera, and M. Karplus, *An All-Atom Empirical Energy Function for the Simulation of Nucleic-Acids*. Journal of the American Chemical Society, 1995. **117**(48): p. 11946-11975.
20. Jorgensen, W.L., D.S. Maxwell, and J. Tirado-Rives, *Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids*. J Am Chem Soc, 1996. **118**: p. 11225-11236.
21. Schuler, L.D., X. Daura, and W.F. van Gunsteren, *An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase*. J Comp Chem, 2001. **22**(11): p. 1205-1218.
22. Scott, W.R.P., et al., *The GROMOS biomolecular simulation Program package*. J Phys Chem A, 1999. **103**(19): p. 3596-3607.
23. Cornell, W.D., et al., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules*. J Am Chem Soc, 1995. **117**: p. 5179.
24. MacKerell, A.D., Jr., M. Feig, and C.L. Brooks, 3rd, *Improved treatment of the protein backbone in empirical force fields*. J Am Chem Soc, 2004. **126**(3): p. 698-9.
25. Verlet, L., *Computer "experiments" on classical fluids. I. thermodynamical properties of Lennard-Jones molecules*. Phys Rev, 1967. **159**(1): p. 98-103.
26. Hockney, R.W., *The potential calculation and some applications*. Methods Comput Phys, 1970. **9**: p. 135.
27. Swope, W.C., et al., *A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters*. J Chem Phys, 1982. **76**: p. 637-649.

28. Van Gunsteren, W.F. and H. Berendsen, *Algorithms for macromolecular dynamics and constraint dynamics* Mol Phys, 1977. **34**: p. 1311-1327.
29. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water* J Chem Phys, 1983. **79**: p. 926-936.
30. Mahoney, M.W. and W.L. Jorgensen, *A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions*. J Chem Phys, 2000. **112**: p. 8910-8923.
31. Berendsen, H., J. Grigera, and T. Straatsma, *The missing term in effective pair potentials* J Phys Chem 1987. **91**: p. 6269-6271.
32. Freddolino, P.L., et al., *Molecular dynamics simulations of the complete satellite tobacco mosaic virus*. Structure, 2006. **14**(3): p. 437-449.
33. Roux, B. and T. Simonson, *Implicit solvent models*. Biophys Chem, 1999. **78**(1-2): p. 1-20.
34. Fraternali, F. and W.F. vanGunsteren, *An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution*. J Mol Biol, 1996. **256**(5): p. 939-948.
35. Feig, M. and C.L. Brooks, *Recent advances in the development and application of implicit solvent models in biomolecule simulations*. Curr Opin Struct Biol, 2004. **14**(2): p. 217-24.
36. Krol, M., *Comparison of various implicit solvent models in molecular dynamics simulations of immunoglobulin G light chain dimer*. J Comput Chem, 2003. **24**(5): p. 531-46.
37. Ulmschneider, J.P., *Folding simulations of membrane proteins in implicit generalized Born membranes*. Biophys J, 2007: p. 215a-215a.
38. Parsegian, A., *Energy of an ion crossing a low dielectric membrane: solutions to four relevant electrostatic problems*. Nature, 1969. **221**: p. 844-846.
39. Sitkoff, D., K.A. Sharp, and B. Honig, *Accurate calculation of hydration free energies using macroscopic solvent models*. J Phys Chem, 1994. **98**(7): p. 1978-1988.
40. Sitkoff, D., N. Ben-Tal, and B. Honig, *Calculation of alkane to water free energy using continuum solvent models*. J Phys Chem, 1996. **100**: p. 2744-2752.
41. Feig, M., et al., *Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures*. J Comput Chem, 2004. **25**(2): p. 265-84.

42. Ulmschneider, M.B., et al., *A generalized Born implicit-membrane representation compared to experimental insertion free energies*. Biophys J, 2007. **92**(7): p. 2338-2349.
43. Feig, M., ed. *Modeling solvent environment*. 2010, Wiley-vch: Weinheim.
44. Bashford, D. and D.A. Case, *Generalized born models of macromolecular solvation effects*. Annu Rev Phys Chem, 2000. **51**: p. 129-52.
45. Still, W., et al., *Semianalytical treatment of solvation for molecular mechanics and dynamics* J Am Chem Soc, 1990. **112**: p. 6127-6129.
46. Constanciel, R. and R. Conteras, *Self consistent field theory of solvent effects representation by continuum models: Introduction of desolvation contribution*. Theor Chim Acta, 1984. **65**: p. 1-11.
47. Onufriev, A., D.A. Case, and D. Bashford, *Effective Born radii in the generalized Born approximation: the importance of being perfect*. J Comput Chem, 2002. **23**(14): p. 1297-304.
48. Schaefer, M. and M. Karplus, *A comprehensive analytical treatment of continuum Eelectrostatics*. J Phys Chem, 1996. **100**: p. 1578-1599.
49. Jackson, J.D., *Classical electrodynamics*. 2000, New York: Wiley.
50. Lee, M.S., F.R. Salsbury, Jr., and C.L. Brooks, *Novel generalized born methods*. J Chem Phys, 2002. **116**: p. 10606-10614.
51. Lee, M.S., et al., *New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations*. J Comput Chem, 2003. **24**(11): p. 1348-56.
52. Ooi, T., et al., *Accessible surface-areas as a measure of the thermodynamic parameters of hydration of peptides*. Proc Natl Acad Sci U S A, 1987. **84**(10): p. 3086-3090.
53. Eisenberg, D. and A.D. McLachlan, *Solvation energy in protein folding and binding*. Nature, 1986. **319**(6050): p. 199-203.
54. Ulmschneider, M., et al., *Simulating protein folding in implicit generalized born membranes*. Biophys J, 2007: p. 73a-73a.
55. Spassov, V.Z., L. Yan, and S. Szalma, *Introducing an implicit membrane in generalized Born/solvent accessibility continuum solvent models*. J Phys Chem B, 2002. **106**(34): p. 8726-8738.

56. Berneche, S., M. Nina, and B. Roux, *Molecular dynamics simulation of melittin in a dimyristoylphosphatidylcholine bilayer membrane*. Biophys J, 1998. **75**(4): p. 1603-1618.
57. Bennaïm, A. and Y. Marcus, *Solvation Thermodynamics of Nonionic Solutes*. J Chem Phys, 1984. **81**(4): p. 2016-2027.
58. Cheng, X.L., V. Hornak, and C. Simmerling, *Improved conformational sampling through an efficient combination of mean-field simulation approaches*. J Phys Chem B, 2004. **108**(1): p. 426-437.
59. Pitera, J.W. and W. Swope, *Understanding folding and design: Replica-exchange simulations of "Trp-cage" fly miniproteins*. Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7587-7592.
60. Zhou, R.H. and B.J. Berne, *Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water?* Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12777-12782.
61. Efremov, R.G., et al., *A solvent model for simulations of peptides in bilayers. II. Membrane-spanning alpha-helices*. Biophys J, 1999. **76**(5): p. 2460-71.
62. Im, W., M. Feig, and C.L. Brooks, 3rd, *An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins*. Biophys J, 2003. **85**(5): p. 2900-18.
63. Efremov, R.G., et al., *A solvent model for simulations of peptides in bilayers. I. Membrane-promoting alpha-helix formation*. Biophys J, 1999. **76**(5): p. 2448-59.
64. Kessel, A., D.P. Tieleman, and N. Ben-Tal, *Implicit solvent model estimates of the stability of model structures of the alamethicin channel*. Eur Biophys J, 2004. **33**(1): p. 16-28.
65. Bechor, D. and N. Ben-Tal, *Implicit solvent model studies of the interactions of the influenza hemagglutinin fusion peptide with lipid bilayers*. Biophys J, 2001. **80**(2): p. 643-55.
66. Fennell, C.J., C.W. Kehoe, and K.A. Dill, *Modeling aqueous solvation with semi-explicit assembly*. Proc Natl Acad Sci U S A, 2011. **108**(8): p. 3234-3239.
67. Kelly, C.P., C.J. Cramer, and D.G. Truhlar, *A density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters*. J Chem Theory Comput, 2005. **1**(6): p. 1133-1152.
68. Purisima, E. and T. Sulea, *Restoring Charge Asymmetry in Continuum Electrostatics Calculations of Hydration Free Energies*. J Phys Chem B, 2009. **113**(24): p. 8206-8209.

69. Brooks, C.L. and M. Karplus, *Solvent effects on protein motion and protein effects on solvent motion: Dynamics of the active region of lysozyme*. J Mol Biol, 1989. **208**(1): p. 159-181.
70. Hansson, T., C. Oostenbrink, and W.F. Van Gunsteren, *Molecular dynamics simulations*. Curr Opin Struct Biol, 2002. **12**(2): p. 190-196.
71. Forrest, L.R. and M.S.P. Sansom, *Membrane simulations: bigger and better?* Curr Opin Struct Biol, 2000. **10**(2): p. 174-181.
72. Stern, H.A. and S.E. Feller, *Calculation of the dielectric permittivity profile for a nonuniform system: Application to a lipid bilayer simulation*. J Chem Phys, 2003. **118**(7): p. 3401-3412.
73. Tanizaki, S. and M. Feig, *A generalized Born formalism for heterogeneous dielectric environments: application to the implicit modeling of biological membranes*. J Chem Phys, 2005. **122**(12): p. 124706-124713.
74. Feig, M., W. Im, and C.L. Brooks, 3rd, *Implicit solvation based on generalized Born theory in different dielectric environments*. J Chem Phys, 2004. **120**(2): p. 903-11.
75. Marrink, S.J. and H. Berendsen, *Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations*. J Phys Chem, 1996. **100**: p. 16729-16738.
76. Sayadi, M., S. Tanizaki, and M. Feig, *Effect of membrane thickness on conformational sampling of phospholamban from computer simulations*. Biophys J, 2010. **98**(5): p. 805-814.
77. MacCallum, J., D. Bennett, and P. Tieleman, *Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiment*. J Gen Physiol, 2007. **129**: p. 371-377.
78. Panahi, A. and M. Feig, *Conformational sampling of influenza fusion peptide in membrane bilayers as a function of termini and protonation states*. J Phys Chem B, 2010. **114**(3): p. 1407-1416.
79. Jaskierny, A.J., A. Panahi, and M. Feig, *Effect of flanking residues on the conformational sampling of the internal fusion peptide from Ebola virus*. Proteins, 2011. **79**(4): p. 1109-1117.
80. Zamoan, J., et al., *NMR solution structure and topological orientation of monomeric phospholamban in dodecylphosphocholine micelles*. Biophys J, 2003. **85**(4): p. 2589-98.
81. Helfrich, W., *Elastic properties of lipid bilayers: theory and possible experiments*. Z. Naturforsch, 1973. **28**: p. 693-703.

82. Nielsen, C., M. Goulian, and O.S. Anderson, *Energetics of inclusion-induced bilayer deformations*. Biophys J, 1998. **74**: p. 1966-1983.
83. Harroun, T.A., et al., *Theoretical analysis of hydrophobic matching and membrane-mediated interactions in lipid bilayers containing gramicidin*. Biophys J, 1999. **76**: p. 3176-3185.
84. Grabe, M., et al., *Protein interactions and membrane geometry*. Biophys J, 2003. **84**: p. 854-868.
85. Kozlovsky, Y., J. Zimmerberg, and M.M. Kozlov, *Orientation and interaction of oblique cylindrical inclusions embedded in a lipid monolayer: a theoretical model for viral fusion peptides*. Biophys J, 2004. **87**: p. 999-1012.
86. Huang, H.W., *Deformation of free energy of bilayer membrane and its effect on gramicidin channel lifetime*. Biophys J, 1986. **50**: p. 1061-1070.
87. Nielsen, C. and O.S. Anderson, *Inclusion-induced bilayer deformations: effect of monolayer equilibrium curvature*. Biophys J, 2000. **79**: p. 2583-2604.
88. Choe, S., K. Hecht, and M. Grabe, *A continuum method for determining membrane protein insertion energies and the problem of charged residues*. J Gen Physiol, 2008. **131**: p. 563-573.
89. Callenberg, K.M., N.R. Latorraca, and M. Grabe, *Membrane bending is critical for the stability of voltage sensor segments in the membrane*. J Gen Physiol, 2012. **140**: p. 55-68.
90. Naji, A., P.J. Atzberger, and F.L. Brown, *Hybrid elastic and discrete-particle approach to biomembrane dynamics with application to the mobility of curved integral membrane proteins*. Phys Rev Lett, 2009. **102**: p. 138102/1-138102/4.
91. Zhou, Y.C., L. Benzhou, and A.G. Alemayehu, *Continuum electromechanical modeling of protein-membrane interactions*. Phys Rev E, 2010. **82**: p. 41923/1-41923/5.
92. Miloshevsky, G.V., et al., *Electroelastic coupling between membrane surface fluctuations and membrane-embedded charges: continuum multielectric treatment*. J Chem Phys, 2010. **132**: p. 234707/1-11.
93. Dorairaj, S. and T.W. Allen, *On the thermodynamic stability of a charged arginine side chain in a transmembrane helix*. Proc Natl Acad Sci U S A, 2007. **104**: p. 4943-4948.
94. Huang, Q., C.L. Chen, and A. Herrmann, *Bilayer conformation of fusion peptide of influenza virus hemagglutinin: a molecular dynamics simulation study*. Biophys J, 2004. **87**(1): p. 14-22.

95. Vaccaro, L., et al., *Plasticity of influenza haemagglutinin fusion peptides and their interaction with lipid bilayers*. Biophys J, 2005. **88**(1): p. 25-36.
96. Lague, P., B. Roux, and R.W. Pastor, *Molecular dynamics simulations of the influenza hemagglutinin fusion peptide in micelles and bilayers: conformational analysis of peptide and lipids*. J Mol Biol, 2005. **354**(5): p. 1129-41.
97. Volynsky, P.E., et al., *Effect of lipid composition on the "membrane response" induced by a fusion peptide*. Biochemistry, 2005. **44**(44): p. 14626-37.
98. Jang, H., et al., *How to lose a kink and gain a helix: pH independent conformational changes of the fusion domains from influenza hemagglutinin in heterogeneous lipid bilayers*. Proteins, 2008. **72**(1): p. 299-312.
99. Hansmann, U.H. and Y. Okamoto, *New Monte Carlo algorithms for protein folding*. Curr Opin Struct Biol, 1999. **9**(2): p. 177-83.
100. Sugita, Y. and Y. Okamoto, *Replica exchange molecular dynamics method for protein folding*. Chem Phys Lett, 1999. **314**: p. 141.
101. Wilson, I.A., J.J. Skehel, and D.C. Wiley, *Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution*. Nature, 1981. **289**(5796): p. 366-73.
102. Steinhauer, D.A., *Role of hemagglutinin cleavage for the pathogenicity of influenza virus*. Virology, 1999. **258**(1): p. 1-20.
103. Bullough, P.A., et al., *Structure of influenza haemagglutinin at the pH of membrane fusion*. Nature, 1994. **371**(6492): p. 37-43.
104. Bizebard, T., et al., *Structure of influenza virus haemagglutinin complexed with a neutralizing antibody*. Nature, 1995. **376**(6535): p. 92-4.
105. Skehel, J.J., et al., *Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion*. Proc Natl Acad Sci U S A, 1982. **79**(4): p. 968-72.
106. Gruenke, J.A., et al., *New insights into the spring-loaded conformational change of influenza virus hemagglutinin*. J Virol, 2002. **76**(9): p. 4456-66.
107. Cross, K.J., et al., *Composition and functions of the influenza fusion peptide*. Protein Peptide Lett, 2009. **16**: p. 766-778.
108. Kemble, G.W., T. Danieli, and J.M. White, *Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion*. Cell, 1994. **76**: p. 383-391.
109. Tse, F.W., A. Iwata, and W. Almers, *membrane flux through the pore formed by a fusogenic viral envelope protein during cell fusion*. J Cell Biol, 1993. **121**: p. 543-552.

110. Schoch, A.E. and R. Blumenthal, *Role of the fusion peptide sequence in the initial stages of influenza hemagglutinin-induced cell fusion*. J Biol Chem, 1993. **268**: p. 9267-9274.
111. Spruce, A.E., A. Iwata, and W. Almers, *the first milliseconds of the pore formed by a fusogenic viral envelope protein during membrane fusion* Proc Natl Acad Sci U S A, 1991. **88**: p. 3623-3627.
112. Epand, R.M. and R.F. Epand, *Modulation of membrane curvature by peptides*. Biopolymers, 2000. **55**(5): p. 358-63.
113. Epand, R.M., et al., *Membrane interactions of mutated forms of the influenza fusion peptide*. Biochemistry, 2001. **40**(30): p. 8800-7.
114. Zhou, Z., et al., *¹⁵N NMR study of the ionization properties of the influenza virus fusion peptide in zwitterionic phospholipid dispersions*. Biophys J, 2000. **78**(5): p. 2418-25.
115. Han, X., et al., *Membrane structure and fusion-triggering conformational change of the fusion domain from influenza hemagglutinin*. Nat Struct Biol, 2001. **8**(8): p. 715-20.
116. Li, Y.L., et al., *Membrane structures of the hemifusion-inducing fusion peptide mutant GLS and the fusion-blocking mutant GLV of influenza virus hemagglutinin suggest a mechanism for pore opening in membrane fusion*. J Virol, 2005. **79**(18): p. 12065-12076.
117. Durrer, P., et al., *H⁺ Induced membrane insertion of influenza virus hemagglutinin involves the HA2 amino terminal fusion peptide but not the coiled coil region*. J Biol Chem, 1996. **273**(23): p. 13417.
118. Tsurudome, M., et al., *Lipid Interactions of the Hemagglutinin Ha2 Nh2-Terminal Segment during Influenza Virus-Induced Membrane-Fusion*. J Biol Chem, 1992. **267**(28): p. 20225-20232.
119. Han, X., et al., *Interaction of mutant influenza virus hemagglutinin fusion peptides with lipid bilayers: Probing the role of hydrophobic residue size in the central region of the fusion peptide*. Biochemistry, 1999. **38**(45): p. 15052-15059.
120. Ge, M. and J.H. Freed, *Fusion Peptide from Influenza Hemagglutinin Increases Membrane Surface Order: An Electron-Spin Resonance Study*. Biophys J, 2009. **96**(12): p. 4925-4934.
121. Lai, A.L. and L.K. Tamm, *Locking the kink in the influenza hemagglutinin fusion domain structure*. J Biol Chem, 2007. **282**(33): p. 23946-23956.
122. Han, X. and L.K. Tamm, *A host-guest system to study structure-function relationships of membrane fusion peptides*. Proc Natl Acad Sci U S A, 2000. **97**(24): p. 13097-102.

123. Sun, Y. and D.P. Weliky, *Secondary Structure And Membrane Insertion of the Membrane-Associated Influenza Fusion Peptide Probed by Solid State Nuclear Magnetic Resonance*, in *Chemistry Department*. 2009, Michigan State University: East Lansing, MI.
124. Lorieau, J.L., J.M. Louis, and A. Bax, *The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:water interface*. *Proc Natl Acad Sci U S A*, 2010. **107**(25): p. 11341-11346.
125. Feldmann, H., H.D. Klenk, and A. Sanchez, *Molecular biology and evolution of filoviruses*. *Arch Virol*, 1993. **7**: p. 81-100.
126. Fields, B.N., et al., *Fields virology*, ed. 3. 1996, Philadelphia: Lippincott.
127. Malashkevich, V.N., et al., *Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-Å resolution*. *Proc Natl Acad Sci U S A*, 1999. **96**: p. 2662-2667.
128. Ito, H., et al., *Mutational analysis of the putative fusion domain of Ebola virus glycoprotein*. *J Virol*, 1999. **73**: p. 8907-8912.
129. Ruiz-Arguello, M.B., et al., *Phosphatidyl-inositol-dependent membrane fusion induced by a putative fusogenic sequence of Ebola virus*. *J Virol*, 1998. **72**: p. 1775-1781.
130. Freitas, M.S., et al., *Structure of the Ebola fusion peptide in a membrane-mimetic environment and the interaction with lipid rafts*. *J Biol Chem*, 2007. **282**: p. 27306-27314.
131. Luneberg, J., et al., *Structure and topology of the influenza virus fusion peptide in lipid bilayers*. *J Biol Chem*, 1995. **270**(46): p. 27606-14.
132. Macosko, J.C., C.H. Kim, and Y.K. Shin, *The membrane topology of the fusion peptide region of influenza hemagglutinin determined by spin-labeling EPR*. *J Mol Biol*, 1997. **267**(5): p. 1139-48.
133. <http://www.rcsb.org>.
134. Lai, A.L., et al., *Fusion peptide of influenza hemagglutinin requires a fixed angle boomerang structure for activity*. *J Biol Chem*, 2006. **281**(9): p. 5760-70.
135. Sammalkorpi, M. and T. Lazaridis, *Modeling a spin-labeled fusion peptide in a membrane: implications for the interpretation of EPR experiments*. *Biophys J*, 2007. **92**(1): p. 10-22.
136. Sammalkorpi, M. and T. Lazaridis, *Configuration of influenza hemagglutinin fusion peptide monomers and oligomers in membranes*. *Biochim Biophys Acta*, 2007. **1768**(1): p. 30-8.

137. Brooks, B., et al., *CHARMM- a program for macromolecular energy, minimization and dynamics calculations*. J Comp Chem, 1983. **4**: p. 187-217.
138. Feig, M., J. Karanicolas, and C.L. Brooks, *MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology*. J Mol Graph Model, 2004. **22**(5): p. 377-95.
139. Foloppe, N. and A.D. MacKerell, Jr., *All atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data*. J Comp Chem, 2000. **21**: p. 86-104.
140. Ryckaert, J., G. Ciccotti, and H. Berendsen, *Numerical integration of cartesian equations of motions of a system with constraints-molecular dynamics of N-alkanes*. J Comp Chem, 1977. **23**: p. 327-340.
141. Brooks, C.L., M. Berkowitz, and S. Adelman, *Generalized Langevin theory for many-body problems in chemical dynamics gas surface collisions, vibrational energy relaxation in solids and recombination reactions in liquids*. J Chem Phys, 1980. **73**: p. 4353.
142. Feig, M., *Kinetics from Implicit Solvent Simulations of Biomolecules as a Function of Viscosity*. J Chem Theory Comput, 2007. **3**(5): p. 1734-1748.
143. Chocholousova, J. and M. Feig, *Implicit solvent simulations of DNA and DNA-protein complexes: agreement with explicit solvent vs experiment*. J Phys Chem B, 2006. **110**(34): p. 17240-51.
144. Kumar, S., et al., *The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method*. J Comp Chem, 1992. **13**: p. 1011-1021.
145. Kumar, S., et al., *Multidimensional free-energy calculation using the weighted histogram analysis method*. J Comp Chem, 1995. **16**: p. 1339-1350.
146. Chen, J., W. Im, and C.L. Brooks, 3rd, *Balancing solvation and intramolecular interactions: toward a consistent generalized Born force field*. J Am Chem Soc, 2006. **128**(11): p. 3728-36.
147. David L. Nelson, M.M.C., *Principles of Biochemistry*. 4 ed. 2005, New York: W. H. Freeman and company.
148. Delano, W.L. *The PyMol Molecular Graphics System*. 2002; Available from: <http://www.pymol.org>.
149. Chang, D.K., et al., *Self-association of glutamic acid-rich fusion peptide analogs of influenza hemagglutinin in the membrane-mimic environments: effects of positional difference of glutamic acids on side chain ionization constant and intra- and inter-*

- peptide interactions deduced from NMR and gel electrophoresis measurements.* Biochim Biophys Acta, 2005. **1712**(1): p. 37-51.
150. Jahring, P.B., et al., Arch Virol, 1995. **10**: p. 289-292.
 151. Malashkevich, V.N., M. Singh, and P.S. Kim, *The trimer-of-hairpins motif in membrane fusion: Visna virus.* Proc Natl Acad Sci U S A, 2001. **98**(15): p. 8502-6.
 152. Lescar, J., et al., *The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH.* Cell, 2001. **105**(1): p. 137-48.
 153. Yao, Y., et al., *Membrane fusion activity of vesicular stomatitis virus glycoprotein G is induced by low pH but not by heat or denaturant.* Virology, 2003. **310**(2): p. 319-32.
 154. Suarez, T., et al., *Calcium-dependent conformational changes of membrane-bound Ebola fusion peptide drive vesicle fusion.* FEBS Lett, 2003. **535**(1-3): p. 23-8.
 155. Wong, T.C., Biochim Biophys Acta, 2003. **1609**: p. 45-54.
 156. Gomara, M.J., et al., *Roles of a conserved proline in the internal fusion peptide of Ebola glycoprotein.* FEBS Lett, 2004. **569**(1-3): p. 261-6.
 157. Orzaez, M., et al., *Influence of proline residues in transmembrane helix packing.* J Mol Biol, 2004. **335**(2): p. 631-40.
 158. Olson, M.A., I.-C. Yeh, and M.S. Lee. *Molecular Dynamics Simulations of Folding and Insertion of the Ebola virus Fusion Peptide into a Membrane Bilayer.* in *International Conference on Bioinformatics and Computational Biology*. 2008. Las Vegas, NV.
 159. Brooks, B.R., et al., *CHARMM: the biomolecular simulation program.* J Comput Chem, 2009. **30**(10): p. 1545-614.
 160. Tanizaki, S. and M. Feig, *Molecular dynamics simulations of large integral membrane proteins with an implicit membrane model.* J Phys Chem B, 2006. **110**(1): p. 548-56.
 161. Chocholousova, J. and M. Feig, *Balancing an accurate representation of the molecular surface in generalized born formalisms with integrator stability in molecular dynamics simulations.* J Comput Chem, 2006. **27**(6): p. 719-29.
 162. von Heijne, G., *The membrane protein universe: what's out there and why bother?* J Intern Med, 2007. **261**(6): p. 543-57.
 163. Wallin, E. and G. von Heijne, *Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms.* Protein Sci, 1998. **7**(4): p. 1029-38.

164. Boyd, D., C. Schierle, and J. Beckwith, *How many membrane proteins are there?* Protein Sci, 1998. **7**(1): p. 201-5.
165. Singam, E.R., et al., *Molecular dynamic simulation studies on the effect of one residue chain staggering on the structure and stability of heterotrimeric collagen-like peptides with interruption.* Biopolymers, 2012. **97**(11): p. 847-63.
166. Darvas, M., et al., *Anesthetic molecules embedded in a lipid membrane: a computer simulation study.* Phys Chem Chem Phys, 2012. **14**(37): p. 12956-69.
167. Nagarajan, A., J.P. Andersen, and T.B. Woolf, *Coarse-Grained Simulations of Transitions in the E2-to-E1 Conformations for Ca ATPase (SERCA) Show Entropy-Enthalpy Compensation.* J Mol Biol, 2012. **422**(4): p. 575-93.
168. Ulmschneider, J.P. and M.B. Ulmschneider, *Sampling efficiency in explicit and implicit membrane environments studied by peptide folding simulations.* Proteins, 2008. **75**(3): p. 586-597.
169. Baker, N.A., *Improving implicit solvent simulations: a Poisson-centric view.* Curr Opin Struct Biol, 2005. **15**(2): p. 137-43.
170. Callenberg, K.M., et al., *APBSmem: a graphical interface for electrostatic calculations at the membrane.* PLoS One, 2010. **5**(9): p. 1-11.
171. Spolar, R.S., J.H. Ha, and M.T. Record, *Hydrphobic effect in protein folding and other noncovalent processes involving proteins.* Proc Natl Acad Sci U S A, 1989. **86**: p. 8382-8385.
172. Gallicchio, E. and R.M. Levy, *AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling.* J Comput Chem, 2004. **25**(4): p. 479-99.
173. Zacharias, M., *Continuum solvent modeling of nonpolar solvation: improvement by separating surface area dependent cavity and dispersion contributions.* J Phys Chem A, 2003. **107**: p. 3000-3004.
174. Fogolari, F., A. Brigo, and H. Molinari, *Protocol for MM/PBSA molecular dynamics simulations of proteins.* Biophys J, 2003. **85**(1): p. 159-66.
175. Luecke, H., et al., *Structure of bacteriorhodopsin at 1.55 Å resolution.* J Mol Biol, 1999. **291**(4): p. 899-911.
176. Luo, R., L. David, and M.K. Gilson, *Accelerated Poisson-Boltzmann calculations for static and dynamic systems.* J Comput Chem, 2002. **23**(13): p. 1244-53.

177. Sayadi, M. and M. Feig, *Role of conformational sampling of Ser16 and Thr17-phosphorylated phospholamban in interactions with SERCA*. Biochim Biophys Acta, Biomembr, 2013. **1828**(2): p. 577-585.
178. Lindahl, E. and O. Edholm, *Mesosopic undulations and thickness fluctuations in lipid bilayers from molecular dynamics simulations*. Biophys J, 2000. **79**(1): p. 426-433.
179. Freites, J.A., D.J. Tobias, and S.H. White, *Inference connections of a transmembrane voltage sensor*. Proc Natl Acad Sci U S A, 2005. **102**: p. 15059-15064.
180. Aranda-Espinoza, H., et al., *Interaction between inclusions embedded in membranes*. Biophys J, 1996. **71**(2): p. 648-56.
181. Partenskii, M.B. and P.C. Jordan, *Membrane deformation and the elastic energy of insertion: Perturbation of membrane elastic constants due to peptide insertion*. J Chem Phys, 2002. **117**(23): p. 10768-10776.
182. Lundbaek, J.A. and O.S. Andersen, *Spring constants for channel-induced lipid bilayer deformations. Estimates using gramicidin channels*. Biophys J, 1999. **76**(2): p. 889-95.
183. Goulian, M., et al., *Gramicidin channel kinetics under tension*. Biophys J, 1998. **74**(1): p. 328-37.
184. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome*. Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10037-41.
185. Best, R., et al., *Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles*. J Chem Theory Comput, 2012. **8**: p. 3257-3273.
186. Nose, S., *A Unified Formulation of the Constant Temperature Molecular-Dynamics Methods*. J Chem Phys, 1984. **81**(1): p. 511-519.
187. Hoover, W.G., *Canonical Dynamics - Equilibrium Phase-Space Distributions*. Phys Rev A, 1985. **31**(3): p. 1695-1697.
188. Townsley, L.E., et al., *Structures of gramicidins A, B, and C incorporated into sodium dodecyl sulfate micelles*. Biochemistry, 2001. **40**(39): p. 11676-11686.
189. Neale, C., W.F.D. Bennet, and D.P. Tieleman, *Statistical convergence of equilibrium properties in simulations of molecular solutes embedded in lipid bilayers*. J Chem Theory Comput, 2011. **7**: p. 4175-4188.
190. Andersen, O.S. and R.E. Koeppe, *Bilayer thickness and membrane protein function: an energetic perspective*. Annu Rev Biophys Biomol Struct, 2007. **36**: p. 107-30.

191. Liu, Z., C.S. Gandhi, and D.C. Rees, *Structure of a tetrameric MscL in an expanded intermediate state*. Nature, 2009. **461**(7260): p. 120-4.
192. Betanzos, M., et al., *A large iris-like expansion of a mechanosensitive channel protein induced by membrane tension*. Nat Struct Biol, 2002. **9**(9): p. 704-10.
193. Kandasamy, S.K. and R.G. Larson, *Molecular dynamics simulations of model trans-membrane peptides in lipid bilayers: a systematic investigation of hydrophobic mismatch*. Biophys J, 2006. **90**(7): p. 2326-43.
194. de Planque, M.R., et al., *Interfacial anchor properties of tryptophan residues in transmembrane peptides can dominate over hydrophobic matching effects in peptide-lipid interactions*. Biochemistry, 2003. **42**(18): p. 5341-8.
195. Kol, M.A., et al., *Phospholipid flop induced by transmembrane peptides in model membranes is modulated by lipid composition*. Biochemistry, 2003. **42**(1): p. 231-7.
196. Weiss, T.M., et al., *Hydrophobic mismatch between helices and lipid bilayers*. Biophys J, 2003. **84**(1): p. 379-85.
197. Strandberg, E., et al., *Tilt angles of transmembrane model peptides in oriented and non-oriented lipid bilayers as determined by ²H solid-state NMR*. Biophys J, 2004. **86**(6): p. 3709-21.
198. Siegel, D.P., et al., *Transmembrane peptides stabilize inverted cubic phases in a biphasic length-dependent manner: implications for protein-induced membrane fusion*. Biophys J, 2006. **90**(1): p. 200-11.
199. Sparr, E., et al., *Self-association of transmembrane alpha-helices in model membranes: importance of helix orientation and role of hydrophobic mismatch*. J Biol Chem, 2005. **280**(47): p. 39324-31.
200. Im, W. and C.L. Brooks, *Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations*. Proc Natl Acad Sci U S A, 2005. **102**(19): p. 6771-6.
201. Holt, A., et al., *Is there a preferential interaction between cholesterol and tryptophan residues in membrane proteins?* Biochemistry, 2008. **47**(8): p. 2638-49.
202. Esteban-Martin, S., et al., *Orientational landscapes of peptides in membranes: prediction of (2)H NMR couplings in a dynamic context*. Biochemistry, 2009. **48**(48): p. 11441-8.
203. Ozdirekcan, S., et al., *On the orientation of a designed transmembrane peptide: toward the right tilt angle?* J Am Chem Soc, 2007. **129**(49): p. 15174-81.
204. Bond, P.J., et al., *Coarse-grained molecular dynamics simulations of membrane proteins and peptides*. J Struct Biol, 2007. **157**(3): p. 593-605.

205. Wan, C.K., W. Han, and Y.D. Wu, *Parametrization of PACE force field for membrane environment and simulation of helical peptides and helix-helix association*. J Chem Theory Comput, 2012. **8**: p. 300-313.
206. Kim, T. and W. Im, *Revisiting hydrophobic mismatch with free energy simulation studies of transmembrane helix tilt and rotation*. Biophys J, 2010. **99**(1): p. 175-83.
207. Strandberg, E., et al., *Orientation and dynamics of peptides in membranes calculated from 2H-NMR data*. Biophys J, 2009. **96**(8): p. 3223-32.
208. de Planque, M.R., et al., *Sensitivity of single membrane-spanning alpha-helical peptides to hydrophobic mismatch with a lipid bilayer: effects on backbone structure, orientation, and extent of membrane incorporation*. Biochemistry, 2001. **40**(16): p. 5000-10.
209. Holt, A., et al., *Tilt and rotation angles of a transmembrane model peptide as studied by fluorescence spectroscopy*. Biophys J, 2009. **97**(8): p. 2258-66.
210. Helfrich, P. and E. Jakobsson, *Calculation of deformation energies and conformations in lipid membranes containing gramicidin channels*. Biophys J, 1990. **57**(5): p. 1075-84.
211. Kim, T., et al., *Influence of hydrophobic mismatch on structures and dynamics of gramicidin a and lipid bilayers*. Biophys J, 2012. **102**(7): p. 1551-60.
212. Kim, T., S. Jo, and W. Im, *Solid-state NMR ensemble dynamics as a mediator between experiment and simulation*. Biophys J, 2011. **100**: p. 2922-2928.
213. Jo, S. and W. Im, *Transmembrane helix orientation and dynamics: insights from ensemble dynamics with solid-state NMR observables*. Biophys J, 2011. **100**(12): p. 2913-21.
214. Legare, S. and P. Lague, *The influenza fusion peptide adopts a flexible flat V conformation in membranes*. Biophys J, 2012. **102**(10): p. 2270-8.
215. Peskin, C.S., G.M. Odell, and G.F. Oster, *Cellular motions and thermal fluctuations: the Brownian ratchet*. Biophys J, 1993. **65**(1): p. 316-24.
216. Marrink, S.J. and A.E. Mark, *Effect of undulations on surface tension in simulated bilayers*. Journal of Physical Chemistry B, 2001. **105**(26): p. 6122-6127.
217. Chiu, S.W., et al., *Structure of sphingomyelin bilayers: a simulation study*. Biophys J, 2003. **85**(6): p. 3624-35.
218. Stevens, M.J., *Coarse-grained simulations of lipid bilayers*. J Chem Phys, 2004. **121**(23): p. 11942-8.

219. Johansson, A.C. and E. Lindahl, *Amino acid solvation structure in in transmembrane helices from molecular dynamics simulations*. Biophys J, 2006. **91**: p. 4450-4463.
220. Wilson, M.A. and A. Pohorille, *Mechanism of unassisted ion transport across membrane bilyer*. J Am Chem Soc, 1996. **118**: p. 6580-6587.
221. Li, L.B., I. Vorobyov, and T.W. Allen, *Potential of mean force and pKa profile calculation for a lipid membrane-exposed arginine side chain*. J Phys Chem B, 2008. **112**: p. 9574-9587.
222. Vorobyov, I., B. Bekker, and T.W. Allen, *Electrostatics of deformable lipid membrane*. Biophys J, 2010. **98**: p. 2904-2913.
223. Berneche, S. and B. Roux, *Molecular dynamics of the KcsA K(+) channel in a bilayer membrane*. Biophys J, 2000. **78**(6): p. 2900-17.
224. Sayadi, M., in *Department of chemistry*. 2011, Michigan State University: East Lansing.
225. Cheng, S.F., A.B. Kantchev, and D.K. Chang, *Fluorescence evidence for a loose self-assembly of the fusion peptide of influenza virus HA2 in the lipid bilayer*. Mol Membr Biol, 2003. **20**(4): p. 345-51.
226. Han, X. and L.K. Tamm, *pH-dependent self-association of influenza hemagglutinin fusion peptides in lipid bilayers*. J Mol Biol, 2000. **304**(5): p. 953-65.