Molecular Simulation of the Passive Permeation of Small Peptides across Lipid Bilayers

By

Brent L. Lee

Submitted to the graduate degree program in Chemistry and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Chair: Dr. Krzysztof Kuczera

Dr. Marco Caricato

Dr. Carey K. Johnson

Dr. Brian B. Laird

Dr. Susan M. Williams

Date Defended: July 22nd, 2019

The Dissertation Committee for Brent L. Lee certifies that this is the approved version of the following dissertation:

Molecular Simulation of the Passive Permeation of Small Peptides across Lipid Bilayers

Chair: Dr. Krzysztof Kuczera

Date approved: July 26th, 2019

Abstract:

The passive permeation of small peptides across lipid bilayers was studied by using molecular dynamics simulations and umbrella sampling. The knowledge gained in this work furthers our understanding of permeation across cell membranes and provides insight into the intelligent design of future pharmaceutical compounds.

The passive permeation of the three resonant amino acids – phenylalanine, tyrosine, and tryptophan – in blocked form was studied in a bilayer consisting of 50 DOPC lipid molecules. The potential of mean force displays a free energy minimum at the interface, followed by an energy barrier at the center of the bilayer. Translational diffusion constants are surprisingly flat; however, the reorientation of the entire molecule and the amino acid sidechains indicates a significant rotational barrier. A conformational and clustering analysis of phi, psi, chi-1, and chi-2 angles demonstrates that each amino acid adopts different conformations based upon its bilayer depth. Radial distribution functions, coordination numbers, and the number of solvating water molecules were also examined.

The phenylalanine dipeptide was then studied as it permeates lipid bilayers consisting of either 50 DOPC, 50 POPC, or 40 POPC lipid molecules. DOPC lipid molecules are more disordered than POPC lipids. In DOPC, the potential of mean force is therefore broader and more rotational conformations were sampled. All other analyses confirmed our prior, general observations and were surprisingly insensitive to either lipid type or system size.

Position dependent diffusion constants were then calculated for the permeation of the phenylalanine dipeptide using the Fluctuation-Dissipation theorem, Green-Kubo expressions,

Einstein relations, the Hummer Displacement method, and a numerical approximation to the Smoluchowski equation. We found the numerical approximation method to be the most reliable, although the Fluctuation-Dissipation theorem also yields acceptable results when unconstrained simulations were conducted.

Finally, the prior analyses were applied to the permeation of wh5, one of the smallest peptides capable of forming an alpha helix. For most of the permeation process, the alpha helix remains intact and only begins to unravel at select distances in the aqueous region and at the center of the lipid bilayer. The presence of the lipid bilayer influences the tertiary structure of wh5.

Acknowledgements:

I would like to thank my parents and my sister to whom words cannot express my appreciation for their support over the years: Lawrence A. Lee, Constance M. Fink, and Holly E. Youmans. I would also like to thank Dr. Kuczera for his insight, skill, understanding, and, of course, humor. Finally, anyone who has had a positive impact on this document, my time here at KU, and my life, in general, deserves a special word of thanks. I contemplated including a detailed list here; but, unfortunately, the size of this document would easily double! Thank you!

Table of Contents

Abstract iii
Acknowledgements v
Table of Contents vi
List of Figures x
List of Tables xiii

Chapter 1: Introduction and Background 1

1.1. Introduction 2
1.2. Classical Molecular Dynamics Algorithms 7
1.3. Free Energy Simulations 13
1.3.1. Widom Particle Insertion 14
1.3.2. Thermodynamic Integration 15
1.3.3. Umbrella Sampling 17
1.3.4. Metadynamics 17
1.3.5. Adaptive Biasing Force 18
1.3.6. Milestoning 18
1.3.7. Method Comparison 19
1.3.8. Applications 21
1.3.9. Recent Developments 27
1.4. Umbrella Sampling Theory 28
1.5. References 41

Chapter 2: Permeation of the Three Aromatic Dipeptides through Lipid Bilayers: 50

2.1. Overview 51
2.2. Introduction 52
2.3. Methods 55
2.4. Results and Discussion 61
2.4.1. Potentials of Mean Force 61
2.4.2. Translational Diffusion 63
2.4.3. Permeability Measures 65
2.4.4. Rotational Diffusion 68
2.4.5. Insertion Angle 71
2.4.6. Backbone Conformations 72
2.4.7. Sidechain Conformations 75
2.4.8. Peptide Structures – Clustering. 77
2.4.9. Molecular Shape and Size 79
2.4.10. Specific Interactions 79
2.5. Conclusions 82
2.6. Supplementary Information 85
2.6.1. Initial Membrane Electron Density 86
2.6.2. PMF Quartiles 86
2.6.3. Alternative Diffusion Constant Methods 88
2.6.4. Radius of Gyration 90
2.6.5. Solvent Accessible Surface Area 91
2.7. References 94

Chapter 3: Effect of System Size and Lipid Type on the Passive Permeation of the Phenylalanine Dipeptide 101

3.1. Overview 102
3.2. Introduction 103

3.3. Methods.....107
3.4. Results and Discussion 112
3.4.1. Potential of Mean Force.....112
3.4.2. Translational Diffusion 115
3.4.3. Permeability Measures 117
3.4.4. Rotational Sidechain Diffusion 119
3.4.5. Insertion Angle.....122
3.4.6. Backbone Conformations.....125
3.4.7. Sidechain Conformations 127
3.4.8. Peptide Structures and Clustering.....129
3.4.9. Molecular Shapes and Sizes 131
3.4.10. Specific Interactions.....132
3.5. Conclusions.....136
3.6. Supplementary Information.....138
3.6.1. Electron Density Profile.....138
3.6.2. PMF Quartiles 139
3.6.3. Radius of Gyration 142
3.6.5. Solvent Accessible Surface Area.....143
3.7. References 147

Chapter 4: On the Calculation of Diffusion Constants in Passive Membrane Permeation Studies 155

4.1. Overview.....156
4.2. Introduction 157
4.3. Methods.....167
4.4. Results and Discussion 173
4.4.1. The Fluctuation Dissipation Theorem.....173

4.4.2. Green-Kubo Relations 181
4.4.3. Einstein Relations 191
4.4.4. The Hummer Displacement Method 196
4.4.5. Numerical Approximation to the Smoluchowski Equation 197
4.5. Conclusions 199
4.6. References 203

Chapter 5: Permeation of a Small, Helical Peptide: *wh5*. 208

5.1. Overview 209
5.2. Introduction 209
5.3. Methods 211
5.4. Results and Discussion 213
5.5. Conclusions 221
5.6. References 223

Chapter 6: Conclusions and Future Directions 225

6.1. Conclusions 226
6.2. Future Directions 230
6.2.1. Near Future 229
6.2.2. Far Future 232
6.3. References 233

List of Figures

Chapter 1:

Figure 1.1. The anatomy of a phosphatidylcholine. 3

Chapter 2:

Figure 2.1. Chemical structures of studied systems. 55
Figure 2.2. The potential of mean force. 62
Figure 2.3. Translational diffusion constants. 65
Figure 2.4. Sidechain rotational correlation times. 70
Figure 2.5. Insertion angles. 72
Figure 2.6. Ramachandran plots. 74
Figure 2.7. Peptide sidechain conformations. 76
Figure 2.8. Representative structures from trajectory clustering. 78
Figure 2.9. Distributions of O to N distances between blocking groups. 79
Figure 2.10. Average number of water molecules within 0.3 nm of the peptides. 80
Figure 2.11. Average coordination numbers of the three peptides. 82
Figure 2.12. Comparison of the total electron density. 86
Figure 2.13. NAFA potential of mean force quartiles. 87
Figure 2.14. NAYA potential of mean force quartiles. 87
Figure 2.15. NATA potential of mean force quartiles. 88
Figure 2.16. Diffusion constants from unconstrained force fluctuations. 89
Figure 2.17. Diffusion constants from the fluctuation of the umbrella pulling forces. 90
Figure 2.18. Radius of gyration. 91
Figure 2.19. Hydrophilic accessible surface area. 92
Figure 2.20. Hydrophobic accessible surface area. 92
Figure 2.21. Total accessible surface area. 93

Chapter 3:

Figure 3.1. Molecules involved with this study. 107
Figure 3.2. The free energy required to pull the phenylalanine dipeptide. 113
Figure 3.3. Translational diffusion constants. 117
Figure 3.4. The autocorrelation time for the rotation of the phenylalanine sidechain. 121
Figure 3.5. The insertion angle of the dipeptide. 123
Figure 3.6. The correlation times for the reorientation of the insertion angle. 124
Figure 3.7. Ramachandran plots and representative structures from a clustering analysis. 126
Figure 3.8. X_1 and X_2 sidechain angles for the phenylalanine dipeptide. 128
Figure 3.9. The distance between the oxygen and nitrogen atoms. 131
Figure 3.10. The number of water molecules within 0.3 nm. 133
Figure 3.11. Radial distribution functions. 134
Figure 3.12. Coordination numbers. 136
Figure 3.13. Electron density profiles. 139
Figure 3.14. Potential of mean force quartiles in the DOPC 50 system. 140
Figure 3.15. Potential of mean force quartiles in the POPC 40 system. 141
Figure 3.16. Potential of mean force quartiles in the POPC 50 system. 142
Figure 3.17. The radius of gyration. 143
Figure 3.18. The hydrophilic accessible surface area. 144
Figure 3.19. The hydrophobic accessible surface area. 145
Figure 3.20. The total solvent accessible surface area. 146

Chapter 4:

Figure 4.1. $D_z(z)$ calculated from the Fluctuation-Dissipation theorem using umbrella sampling pulling forces from GROMACS simulations. 174
Figure 4.2. $D_z(z)$ calculated from the Fluctuation-Dissipation theorem using force data from a *z*-constrained CHARMM simulation. 177
Figure 4.3. $D_z(z)$ calculated from the Fluctuation-Dissipation theorem using forces from a free, unconstrained CHARMM simulation. 179
Figure 4.4. $D_z(z)$ calculated from a Green-Kubo expression using velocity data from unconstrained GROMACS simulations. 182

Figure 4.5. Autocorrelation functions for *z*-velocities for all ten trajectories from a 20 ps free, unconstrained GROMACS simulation starting at *z* = 1.0 nm. 184

Figure 4.6. Integrated *z*-velocity autocorrelation functions. 185

Figure 4.7. $D_{xy}(z)$ calculated from a 2D Green-Kubo expression using velocity data from unconstrained GROMACS simulations. 186

Figure 4.8. $D_{3d}(z)$ calculated from a three dimensional Green-Kubo expression using velocity data from unconstrained GROMACS simulations. 187

Figure 4.9. $D_{3d}(z)$ calculated from a three dimensional Green-Kubo expression using velocity data from unconstrained GROMACS simulations. 189

Figure 4.10. $D_z(z)$ calculated from a 2D Green-Kubo expression using velocity data from z-constrained CHARMM simulations. 190

Figure 4.11. The mean square displacement for the phenylalanine dipeptide is depicted from a free GROMACS simulation at *z* = 0 for the POPC 50 system. 193

Figure 4.12. Mean square displacement, expanded view. 194

Figure 4.13. $D_z(z)$ calculated from an Einstein relation using position data from an unconstrained, free GROMACS simulation. 195

Figure 4.14. $D_z(z)$ calculated from the Hummer Displacement method using position data from umbrella sampling GROMACS simulations. 197

Figure 4.15. $D_z(z)$ calculated from a numerical solution to the Smoluchowski equation from position data from umbrella sampling GROMACS simulations. 199

Chapter 5:

Figure 5.1. A schematic diagram of the peptide *wh5*. 210

Figure 5.2. The locations of hydrogen bonds 1, 2, and 3 are displayed on *wh5*. 211

Figure 5.3. The potential of mean force for *wh5*. 214

Figure 5.4. The root mean square deviation of the alpha carbons of *wh5*. 215

Figure 5.5. The distance between the two ends of the *wh5* molecule. 216

Figure 5.6. The distance between the center of mass of the tryptophan and histidine sidechains in *wh5*. 217

Figure 5.7. The radius of gyration for *wh5*. 218

Figure 5.8. The number of hydrogen bond pairs within 0.35 nm of *wh5*. 219

Figure 5.9. The oxygen to nitrogen distance for each of the three hydrogen bonds that hold together the *wh5* alpha helix. 221

List of Tables

Table 1.1. Comparison of the advantages and disadvantages of common molecular dynamics algorithms. 12

Table 2.1. The permeation coefficients and mean passage time for NAFA, NAYA and NATA. 66

Chapter 1

Introduction and Background

1.1. Introduction

The diffusion of molecules through membranes plays an extremely important role in biology and the intelligent design of pharmaceutical compounds. According to Deamer *et. al.*, the first unicellular life began from the self-assembly of vesicles formed by phospholipid bilayers.[1] Over time, other molecules and even other vesicles merged with each other to form multi-cellular life. Specifically, Adamala and Szostak have theorized that short, hydrophobic peptide strands, such as *N*-acetyl-L-phenylalanine leucinamide ($AcPheLeuNH_2$), are naturally capable of bridging the gap between lipid bilayers.[2] These peptide strands permeate a lipid bilayer and, due to their hydrophobic amino acid residues, stick to the interior where they can then catalyze the growth of larger phospholipid bilayers. This demonstrates the important intertwining of phospholipid bilayers and peptide structure. However, none of this would have been possible without the unique properties of lipid bilayers.

The most common lipid bilayers found in plant and animal tissues are phosphatidylcholines (PC). One such phosphatidylcholine is 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC), as depicted in Figure 1.1. Each phospholipid contains a headgroup, in this case choline, followed by a phosphate group, then followed by a glycerol linking group, and concluding with an acyl chain, also known as a fatty acid. The first three groups are polar and possess a lower interfacial free energy by being closer to polar solvents such as water; whereas, the acyl chain consists of a wide variety of hydrocarbons in terms of length, saturation, and branching. In aqueous environments, such as the cytoplasm within cells or blood plasma in

exterior regions, the acyl hydrocarbon tails minimize their interfacial free energy by being close to each other and will naturally self-assemble into a lipid bilayer.

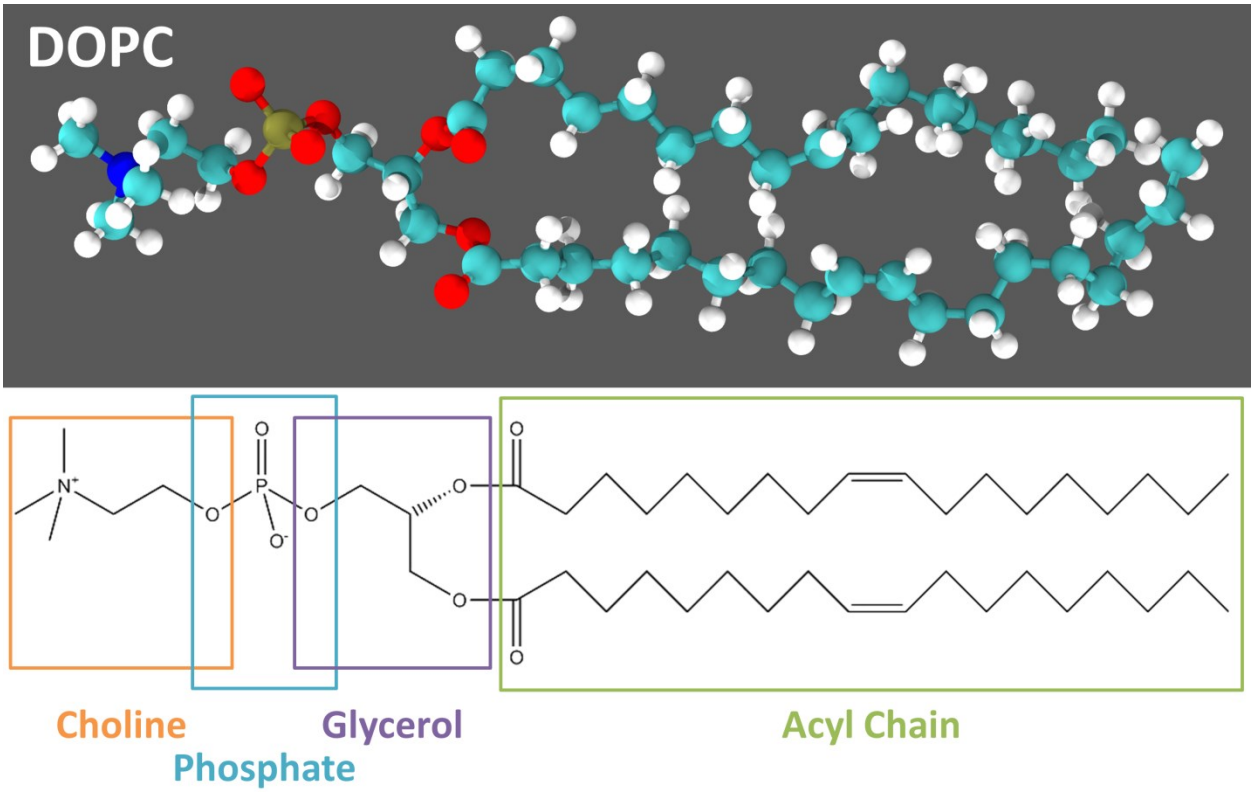


Figure 1.1. The anatomy of a phosphatidylcholine, specifically 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC).

Of course, cell membranes are far more complicated than a simple lipid bilayer. They contain a wide variety of other chemical compounds such as carbohydrates, cholesterol, and a wide variety of proteins. Carbohydrates are often attached to the surface of lipid molecules for the purposes of cell identification and signaling. Cholesterol and other phospholipids are present in cell membranes and create interesting phase equilibria, such as lipid rafts and regions of enhanced transport properties.[3] A wide variety of large, functional proteins are also embedded within cell membranes. For example, cellular respiration within mitochondria relies

upon a rather complicated interplay between cell membranes; respiratory complexes I, II, III, and IV; and, cytochrome C – all of which are present at the surface of a phospholipid membrane.[4]

Instead of taking a complicated, macroscopic approach towards studying membrane permeation, we focus on a more direct atomistic approach from the bottom up by focusing on pure lipid bilayers and simple peptides. Specifically, why study the passive permeation of small peptides across lipid bilayers? As mentioned earlier, from an evolutionary perspective, unicellular life began as a simple lipid bilayer that, over time, trapped other molecules within itself to gradually form the complex cellular structures that we know today.[1-2] However, until within the past couple decades, relatively little was known about precisely how molecules traverse a lipid bilayer. Several theories were then developed that characterized cellular permeation as a function of the natural diffusive properties of a molecule and of the energy required to pass through a lipid bilayer.[5-7] This general knowledge was then used to calculate a wide variety of useful of information such as partitioning coefficients, diffusion rates, permeability coefficients, and other kinetic rate parameters that are important for understanding drug absorption.[8-10] As a result, the simulation of membrane permeation is currently an important aid in the systematic design of novel pharmaceutical compounds. Nature clearly demonstrates the importance of this process because over 30% of all proteins encoded by DNA are designed to manipulate cellular permeation,[11] and drug companies are following suit with over 70% of modern drug targets also aimed at influencing cellular permeation.[12] The molecular simulation of passive membrane permeation allows us to draw

important conclusions and mechanistic insights into the nature of cellular processes and improves our ability to manipulate these processes towards a practical end.

Numerous experimental studies on membrane systems have produced a wealth of information that lays out the foundation of our knowledge of these complex environments, and may also be used to validate simulation data. For example, the lipid membrane undergoes a wide variety of structural and dynamic processes that have been explored through x-ray and neutron scattering.[13-14] These methods can also be used to determine the surface area per headgroup and electron density.[15-16] Lipid dynamics and alkane tail order parameters have been determined by NMR spectroscopy.[17] Infrared spectroscopy has been used to study the dynamics of water inside of a lipid bilayer,[18-20] and fluorescence spectroscopy is commonly used to probe a wide variety of non-lipid molecules inside of membranes.[21-22] The rate of membrane permeation by ligands has been analyzed using the Parallel Artificial Membrane Permeation Assay (PAMPA) method,[23-26] which follows the Caco-2 model of drug absorption through either the intestines or the blood-brain barrier.[27]

Molecular simulations of membranes provide both conceptual and practical information concerning cellular permeation that, in many ways, is complementary to experimental findings. The starting point is reproduction of observed properties in order to confirm the realistic nature of models. Going beyond that, simulations provide a powerful tool for probing the atomistic and molecular properties of membrane structure, dynamics and permeation. Atomistic insights yield valuable mechanistic information on structure, dynamics and interactions of the components. The density of the membrane can be broken up into its constituent pieces, such

as the headgroups, tails, glycerol groups, hydrogen atoms, and carbon atoms.[5] Translational and rotational diffusion of molecules at different depths within the membrane can be calculated.[28-30] For flexible molecules, conformational changes during permeation may be followed, as in the case of Ramachandran plots for peptide systems.[28] The overall orientation of a molecule as it approaches, enters, and permeates a cell membrane may be followed.[28] However, one of the most commonly calculated and discussed properties in membrane permeation studies is the free energy, *G(z)*, as a function of insertion depth of the permeant in the membrane, *z*. This is otherwise known as the potential of mean force (PMF).

Throughout this dissertation, we consider many atomistic insights and additional avenues of inquiry. Chapter 1 concludes in the following sections with some basic background information concerning molecular dynamics algorithms, umbrella sampling, free energy calculations, and experimental methods for studying passive permeation. Chapter 2 then explores the passive permeation of the aromatic dipeptides – phenylalanine, tyrosine, and tryptophan – across a DOPC lipid bilayer consisting of 50 lipid molecules. Chapter 3 explores the influence of different lipid bilayers (monounsaturated vs. saturated) and sizes (50 lipids vs. 40 lipids) on many of the phenomena observed in Chapter 2 within the context of phenylalanine permeation. Chapter 4 then attempts to unravel the complicated nature of calculating diffusion constants within passive permeation processes. Chapter 5 broadens the discussion towards the permeation of a larger peptide, *wh5*, which is capable of forming an alpha helix within itself. Finally, Chapter 6 presents conclusions and offers avenues for future research.

1.2. Classical Molecular Dynamics Algorithms

The primary investigatory tools used throughout this dissertation are molecular dynamics simulations. An incredibly rich and detailed body of literature exists concerning this subject, and we simply present a brief, general introduction to some of the common methods and algorithms that one can use.

Molecular dynamics (MD) simulations are used to determine the properties of a system based upon the trajectories of a collection of molecules. Towards that end, there are a variety of algorithms for determining the motion of molecules. These algorithms attempt to solve, numerically, the equations of motion for the system.

The most common algorithms are finite difference algorithms where a continuous differential equation is replaced by finite time lengths. Most of these methods start out with a Taylor series approximation of position, x, as a function of some very small changes in time, Δt:

$$x(t + \Delta t) = x(t) + x'(t)(\Delta t) + \frac{1}{2}x''(t)(\Delta t)^2 + \frac{1}{3!}x'''(t)(\Delta t)^3 + \cdots. \qquad (1.1)$$

where t represents time and where the apostrophes represent derivatives with respect to time. When this Taylor series expansion is truncated at the first order with respect to Δt, then the simplest MD algorithm, Euler's method, is created. Euler's method uses the following algorithm:[31-32]

1.) Assume a known starting position, $x(t_o)$.

2.) Define a time step, Δt, in order to advance the simulation.

3.) Assume a constant rate of change in position, $x'(t)$, over the time step.

4.) Calculate the initial speed through a Boltzmann distribution of velocities at the desired

temperature.

5.) Find the new position by using the truncated Taylor series expansion:

$$x(t + \Delta t) = x(t) + x'(t)\Delta t \tag{1.2}$$

6.) Re-evaluate the speed at the new position, $x(t + \Delta t)$, by using a potential of interaction between atoms and molecules.

7.) Repeat steps 5 - 6 until finished.

In addition, the same method can be described in terms of velocity since the speed of atoms and molecules is directly related to the interaction potential that it experiences:

1.) Assume a known starting velocity, $v(t_o)$.

2.) Define a time step, Δt, in order to advance the simulation.

3.) Assume a constant rate of change in velocity, $v'(t)$, over the time step.

4.) Calculate the initial acceleration through some means, such as a potential of interaction between atoms and molecules.

$$F = ma = m\frac{dv(t)}{dt} = -\frac{dV}{dx} \tag{1.3}$$

5.) Find the new velocity by using the truncated Taylor series expansion:

$$v(t + \Delta t) = v(t) + v'(t)\Delta t \tag{1.4}$$

6.) Re-evaluate the position based on the new velocity, $v(t + \Delta t)$.

7.) Repeat steps 5 - 6 until finished.

Unfortunately, Euler's method involves a rather large amount of truncation error because higher order terms within the Taylor series expansion are neglected. Also, velocity is not constant over the interval, Δt.

Runge-Kutta (RK) methods are based upon Euler's basic algorithm. However, these methods evaluate the velocity at several different points along the time interval. For example, a 4th order RK method calculates the velocity at four different points along the time interval and then uses those points to determine the next position of the atom or molecule. The algorithm proceeds, as follows:

1.) Assume an initial position, $x(t_o)$.

2.) Define a time step, Δt.

3.) Represent the velocity as a function of both time and the position: $x'(t) = f(t, x)$.

4.) The new position is then defined as:

$$x(t + \Delta t) = x(t) + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \tag{1.5}$$

where

$$k_1 = f(t, x) \tag{1.6}$$

$$k_2 = f\left(t + \frac{1}{2}\Delta t, x + \frac{\Delta t}{2}k_1\right) \tag{1.7}$$

$$k_3 = f\left(t + \frac{1}{2}\Delta t, x + \frac{\Delta t}{2}k_2\right) \tag{1.8}$$

$$k_4 = f(t + \Delta t, x + \Delta t\, k_3) \tag{1.9}$$

In this manner, the position is related to the velocity at several different points.

5.) Advance time by another increment of Δt and repeat step 4.)

The previous example is rather special because the weighted coefficients make it analogous to Simpson's rule; in other words, the velocity is represented not as a constant over the time interval, but as a 2nd order polynomial. More complicated forms of Runge-Kutta algorithms exist that change the weighting, include variable weighting parameters, add additional parameters, or change the functional form of step 4.) Although RK methods possess less truncation and local error, they require additional force and velocity calculations at each time step. The calculation of force is the most computationally intensive part of an MD simulation, and RK algorithms are, therefore, very slow and are only used when a high degree of accuracy and precision is required.

Another drawback of all Euler based methods is that they are not time reversible. In the Euler method, if time is advanced:

$$x(t + \Delta t) = x(t) + x'(t)\Delta t \tag{1.10}$$

When time is reversed:

$$x(t) = x(t + \Delta t) - x'(t + \Delta t)\Delta t \tag{1.11}$$

The velocities in both the forward and reverse time steps are different because they are evaluated at different times. As a result, Euler based methods are not time reversible. The volume of phase space is constantly increasing with each increasing time step and iteration of the simulation. As a result, the system being studied may not be accurately described, particularly after a large number of time intervals, and system properties, such as energy, are not conserved.

The most common symplectic algorithm that corrects these issues is the Verlet method. The Verlet method uses the current and previous positions to predict the future position by combining the Taylor series expansions for both the forward and reverse time steps:

1.) The forward Taylor series time step is:

$$x(t+\Delta t) = x(t) + x'(t)\Delta t + \frac{1}{2}x''(t)\Delta t^2 + \frac{1}{3!}x'''(t)\Delta t^3 + \cdots. \quad (1.12)$$

2.) The reversed Taylor series time step is:

$$x(t-\Delta t) = x(t) - x'(t)\Delta t + \frac{1}{2}x''(t)\Delta t^2 - \frac{1}{3!}x'''(t)\Delta t^3 + \cdots. \quad (1.13)$$

Add the previous two equations together and move the reversed time step term over to the right side. The odd order derivatives also cancel each other out and:

$$x(t+\Delta t) = 2x(t) - x(t-\Delta t) + \frac{1}{2}x''(t)\Delta t^2 + \cdots. \quad (1.14)$$

Due to the cancellation of the odd powered terms, this formulation of the Verlet method is 3rd order with respect to truncation error in the time step.

Furthermore, the velocity can be approximated directly from the position, as Verlet originally proposed:

$$v(t+\Delta t) = \frac{x(t+\Delta t) - x(t-\Delta t)}{2\Delta t} \quad (1.15)$$

Because the velocities are not calculated directly from the simulation, additional error is generated to the second order when calculating energies. This can be corrected by using modified Verlet Algorithms such as the 'Leap Frog' variant. In this algorithm, the position and velocity data is updated in alternating sequences: the position is calculated on every whole time interval, Δt, and the velocity is calculated on every half time interval, $1/2\Delta t$. For example,

if the position is calculated at t = 0, the velocity is then calculated at t = 1/2Δt, and the position is next calculated at $t = \Delta t$. In this manner, the position and velocity data jump or 'leap frog' over each other in terms of which one is most recent in time.[31] For these reasons, where possible, the Leap Frog algorithm has been used for most of the simulations in this Dissertation.

Another common type of symplectic algorithms are various Predictor Corrector (PC) methods. In PC methods, the following process is used:

1.) Predict the position, velocity, and acceleration at time $t + \Delta t$ using current values.

2.) Evaluate the forces present in the system from Newton's equations of motion, $F = ma$, from the new position at time $t + \Delta t$.

3.) Correct the predicted positions, velocities, and accelerations from both the predicted and prior values.

More complicated PC methods can be created by adding in additional prediction, correction, and evaluation steps that are similar to the above process.

Table 1.1. Comparison of the advantages and disadvantages of common molecular dynamics algorithms.

Algorithm	Advantages	Disadvantages
Euler	Simple, fast	1st Order method 2nd Order truncation error Stability issues
Runge-Kutta	Higher order More accurate than Euler Stable Small local fluctuations	Slow Global energy drift
Verlet	Simple Symplectic Conserves energy well 3rd Order	Large local fluctuations Does not naturally calculate velocities
Predictor Corrector	Best conservation of energy	Complex

1.3. Free Energy Simulations

This section has been adapted from a review article, with permission from the publisher, Taylor and Francis.[33]

As mentioned in Section 1.1, one of the most frequently calculated and discussed properties from passive membrane permeation studies is the potential of mean force, *G(z)*. The coordinate most often used in permeation simulations is the insertion depth, *z*, describing the distance of the permeant center-of-mass (COM) from the membrane center. *z = 0* refers to the center of the lipid system, typically a bilayer. *G(z)* gives the reversible work for moving the permeant to position *z* and is defined within an arbitrary constant.[34-35] The relative free energies associated with the permeant presence in each part of the membrane vary significantly, based upon the chemical environment. The exterior of the membrane is typically aqueous — a highly polar environment. The interior of the membrane, however, is composed of hydrocarbon chains from the tails of the lipid molecule — a highly nonpolar, hydrophobic environment. The interface between the model membrane and surrounding water molecules represents a balance between the two. For example, amino acids and most biological molecules contain both hydrophilic and hydrophobic components. In many cases, the short peptides studied so far behave similarly to polar organic molecules, exhibiting a lower free energy at the interface and a higher free energy within both the center of the membrane and the aqueous region.[7, 28] Quantitatively, this information can then be used to calculation permeation and partition coefficients for use in other analyses.[5, 8]

Throughout this work, umbrella sampling has been used to calculate the potential of mean force; however, as seen in Chapter 2 and Chapter 3, this approach has a variety of limitations. As such, throughout the remainder of this section, a review of some of the current and competing free energy methods, such as particle insertion, thermodynamic integration, metadynamics, adaptive biasing force, and milestoning are briefly reviewed and described, as well as a bit more detail concerning modern applications of umbrella sampling. For a more detailed description of theory behind umbrella sampling, please see section 1.4 of this chapter.

1.3.1. - Widom Particle Insertion.

The Widom particle insertion method is based on the relation between the excess chemical potential of a solute in solution, $\Delta\mu^{ex}$, and the solute-solvent interaction energy ΔE:

$$\Delta\mu^{ex} = \mu - \mu^{id} = \left\langle exp(-{}^{\Delta E}/_{RT}) \right\rangle_0 \quad (1.16)$$

where μ is the solute chemical potential, μ^{id} is the reference ideal gas contribution, and $\langle\ldots\rangle_0$ is the average over unperturbed solvent configurations and random solute insertion attempts, R is the gas constant, and T is the absolute temperature.[36-38] For larger solutes, the original Widom approach is inefficient, because a majority of random insertions lead to large values of ΔE, due to atomic overlap between solute and solvent. To overcome this problem, the cavity insertion Widom method was developed, in which the solute is inserted into existing solvent cavities of appropriate size, and the formula for $\Delta\mu^{ex}$ is corrected by the probability of observing such cavities in the pure solvent.[37, 39-40] To calculate the potential of mean force, the chemical

potential of the solute is calculated as a function of *z* by insertion into different regions of a solvated lipid bilayer.

1.3.2. - Thermodynamic Integration.

This method is based on the equation for the derivative of the free energy with respect to an external parameter.[41-42] The initial derivation applied to "alchemical-type" processes, where the system was smoothly changed from a wild-type state (with potential energy U_0) to a mutant state (with potential energy U_1) by varying a coupling parameter λ. As the parameter is varied from $\lambda = 0$ to $\lambda = 1$, the potential energy changes from U_0 to U_1. In the simplest, linear coupling scheme:

$$U(\lambda) = (1-\lambda)U_0 + \lambda U_1 \tag{1.17}$$

The primary thermodynamic integration (TI) equation is then:[41-43]

$$\left(\frac{\partial G}{\partial \lambda}\right)_{\lambda_0} = \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_{\lambda_0} \tag{1.18}$$

where the slope of G with respect to the coupling parameter at $\lambda = \lambda_0$ is evaluated as the average of the derivative of $U(\lambda)$ over a sample of structures generated with potential $U(\lambda_0)$. In the case of linear coupling:

$$\frac{\partial U}{\partial \lambda} = U_1 - U_0. \tag{1.19}$$

When thermodynamic integration was extended to calculations of the derivative of *G* with respect to an internal coordinate ξ (such as a distance, dihedral angle, or a more general collective variable), Equation 1.18 was modified to:

$$\left(\frac{\partial G}{\partial \xi}\right)_{\xi_0} = \left\langle \frac{\partial U}{\partial \xi} \right\rangle_{\xi_0} + RT \left\langle \frac{\partial ln|J|}{\partial \xi} \right\rangle_{\xi_0} \tag{1.20}$$

where the additional term contains the Jacobian, *J*, for the transformation of coordinates between Cartesian coordinates and the full set of generalized coordinates including ξ.[44-45] In the case of simulating transmembrane permeation, the structural coordinate is typically chosen to be '*z*', as defined earlier. Therefore, the *J* term vanishes, and the thermodynamic integration formula may be written as:[46]

$$\left(\frac{\partial G}{\partial z}\right)_{z_0} = \left\langle \frac{\partial U}{\partial z} \right\rangle_{z_0} = -\langle F_z \rangle_{z_0} \tag{1.21}$$

Here, the derivative of the free energy with respect to '*z*' is expressed as the negative of the average *z*-component force acting on the permeant center of mass over a sample of structures corresponding to a fixed value of $z = z_0$. The sample is obtained by performing molecular dynamics simulations with a holonomic constraint on the permeant position. The PMF, *G(z)*, is obtained by integrating the derivative. The advantage of the TI approach is that the fluctuations of the instantaneous values of the force, $\Delta F_z(z,t) = F_z(z,t) - \langle F_z(z,t) \rangle$, may be used to calculate the position-dependent diffusion coefficient, *D(z)*, according to:[47]

$$D(z) = \frac{(k_B T)^2}{\int_0^\infty \langle \Delta F_z(z,0) * \Delta F_z(z,t) \rangle dt} \tag{1.22}$$

1.3.3. - Umbrella Sampling.

In umbrella sampling, a harmonic restraint potential is introduced, $U_r(z) = ½ f(z-z_0)^2$, where f is a force constant and z_0 defines the location of the window where conformational sampling will take place. The system is simulated in the biased potential $U^* = U + U_r$, where U is the normal molecular potential described by the molecular mechanics force field.[48-49] A typical application involves a molecular dynamics simulation with the biased potential U^*, which samples a biased distribution of z values, $P^*(z)$, and a corresponding biased free energy, $G^*(z) = -RT \ln P^*(z)$. The corresponding unbiased PMF G(z) is recovered by subtracting the known function U_r:

$$G(z) = -RT \ln P^*(z) - \frac{1}{2} f (z - z_0)^2 + C \qquad (1.23)$$

where C is an arbitrary constant. Due to the presence of the restraint potential, only a limited range of z is sampled in such a simulation, so a series of simulations is performed, sampling different regions, or windows, of the reaction coordinate z. To generate the overall PMF $G(z)$, the unbiased contributions from the different windows have to be matched by adjusting the constants. An elegant way of doing this is the Weighted Histogram Analysis Method (WHAM) approach, which is designed to minimize the statistical error in the sampled distribution $P(z)$ and PMF $G(z)$.[34-35, 49]

1.3.4. - Metadynamics.

Metadynamics is an adaptive approach in which the system is pushed to explore the full range of a reaction coordinate, or more generally, a collective variable.[50] This is done by adding to the molecular potential 'hills' or Gaussian penalty functions, localized at already explored

regions of conformational space. At the end of the process, the biasing potential is the negative of the PMF, and the system motion along the collective variable is on a flat effective potential landscape, allowing for the complete sampling of the selected coordinate. Several improvements, such as parallel tempering metadynamics[51] and transition-tempered metadynamics,[52] have been introduced and applied.

1.3.5. - Adaptive Biasing Force.

In the Adaptive Biasing Force (ABF) method, the running average of the force acting along a selected coordinate, ξ, is evaluated in bins (according to Equation 1.20)**,** and subtracted from the instantaneous force due to the molecular potential energy.[53-54] The average effective force acting in the chosen coordinate is then close to zero, allowing for efficient conformational sampling in the selected direction. This method yields the derivative of the free energy with respect to the chosen coordinate, without need for constraints or restraints, combining the Thermodynamic Integration approach with enhanced sampling along the reaction coordinate.[54-55]

1.3.6. - Milestoning.

Milestoning is mainly used in kinetic modeling.[56] In this case, a set of fixed conformational states, or anchors, is defined, which span the conformational space of interest. The milestones themselves are hypersurfaces separating the conformational regions assigned to each anchor. Multiple trajectories are initiated at each milestone and transition probabilities $K_{\alpha\beta}(t)$ are calculated, which describe the probability that a trajectory initiated at milestone α at time zero will pass milestone β for the first time at time t. If there are n_α trajectories initiated at

α and $n_{\alpha\beta}(t)$ trajectories hit β at time t, then $K_{\alpha\beta}(t) \approx n_{\alpha\beta}(t)/n_{\alpha}$.[57] These transition probabilities are then used to calculate mean first passage times between all pairs of milestones, providing the full kinetic description for transitions in the coarse-grained space of anchor states. From the transition matrix, $K_{\alpha\beta}$, the stationary milestone populations q_{α} are calculated and can then be converted to a PMF through standard methods.[57-58]

1.3.7. - Method Comparison.

In general, the different methods described above have been shown to enable the efficient simulation of the complex process of transmembrane permeation of small molecules. As simulations switch to larger and more complex permeants, the Widom method is becoming less popular. While metadynamics calculates the whole probability distribution and PMF, the TI, US, and ABF approaches focus on describing the local properties of the PMF and join the pieces together. This makes the latter methods somewhat more flexible. A recent comparison of US and ABF simulations for the bilayer permeation of model compounds found that they gave mostly comparable results.[59] In theoretical terms, ABF exhibits highly appealing features, combining both enhanced sampling and PMF generation without the need for using restraints or constraints. The TI algorithm necessitates the use of holonomic constraints, which are not widely implemented. However, it has the advantage of generating the instantaneous forces acting along the constrained coordinate, which may be used to obtain position-dependent diffusion rates.

The US approach, based on harmonic restraints, is simpler to implement. Some authors use the US instantaneous restraint force $F_{r,z}(z,t) = f(z-z_0)$ in an analog of Equation 1.22 to

estimate $D(z)$. We believe this is not correct, because while $F_z(z,t)$ describes the force acting on the permeant due to interactions with its environment, $F_{r,z}(z,t)$ only reflects the permeant displacement relative to the window center, z_0. While it may be expected that the average total force in an umbrella sampling simulation should be close to zero. Therefore, $< F_z(z,t)> \approx -< F_{r,z}(z,t)>$, and there is no evident reason for the instantaneous values to be related. Some more practical and theoretical aspects of different PMF calculation methods are discussed in more detail in the following reference.[54] Milestoning is quite complex to implement, but has the advantage of reproducing essentially exact system dynamics across the defined conformational space with minimal assumptions.

For all the methods discussed, simulations must be long enough to sample fluctuations in the degrees of freedom orthogonal to the special coordinate z. This is generally difficult to verify. In most cases, it is assumed that the depth of permeant insertion z is the only "slow" coordinate which requires special treatment. However, several additional slow variables that may require special sampling approaches have been suggested recently. One of these is the reorientation of the permeant, which may become difficult to sample as more complex molecules are simulated.[7, 30, 60-63] For example, rotational relaxation times for aromatic sidechains were found to be 10-100 times slower inside of the membrane as compared to the solution.[28] Another slow variable may involve solvent structural fluctuations. The dragging of water into the membrane by charged amino acids [64-65] and slow water density dynamics [66] have been studied and found to occur on a timescale of tens of microseconds.

1.3.8. - Applications.

Throughout this section, we examine some of the many different ways that the previously mentioned methods can be applied towards both calculating and interpreting the free energy and potential of mean force for membrane permeation. For those interested in a broader and more comprehensive survey of recent literature, we direct you towards two of the more recent review articles on membrane simulations written by Awoonor-Williams *et al.*[67] and Xiang *et al.*[68] The former contains an excellent survey of recent literature; whereas, the latter contains a great biological perspective. For those interested in learning how to design and setup their own membrane simulations, including the selection of computer programs, force fields, ensembles, and lipid types, the following review articles and book chapters by Dickey,[69] Tieleman,[70] and Orsi and Essex[71] are excellent starting points. For those interested in the basic theory behind molecular dynamics simulations, the following reference books build a good theoretical foundation.[31-32]

The particle insertion method is one of the earliest and most computationally efficient methods for calculating the relative free energies of small molecules inside and outside of a membrane.[36, 40] It was first used in membrane transport simulations by Marrink and Berendsen[5] in 1994 to simulate the permeation of water through a lipid bilayer consisting of 64 dipalmitoylphosphatidylcholine (DPPC) molecules, with the GROMOS force field and single point charge (SPC) water model.[72] They proposed a 4 region membrane model for interpreting their results: region 1 consists of a largely aqueous phase with small lipid headgroup density; region 2 consists of the largest lipid headgroup density; region 3 contains the largest density of

lipid tail groups; whereas, region 4 represents the center of the membrane where the greatest amount of free volume is present. Their results indicated a roughly trapezoidal free energy curve where the potential of mean force increases steadily through regions 1, 2, and 3 until it decreases slightly inside the middle, region 4, due to a greater excess volume where opposing lipid tails meet. These results were later confirmed by additional follow up simulations in 1996 on both water and ammonia.[6] The particle insertion method has since been improved upon and used to investigate the free energy of a wide variety of small molecules such as oxygen,[6] ethanol,[73] methane derivatives,[37] and small gasses,[39] as well as the effect of carbon lipid length and water, revisited.[40, 74-75] Unfortunately, due to difficulties in sampling large volume cavities, it is of little use for larger permeants.

Thermodynamic integration is also referred to as the *z*-constraint method. The same constrained force data from umbrella sampling simulations can be used to calculate diffusion constants. The first commonly referenced instance of its use was by Marrink and Berendsen in their second article on the permeation of water and small molecules through a lipid bilayer.[6] Since then, thermodynamic integration has been used to determine the free energy of a wide variety of small molecules and pharmaceutical compounds.[29, 76-79] For example, in 2009, Orsi, Sanderson, and Essex conducted a thermodynamic integration study on several small molecules. Their simulation consisted of 128 dimyristoylphosphatidylcholine (DMPC) lipid molecules using a coarse grain model[80] and 3400 water molecules using a soft sticky dipole model[81] and was run using their own custom software, BRAHMS[80], for up to 80 ns. Their free energy results follow the same trend described at the start of this section. Hydrophilic molecules show an increase in free energy towards the center of the membrane; whereas,

hydrophobic molecules like ethane demonstrate a decrease in free energy at the central location. More interestingly, molecules with both hydrophilic and hydrophobic components demonstrate a sinusoidal pattern where the free energy decreases near the interface (Regions II and III) but increases at the center of the membrane. This is a pattern that is commonly witnessed with biological molecules.

Umbrella sampling has been the most popular method for determining the potential of mean force over the past couple decades and has been used in a wide variety of simulations of amino acids[7, 9, 61, 64], ions[65, 82-83], pharmaceutical compounds[84-86], and various membrane models[87-89]. In our work, we have studied the permeation of three aromatic dipeptides with phenylalanine, tyrosine, and tryptophan sidechains, as well as a small helical peptide, *wh5*. Our simulations comprised either 40 or 50 1,2-Dioleoyl-sn-glycero-3-phosphocholine (DOPC) lipid molecules using the CHARMM 36 force field [90] and TIP3P water molecules.[72, 91] All simulations were setup with CHARMM[92-93] and were then run with GROMACS 4.5.6 software[94] for 50 to 100 ns for each umbrella sampling window. Our results can be found in Chapter 2 (Figure 2.2). All of the amino acids, relative to water, possess a smaller free energy near the interface of the membrane where their hydrophobic moiety can interact with the lipid membrane and where their hydrophilic moiety can interact with the polar lipid headgroups and aqueous region. The larger systems, blocked tryptophan and *wh5*, exhibited relatively large free energy cost to enter the center of the membrane, likely due to the destruction of the solvation shell.[28] It should, however, be noted that many ionic amino acids do not exhibit this pattern and their permeation of a membrane is due to the creation of a water filament with a greatly reduced free energy penalty through the center of the membrane.[64-65, 83] We observed little difference

in the potential of mean force as a result of changes in system size or lipid type. This is in accord with work of Naomi *et. al.*, who demonstrated moderate free energy differences due to even larger changes in system size.[95] The small system size likely minimized the effects of membrane deformations that occur on long timescales[87], and we recommend that future studies ensure that the free energy has converged through a wide variety of parameters such as length and starting conformation.[96] The Ramachandran plots in Chapter 2 (Figure 2.6) demonstrate that molecules also adopt different conformations based upon their location within the membrane; this is a topic that has been explored in depth by other methods, as discussed later on in this section.

Recently, Carpenter *et al.*[10] conducted an umbrella sampling study to predict the permeation of twelve pharmaceutical compounds through the blood brain barrier. Their systems were comprised of 72 DOPC molecules using the Berger coarse grained force field [97] and single point charge water.[72] Each umbrella sampling window was run for 45 ns by using GROMACS 4.5.5.[94] All of the drugs exhibited a free energy minimum at the interface of the membrane, albeit to varying degrees. Perhaps unsurprisingly, the molecules most capable of passing through the blood-brain barrier were those with a free energy lower in the membrane than in water. For example, nordazepam, imipramine, chlorpromazine, and promazine are all experimentally known to pass through the blood brain barrier; whereas, atenolol, salbutamol, salicylate, and cimetidine do not. Those with higher free energy barriers are all capable of extensive hydrogen bonding; whereas, those with lower free energies are largely aliphatic with a small number of hydrogen bond donors and acceptors.

Unfortunately, umbrella sampling, although popular, has issues with convergence, and its derivation assumes only one free energy barrier towards permeation, such as translational movement through the bilayer, when permeant reorientation and membrane deformations may be more important.[62, 87, 96] As a result, a large amount of research over the past decades has gone towards devising new methods that may more accurately or more efficiently describe lipid membrane permeation. One such example is metadynamics, as devised by Laio in 2002[50] and recently revised by Voth in 2016.[52] In 2013, Bochicchio *et al.*[98] compared umbrella sampling with metadynamics and found both approaches to yield similar results for simple polymer systems, while metadynamics proved to be more computationally efficient. Earlier, Jämbeck and Lyubartsev[99] conducted another comparison between umbrella sampling and metadynamics by using ibuprofen, aspirin, and diclofenac. Although their work was inconclusive in terms of which method better describes the potential of mean force, they did demonstrate the importance of sampling internal degrees of freedom by comparing results with both *cis* and *trans* ibuprofen — the *trans* conformation better reproduced experimental results.

Adaptive biasing force is another relatively modern method and is in continued development by Pohorille.[53-55] In 2014, Comer *et al.*[63] used this method to examine a wide variety of questions concerning lipid bilayer simulations, such as the van der Waals cutoff, the choice of water model, and bilayer size on the permeation of water through a POPC lipid bilayer. Their simulations used 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid molecules, the CHARMM 36 force field,[90] and were run with NAMD.[100] They found that varying van der Waals cutoffs between 0.7 and 1.2 nm had little impact on free energy values.

Simulations with TIP4P-Ew water yielded similar free energy profile to TIP3P, with the PMF only about a $kJmol^{-1}$ higher within the membrane center.[101] Finally, increasing the system size from 40 to 100 POPC molecules only slightly changed the free energy of permeation. However, it should be noted that these changes did influence other properties, such as the overall permeability. Although not directly related to membrane permeation, Chipot's earlier work using this method may also be useful to those replicating this approach.[102]

Milestoning represents one of the newest and most novel means for determining the free energy. Specifically, one-dimensional Milestoning (along the permeation coordinate, *z*) was applied to study the transport of blocked tryptophan (NATA) through a model DOPC bilayer by Elber *et al.*,[7] using the Berger parameter set[97] and the MOIL simulation package.[103] The results yielded a permeation time of about 4 hours, in good agreement with experimental estimates of 8 hours from the parallel artificial mebrane permeation assay. Their results obtained with the solubility-diffusion model for NATA also yielded a very long permeation time of around 15 min. The slow permeation time for NATA was traced to a very high free energy barrier, about 75 $kJmol^{-1}$, which is significantly higher than a simple sidechain model with 1-methylindol (around 16 $kJmol^{-1}$) or our recent umbrella sampling results for the same system of 44 $kJmol^{-1}$.[28] Partial solvation of NATA was found throughout the permeation process and indicates the formation of membrane defects.[7] A further study performed milestoning on NATA membrane permeation in the two-dimensional space of insertion depth *z* and insertion angle θ.[30] Here, θ was defined as the angle between the *z* axis and the vector between centers of the NATA backbone and sidechain. Interestingly, the computed free energy barrier was essentially the same as in the case of using *z* only, indicating that NATA reorientation was not a

'hidden slow variable', but was adequately sampled in previous studies. Another Milestoning study of a membrane environment involved simulations of transitions between states corresponding to different number densities.[66] This was used to study the dynamics and distribution of cavities inside the membrane, which may facilitate the permeation of nonpolar molecules. Additionally, water permeation was characterized in the same work. While calculated permeabilities were somewhat larger than previously calculated, an improved estimate of the water transition time across a lipid bilayer (around 10 μs) was obtained. Milestone lifetimes along the permeation coordinate were calculated, with most values being short (single ps and below), except for locations in the leaflet centers, which reached hundreds of ps.

1.3.9. - Recent Developments.

New developments in free energy simulations of membrane permeation are occurring in several directions. One direction involves development and application of new, improved algorithms, such as transition tempered metadynamics,[52] which improve convergence of the free energy, or the modified WHAM method,[104] which enhances data analysis. Another direction is the use of multiple copies of the permeant, which come in several types. At the simplest level, multiple independent copies of the permeating molecule are propagated in a single simulation, with sufficient spacing between copies to avoid interactions.[95] A related approach in this category is to use multiple-walker versions of metadynamics and ABF.[105-106] A slightly different approach applies replica-exchange ideas, propagating multiple trajectories to enhance sampling – including replica-exchange molecular dynamics, bias-exchange

metadynamics, and replica-exchange umbrella sampling.[59, 107] Transition-tempered metadynamics takes a novel approach to the objective of rare event sampling, where Gaussian biasing functions are exponentially tempered based on known barriers in the system such as the center of membranes in permeation simulations or protein folding events.[52, 108] A separate direction is the development of molecular mechanics force fields for lipids; besides the Berger set,[97] improved versions of CHARMM [90] and GROMOS [109-110] parameters have recently been published.

The idea of moving beyond the simple inhomogenous solubility-diffusion model of permeation is undergoing increased scrutiny. On one hand, the possibility of several slow degrees of freedom that require enhanced sampling is being considered. Besides the membrane insertion *z*, these include both permeant orientations and coupled slow solvent-membrane fluctuations. As more complex permeants are being modeled, internal conformational sampling of molecules will have to be considered as well.[28] Finally, memory effects in permeant dynamics are being investigated by the analysis of non-Markovian behavior [111] and the consideration of subdiffusive motion.[112]

1.4. Umbrella Sampling Theory

Umbrella sampling is a very important method that is used throughout this dissertation; and, therefore, a discussion concerning the underlying theory and derivation of the general umbrella sampling equations that were mentioned earlier may prove useful.

Umbrella sampling is a method for determining the energy differences within a system that contains thermodynamically inaccessible regions of phase space that are of interest. A typical molecular dynamics simulation has difficulty sampling regions of phase space that are separated by a high energy barrier. For example, when a protein diffuses across a lipid bilayer, the protein encounters a large energy barrier as it travels from one side of the membrane to the other. Experimental evidence suggests that this process takes around eight hours to occur naturally and unassisted.[7] A reasonably sized molecular dynamics simulation of such a system with 64 modern processors would take around 4 billion years to model this process – roughly the age of the Earth! Thankfully, this problem has been considered in detail by others, and one of the oldest and most common solutions is to use umbrella sampling.

Umbrella Sampling works by adding a bias potential to the system Hamiltonian that lowers the free energy of particular configurations of interest or, more commonly, raises the free energy of regions that are not of interest. Simulations are then repeated for other regions of interest and are allowed to access the aforementioned thermodynamically inaccessible regions of phase space for the unbiased system. Finally, at the end, all of the separate simulations are combined together to represent a continuous function of the free energy for the unbiased and unweighted system that is truly of academic interest.

Unfortunately, the mathematics behind this process are nontrivial and are historically split into two separate pathways. The first description of Umbrella sampling began with Torrie and Valleau and built upon the Acceptance Ratio Method of Bennet *et. al.* before them.[48, 113-114] The derivation begins with basic thermodynamic definitions of the unbiased system of interest

and the biased system over which the simulation will actually be run. The subscript '*u*' denotes the unbiased system and the subscript '*b*' denotes the biased system. In a canonical ensemble where *N*, *V*, and *T* are constant, the Helmholtz free energy, *A*, is given by:

$$A_u = -k_B T_u \ln(Q_u) \tag{1.24}$$

where Q_u is the partition function for the unbiased system. Assume a second system of interest exists where:

$$A_b = -k_B T_b \ln(Q_b) \tag{1.25}$$

where A_b and Q_b are the associated variables for this new, biased system. Simplify and take the difference of the two, previous equations and:

$$\Delta A = -k_B T \ln \frac{Q_b}{Q_u} \tag{1.26}$$

assuming that the temperature of the biased and unbiased systems is the same.

For conformational processes, the contributions of kinetic energy to ΔA cancel each other out, and we can express the free energy change in terms of the configuration integral, *Z*.

$$Z = \int e^{-\beta U(r^N)} dr^N \tag{1.27}$$

where $\beta = (k_B T)^{-1}$.

ΔA then becomes:

$$\Delta A = -k_B T \ln \frac{Z_b}{Z_u} \tag{1.28}$$

$$\Delta A = -k_B T \ln\left(\frac{\int e^{-\beta U_b(r^N)} dr^N}{\int e^{-\beta U_u(r^N)} dr^N}\right) \tag{1.29}$$

Now, assume that both integrals are over the same configuration space, as written up above. Notice that the partition function usually contains an integral over all of phase space, over both the positions and momenta. However, the internal energy is not a function of the momentum of the system and the associated integral over all possible momenta cancels out in the above ratio. Next, the biased system differs from the unbiased system by an offset potential, *ΔU*:

$$U_b(r^N) = U_u(r^N) + \Delta U(r^N) \tag{1.30}$$

$$\Delta U(r^N) = U_b(r^N) - U_u(r^N) \tag{1.31}$$

ΔA now becomes:

$$\Delta A = -k_B T \ln\left(\frac{\int e^{-\beta[\Delta U(r^N) + U_u(r^N)]} dr^N}{\int e^{-\beta U_u(r^N)} dr^N}\right) \tag{1.32}$$

Distribute the -β in the numerator and substitute in the given equation for *ΔU* and this becomes:

$$\Delta A = -k_B T \ln\left(\frac{\int e^{-\beta[U_b(r^N) - U_u(r^N) + U_u(r^N)]} dr^N}{\int e^{-\beta U_u(r^N)} dr^N}\right) \tag{1.33}$$

$$\Delta A = -k_B T \ln\left(\frac{\int e^{-\beta U_b(r^N) + \beta U_u(r^N) - \beta U_u(r^N)} dr^N}{\int e^{-\beta U_u(r^N)} dr^N}\right) \tag{1.34}$$

Split the exponent in the numerator into two parts and:

$$\Delta A = -k_B T \ln\left(\frac{\int \left[e^{-\beta U_b(r^N) + \beta U_u(r^N)}\right]\left(\boldsymbol{e^{-\beta U_u(r^N)}}\right) dr^N}{\int \boldsymbol{e^{-\beta U_u(r^N)}} \boldsymbol{dr^N}}\right) \tag{1.35}$$

The bolded parts in the numerator and denominator and their associated integrals represent a thermodynamic average of the probability of finding a certain energy difference between the biased and unbiased states. In other words, what is the probability of finding a particular energy difference between states for a given unbiased state? This can be denoted as the following equation:

$$\Delta A = -k_B T \ln\langle e^{-\beta U_b(r^N)+\beta U_u(r^N)}\rangle_u \tag{1.36}$$

$$\Delta A = -k_B T \ln\langle e^{-\beta \Delta U(r^N)}\rangle_u \tag{1.37}$$

Note: the previous three equations are simply different ways of expressing the same thing. Now, the thermodynamic average is over all possible configurations of the unbiased system. Instead of integrating over all possible configurations, the integral can be expressed over all possible energy differences, *ΔU*, by weighting each energy difference by its probability density, $p_u(\Delta U)$.

$$\Delta A = -k_B T \ln \int_{-\infty}^{\infty} p_u(\Delta U) e^{-\beta \Delta U} d(\Delta U) \tag{1.38}$$

$p_u(\Delta U)$ can then be found by running a simulation, binning configurations, and creating a probability histogram. Although the above equation forms the theoretical foundation for umbrella sampling, its practical uses are rather limited. This equation is only valid when the two distributions overlap each other, as discussed by Bennet in his Acceptance Ratio Method.[114] Unfortunately, a definition of what constitutes 'good overlap' largely depends on the error associated in the final results. Even then, it is rarely possible to achieve decent overlap between the two systems with one and only one simulation run because large negative

values of ΔU will cause the above integral to approach an unreasonable large value and overemphasize configurations that have low probabilities.[31]

To correct these problems, the probability distribution can be represented as a Markov chain of many different states that are averaged over the course of an entire simulation. A Markov chain only depends on the current, initial state and not on the past historical events of the simulation. In order to do this, a new probability distribution, $\pi(r^N)$, is created based upon a weighting function, $w(r^N)$:

$$\pi(r^N) = \frac{w(r'^N)e^{-\beta U_u(r'^N)}}{\int w(r^N)e^{-\beta U_u(r^N)}dr^N} \tag{1.39}$$

The above equation represents the degree of weighted overlap between a simulation run at coordinates r' as compared to the entire weighted simulation; this forms the foundation for stitching together various pieces of the simulation. $w(r^N)$ is then chosen to emphasize the coordinates where both the biased and unbiased systems can be sampled between both pieces of the simulation. In other words, inaccessible states are either discarded or de-emphasized between the different simulation windows. Therefore, the probability of sampling a configuration, r^N, in phase space is now π. Switching to a new probability distribution is a consequence of both systems having different energies, which results in different probability distributions. In order to infer something about the unbiased system from the biased system, this difference needs to be known and taken into account within the relevant thermodynamic averages. The thermodynamic average can now be written as:

$$\langle e^{-\beta\Delta U(r^N)}\rangle_u = \frac{\int e^{-\beta U_b(r^N)}dr^N}{\int e^{-\beta U_u(r^N)}dr^N} \tag{1.40}$$

$$\frac{\int e^{-\beta U_b(r^N)} dr^N}{\int e^{-\beta U_u(r^N)} dr^N} = \frac{\int \frac{\pi(r^N)}{\pi(r^N)} e^{-\beta U_b(r^N)} dr^N}{\int \frac{\pi(r^N)}{\pi(r^N)} e^{-\beta U_u(r^N)} dr^N} = \frac{\int \frac{e^{-\beta U_b(r^N)}}{\pi(r^N)} \pi(r^N) dr^N}{\int \frac{e^{-\beta U_u(r^N)}}{\pi(r^N)} \pi(r^N) dr^N} = \frac{\langle \frac{e^{-\beta U_b(r^N)}}{\pi(r^N)} \rangle_\pi}{\langle \frac{e^{-\beta U_u(r^N)}}{\pi(r^N)} \rangle_\pi} \quad (1.41)$$

$$\langle e^{-\beta \Delta U(r^N)} \rangle_u = \frac{\langle \frac{e^{-\beta U_b(r^N)}}{\pi(r^N)} \rangle_\pi}{\langle \frac{e^{-\beta U_u(r^N)}}{\pi(r^N)} \rangle_\pi} = \frac{\langle \frac{e^{-\beta \Delta U(r^N)}}{\pi(r^N)} \rangle_\pi}{\langle \frac{1}{\pi(r^N)} \rangle_\pi} \quad (1.42)$$

The free energy difference can now be found:

$$\Delta A = -k_B T \ln \langle e^{-\beta \Delta U(r^N)} \rangle_u = -k_B T \ln \left[\frac{\langle \frac{e^{-\beta \Delta U(r^N)}}{\pi(r^N)} \rangle_\pi}{\langle \frac{1}{\pi(r^N)} \rangle_\pi} \right] \quad (1.43)$$

Split up the natural logarithm using the division property of logarithms:

$$\Delta A = -k_B T \ln \left[\frac{\langle \frac{e^{-\beta \Delta U(r^N)}}{\pi(r^N)} \rangle_\pi}{\langle \frac{1}{\pi(r^N)} \rangle_\pi} \right] = -k_B T \ln \left[\langle \frac{e^{-\beta \Delta U(r^N)}}{\pi(r^N)} \rangle_\pi \right] + k_B T \ln \left[\langle \frac{1}{\pi(r^N)} \rangle_\pi \right] \quad (1.44)$$

Again, for practical purposes, the above equation is not particularly useful since rarely is the overall free energy difference between two systems of interest. Also, the determination of π is typically done through trial and error and requires some foreknowledge about the system that is being studied.[31] However, the above does clearly demonstrate that sufficient overlap is required between the separate simulation windows that make up the Markov chain. If there is not sufficient overlap, then $\pi(r^N)$, as defined by Equation 1.39, will be negligible, and the above calculation will deviate towards infinity.

The second historical derivation concerning the mathematics of umbrella sampling is more recent and involves a similar process over a particular coordinate of interest, as

demonstrated by a recent review article by Kaestner.[115] Begin with the canonical partition function:

$$Q = \int e^{-\beta U(r^N)} dr^N \tag{1.45}$$

However, instead of integrating over all of phase space, r^N, a coordinate of interest is removed from the integration by using the process described below. Such a coordinate could be a reaction pathway; torsional degree of freedom; or, of interest to this work, the path an amino acid takes as it diffuses through a lipid bilayer. After removing this coordinate from the integration in Equation 1.45, the partition function is now a function of the coordinate of interest, and a delta function must be introduced into the integral to account for this newfound dependence. Below, let '$a(r^N)$' represent the coordinate of interest as a function of r^N, the system coordinates, and let 'a' represent the currently desired value of the coordinate of interest:

$$p(a) = \int \{\delta[a(r^N) - a] e^{-\beta U(r^N)}\} dr^N \tag{1.46}$$

where *p(a)* represents the partitioning of energy states orthogonal to the coordinate *a*. The above integral now integrates the probability distribution over all degrees of freedom other than the one of interest. For example, if the coordinate of interest is the *z* axis of a simulation, and the desired value of *z* is 2.0, then the above function counts all possible configurations of the system in the *xy* plane at *z* = 2.0. The probability of finding the system at a particular coordinate of interest is then:

$$P(a) = \frac{p(a)}{Q} = \frac{\int \{\delta[a(r^N) - a] e^{-\beta U(r^N)}\} dr^N}{\int e^{-\beta U(r^N)} dr^N} \tag{1.47}$$

Now, introduce the energy of the biased system, as follows:

$$U_b(r^N) = U_u(r^N) + \Delta U(a) \tag{1.48}$$

Note that the biased potential is only a function of the coordinate of interest, a.

$$P^b(a) = \frac{\int \left\{\delta[a(r) - a] e^{-\beta\left(U_u(r^N)+\Delta U(a)\right)} dr^N\right\} dr}{\int e^{-\beta\left(U_u(r^N)+\Delta U(a)\right)} dr^N} \tag{1.49}$$

$$= \frac{\int\{\delta[a(r) - a] e^{-\beta U_u(r^N)} e^{-\beta\Delta U(a)} dr^N\} dr}{\int e^{-\beta\left(U_u(r^N)+\Delta U(a)\right)} dr^N} \tag{1.50}$$

The Boltzmann factor for the biased potential, in purple above, is not a function of r^N; therefore, it can be pulled out of the integral:

$$P^b(a) = e^{-\beta\Delta U(a)} \frac{\int\{\delta[a(r) - a] e^{-\beta U_u(r^N)} dr^N\} dr}{\int e^{-\beta\left(U_u(r^N)+\Delta U(a)\right)} dr^N} \tag{1.51}$$

The above equation can be rearranged and solved for the integral in the numerator:

$$\int \{\boldsymbol{\delta[a(r) - a] e^{-\beta U_u(r^N)} dr^N}\} \boldsymbol{dr} = P^b(a) * e^{\beta\Delta U(a)} * \int e^{-\beta\left(U_u(r^N)+\Delta U(a)\right)} dr^N \tag{1.52}$$

Recall that the probability distribution for the unbiased system is of interest and not of the biased system, as discussed up above.

$$P^u(a) = \frac{\int \{\boldsymbol{\delta[a(r) - a] e^{-\beta U_u(r^N)} dr^N}\} \boldsymbol{dr}}{\int e^{-\beta U_u(r^N)} dr^N} \tag{1.53}$$

From the previous two equations, the two parts in bold are the same, and the unbiased probability distribution can be written as:

$$P^u(a) = P^b(a) * e^{\beta \Delta U(a)} \frac{\int e^{-\beta \Delta U(a)} e^{-\beta U_u(r^N)} dr^N}{\int e^{-\beta U_u(r^N)} dr^N} \quad (1.54)$$

The fraction is simply the thermodynamic average of the weighted, biased Boltzmann factor over the unbiased partition function:

$$P^u(a) = P^b(a) * e^{\beta \Delta U(a)} \langle e^{-\beta \Delta U(a)} \rangle_u \quad (1.55)$$

The above function represents the unbiased partition function for a given coordinate of interest with respect to the potential bias and the biased partition function. The free energy is then given by:

$$A_u(a) = -k_B T \ln[P^u(a)] \quad (1.56)$$

$$= -k_B T \ln\left[P^b(a) * e^{\beta \Delta U(a)} \langle e^{-\beta \Delta U(a)} \rangle_u\right] \quad (1.57)$$

$$= -k_B T \ln[P^b(a)] - k_B T \ln\left[e^{\beta \Delta U(a)}\right] - k_B T \ln\left[\langle e^{-\beta \Delta U(a)} \rangle_u\right] \quad (1.58)$$

$$= -k_B T \ln[P^b(a)] - \Delta U(a) - k_B T \ln\left[\langle e^{-\beta \Delta U(a)} \rangle_u\right] \quad (1.59)$$

If *ΔU(a)* is given by a known function, then its average value over a given simulation is simply a constant value.

$$A_u(a) = -k_B T \ln[P^b(a)] - \Delta U(a) + C \quad (1.60)$$

$$C = -k_B T \ln\left[\langle e^{-\beta \Delta U(a)} \rangle_u\right] \quad (1.61)$$

$A_u(a)$ is usually given the name "Potential of Mean Force." It represents the energy required to move a subset of a simulation along a coordinate of interest. This derivation uses the Helmholtz free energy; and, in condensed phases, the Helmholtz and Gibb's free energies are

approximately the same. Therefore, Equation 1.60 is equivalent to Equation 1.23 in the prior review of free energy methods, Section 1.3.3.

As mentioned before, the probability distribution can be obtained from a histogram of coordinates from a classical molecular dynamics simulation, and *ΔU* is specified as the biased potential. The potential bias is typically a harmonic potential of the form:

$$\Delta U = \frac{1}{2} k \left(a - a_{eq}\right)^2 \tag{1.62}$$

where *k* is the force constant of the harmonic potential, *a* is the coordinate of interest for the current position of the system, and a_{eq} is the equilibrium coordinate about which the potential is centered. When choosing the appropriate biasing potential, a few things need to be considered: first, the force constant should not be too large or too small. If the force constant is too small then the umbrella potential is far too 'wide' and 'shallow', resulting in either the poor sampling of high energy states or unnecessarily long simulation runs to sample an expanded phase space. However, if the force constant is too large, only a small region of interest will be sampled and a large number of simulation runs will be required to guarantee sufficient overlap. All of these concerns result in either a greater amount of error or require a prohibitive amount of computational resources.[116]

A second concern with umbrella sampling is ensuring sufficient sampling of all coordinates orthogonal to the coordinate of interest. As seen up above, umbrella sampling works by determining a probability distribution of coordinates at a given coordinate of interest. However, this requires sufficient sampling of all other coordinates besides the one of interest. For example, at a given protein lipid bilayer separation distance, the center of the protein is

allowed to slide in the *xy* plane (assume the protein lipid distance is along the *z*-axis); in addition to sliding, the protein can also rotate end-over-end. The length of each separate umbrella simulation must be sufficient to not only allow for decent overlap along the coordinate of interest (the *z*-axis in the previous example) but also to allow for slipping and rotating along all orthogonal coordinates (the *xy*-plane in the previous example.) Unfortunately, the timescale of processes such as slipping and rotating must be determined by trial-and-error, previous studies, theory, and/or physical measurements.[116]

The last piece to the umbrella sampling puzzle is the nature of the last term in the potential of mean force, A_u, up above. This term represents a thermodynamic average of the Boltzmann distribution for energy differences in the unbiased probability distribution. More simply put: it is a constant energy offset that is particular to each umbrella sampling run. In order to reconstruct a continuous potential of mean force over all of the separate umbrella sampling windows, the pieces must be 'stitched' together through some method. The most common method for doing this is the Weighted Histogram Analysis Method (WHAM) first developed by Kumar and later modified by Roux and Souaille.[34, 49, 116]

The following summary of the WHAM method closely follows the aforementioned, excellent review article by Kaestner.[115] The primary goal of the WHAM method is to minimize the statistical error between the various probability distributions of different windows. Begin by considering the sum of the probability distributions that are stitched together by an appropriate weighting factor, *w(a)*:

$$P^u(a) = \sum_i^{Windows} w_i(a) P_i^u(a) \tag{1.63}$$

The minimization criteria are such that the derivative of the variance of the probability with respect to each window should equal zero:

$$\frac{\partial \sigma^2(P^u)}{\partial P_i} = 0 \tag{1.64}$$

By normalizing the probability distribution, the two previous equations yield:

$$w_i = \frac{b_i}{\sum_j b_j} \tag{1.65}$$

$$b_i(a) = N_i e^{-\beta \Delta U_i(a) + \beta C_i} \tag{1.66}$$

where N_i is the number of bins for each window *i*.

Recall that:

$$e^{[-\beta C_i]} = \int \left\{ P^u(a) e^{[-\beta \Delta U_i(a)]} \right\} da \tag{1.67}$$

Unfortunately, $P^u(a)$ is not known beforehand, so an iterative method is required between the summation and the integral in order to determine its value. Once $P^u(a)$ has converged, the potential of mean force can be easily calculated from its definition:

$$A_u(a) = -k_B T \ln[P^u(a)] \tag{1.68}$$

For information concerning the practical applications of Umbrella Sampling, please see the following section in his Chapter or Chapters 2, 3, and 5.

1.5. References

1. Deamer, D. W., Origins of life - How leaky were primitive cells? *Nature* **2008,** *454* (7200), 37-38.

2. Adamala, K.; Szostak, J. W., Competition between model protocells driven by an encapsulated catalyst. *Nature Chemistry* **2013,** *5* (6), 495-501.

3. Schmidt, M. L.; Ziani, L.; Boudreau, M.; Davis, J. H., Phase equilibria in DOPC/DPPC: Conversion from gel to subgel in two component mixtures. *The Journal of Chemical Physics* **2009,** *131* (17), 175103.

4. Ow, Y.-L. P.; Green, D. R.; Hao, Z.; Mak, T. W., Cytochrome c: functions beyond respiration. *Nature Reviews Molecular Cell Biology* **2008,** *9*, 532.

5. Marrink, S.; Berendsen, H., Simulation of Water Transport Through a Lipid-Membrane. *Journal of Physical Chemistry* **1994,** *98* (15), 4155-4168.

6. Marrink, S. J.; Berendsen, H. J. C., Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations. *The Journal of Physical Chemistry* **1996,** *100* (41), 16729-16738.

7. Cardenas, A. E.; Jas, G. S.; DeLeon, K. Y.; Hegefeld, W. A.; Kuczera, K.; Elber, R., Unassisted Transport of N-Acetyl-L-tryptophanamide through Membrane: Experiment and Simulation of Kinetics. *The Journal of Physical Chemistry B* **2012,** *116* (9), 2739-2750.

8. Diamond, J. M.; Katz, Y., Interpretation of Nonelectrolyte Partition-Coefficients between Dimyristoyl Lecithin and Water. *Journal of Membrane Biology* **1974,** *17* (2), 121-154.

9. MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P., Distribution of Amino Acids in a Lipid Bilayer from Computer Simulations. *Biophys. J.* **2008,** *94* (9), 3393-3404.

10. Carpenter, Timothy S.; Kirshner, Daniel A.; Lau, Edmond Y.; Wong, Sergio E.; Nilmeier, Jerome P.; Lightstone, Felice C., A Method to Predict Blood-Brain Barrier Permeability of Drug-Like Compounds Using Molecular Dynamics Simulations. *Biophys. J.* **2014,** *107* (3), 630-641.

11. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L., Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology* **2001,** *305* (3), 567-580.

12. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L., Opinion - How many drug targets are there? *Nature Reviews Drug Discovery* **2006,** *5* (12), 993-996.

13. Trapp, M.; Gutberlet, T.; Juranyi, F.; Unruh, T.; Demé, B.; Tehei, M.; Peters, J., Hydration dependent studies of highly aligned multilayer lipid membranes by neutron scattering. *The Journal of Chemical Physics* **2010,** *133* (16), 164505.

14. Edholm, O., Time and length scales in lipid bilayer simulations. In *Computational Modeling of Membrane Bilayers*, Feller, S. E., Ed. Elsevier Academic Press Inc: San Diego, 2008; Vol. 60, pp 91-110.

15. Petrache, H. I.; Tristram-Nagle, S.; Gawrisch, K.; Harries, D.; Parsegian, V. A.; Nagle, J. F., Structure and fluctuations of charged phosphatidylserine bilayers in the absence of salt. *Biophys. J.* **2004,** *86* (3), 1574-1586.

16. Liu, Y.; Nagle, J. F., Diffuse scattering provides material parameters and electron density profiles of biomembranes. *Physical Review E* **2004,** *69* (4).

17. Nevzorov, A. A.; Trouard, T. P.; Brown, M. F., Correlation functions for lipid membrane dynamics obtained from NMR spectroscopy. *Physical Review E* **1997,** *55* (3), 3276.

18. Gruenbaum, S. M.; Skinner, J. L., Vibrational spectroscopy of water in hydrated lipid multi-bilayers. I. Infrared spectra and ultrafast pump-probe observables. *The Journal of Chemical Physics* **2011,** *135* (7), 075101.

19. Gruenbaum, S. M.; Pieniazek, P. A.; Skinner, J. L., Vibrational spectroscopy of water in hydrated lipid multi-bilayers. II. Two-dimensional infrared and peak shift observables within different theoretical approximations. *The Journal of Chemical Physics* **2011,** *135* (16), 164506.

20. Gruenbaum, S. M.; Skinner, J. L., Vibrational spectroscopy of water in hydrated lipid multi-bilayers. III. Water clustering and vibrational energy transfer. *The Journal of Chemical Physics* **2013,** *139* (17), 175103.

21. Kuczera, K.; Unruh, J.; Johnson, C. K.; Jas, G. S., Reorientations of Aromatic Amino Acids and Their Side Chain Models: Anisotropy Measurements and Molecular Dynamics Simulations. *The Journal of Physical Chemistry A* **2010,** *114* (1), 133-142.

22. Loura, L. M. S.; Ramalho, J. P. P., Recent Developments in Molecular Dynamics Simulations of Fluorescent Membrane Probes. *Molecules* **2011,** *16* (7), 5437-5452.

23. Kansy, M.; Senner, F.; Gubernator, K., Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. *Journal of medicinal chemistry* **1998,** *41* (7), 1007-1010.

24. Thompson, M.; Krull, U. J.; Worsfold, P. J., The Analytical Potential of Chemoreception at Bilayer Lipid-Membranes. *Analytica Chimica Acta* **1980,** *117* (Jun), 121-132.

25. Thompson, M.; Krull, U. J.; Worsfold, P. J., The Structure and Electrochemical Properties of a Polymer-Supported Lipid Biosensor. *Analytica Chimica Acta* **1980,** *117* (Jun), 133-145.

26. Thompson, M.; Krull, U. J., The Electroanalytical Response of the Bilayer Lipid-Membrane to Valinomycin - an Empirical-Treatment. *Analytica Chimica Acta* **1982,** *141* (Sep), 49-56.

27. Shah, P.; Jogani, V.; Bagchi, T.; Misra, A., Role of Caco-2 cell monolayers in prediction of intestinal drug absorption. *Biotechnology Progress* **2006,** *22* (1), 186-198.

28. Lee, B. L.; Kuczera, K.; Middaugh, C. R.; Jas, G. S., Permeation of the three aromatic dipeptides through lipid bilayers: Experimental and computational study. *J. Chem. Phys.* **2016,** *144* (24), 14.

29. Orsi, M.; Sanderson, W. E.; Essex, J. W., Permeability of Small Molecules through a Lipid Bilayer: A Multiscale Simulation Study. *The Journal of Physical Chemistry B* **2009,** *113* (35), 12019-12029.

30. Cardenas, A. E.; Elber, R., Computational study of peptide permeation through membrane: searching for hidden slow variables. *Molecular Physics* **2013,** *111* (22-23), 3565-3578.

31. Allen, M. P.; Tildesley, D. J., *Computer simulation of liquids*. Clarendon Press: 1989; p 385.

32. Frenkel, D.; Smit, B., Chapter 4 - Molecular Dynamics Simulations. In *Understanding Molecular Simulation (Second Edition)*, Academic Press: San Diego, 2002; pp 63-107.

33. Lee, B. L.; Kuczera, K., Simulating the free energy of passive membrane permeation for small molecules. *Molecular Simulation* **2017**, 1-11.

34. Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. the Method. *Journal of Computational Chemistry* **1992,** *13* (8), 1011-1021.

35. Souaille, M.; Roux, B., Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer Physics Communications* **2001,** *135* (1), 40-57.

36. Widom, B., Some Topics In Theory Of Fluids. *J. Chem. Phys.* **1963,** *39* (11), 2808-&.

37. Pohorille, A.; Wilson, M. A., Excess chemical potential of small solutes across water-membrane and water-hexane interfaces. *J. Chem. Phys.* **1996,** *104* (10), 3760-3773.

38. Widom, B., Potential-Distribution Theory and the Statistical-Mechanics of Fluids. *Journal of Physical Chemistry* **1982,** *86* (6), 869-872.

39. Jedlovszky, P.; Mezei, M., Calculation of the free energy profile of H2O, O-2, CO, CO2, NO, and CHCl3 in a lipid bilayer with a cavity insertion variant of the Widom method. *Journal of the American Chemical Society* **2000,** *122* (21), 5125-5131.

40. Shinoda, K.; Shinoda, W.; Mikami, M., Efficient free energy calculation of water across lipid membranes. *Journal of Computational Chemistry* **2008,** *29* (12), 1912-1918.

41. Beveridge, D. L.; Dicapua, F. M., Free-Energy via Molecular Simulation - Applications to Chemical and Biomolecular Systems. *Annual Review of Biophysics and Biophysical Chemistry* **1989,** *18*, 431-492.

42. Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M., Hidden Thermodynamics of Mutant Proteins - A Molecular-Dynamics Analysis. *Science* **1989,** *244* (4908), 1069-1072.

43. Kuczera, K.; Gao, J.; Tidor, B.; Karplus, M., Free-Eergy of Sickling - A Simulation Analysis. *Proc. Natl. Acad. Sci. U. S. A.* **1990,** *87* (21), 8481-8485.

44. Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R., Constrained Reaction Coordinate Dynamics For The Simulation Of Rare Events. *Chemical Physics Letters* **1989,** *156* (5), 472-477.

45. Kuczera, K., One- and multidimensional conformational free energy simulations. *Journal of Computational Chemistry* **1996,** *17* (15), 1726-1749.

46. Marrink, S.-J.; Berendsen, H. J. C., Simulation of water transport through a lipid membrane. *The Journal of Physical Chemistry* **1994,** *98* (15), 4155-4168.

47. Kubo, R., Fluctuation-Dissipation Theorem. *Reports on Progress in Physics* **1966,** *29*, 255-&.

48. Torrie, G.; Valleau, J., Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *Journal of Computational Physics* **1977,** *23* (2), 187-199.

49. Roux, B., The calculation of the potential of mean force using computer-simulations. *Computer Physics Communications* **1995,** *91* (1-3), 275-282.

50. Laio, A.; Parrinello, M., Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (20), 12562-12566.

51. Laio, A.; Gervasio, F. L., Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics* **2008,** *71* (12), 22.

52. Sun, R.; Dama, J. F.; Tan, J. S.; Rose, J. P.; Voth, G. A., Transition-Tempered Metadynamics Is a Promising Tool for Studying the Permeation of Drug-like Molecules through Membranes. *Journal of Chemical Theory and Computation* **2016,** *12* (10), 5157-5169.

53. Darve, E.; Pohorille, A., Calculating free energies using average force. *J. Chem. Phys.* **2001,** *115* (20), 9169-9183.

54. Darve, E.; Rodriguez-Gomez, D.; Pohorille, A., Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* **2008,** *128* (14), 13.

55. Wei, C. Y.; Pohorile, A., Permeation of Nucleosides through Lipid Bilayers. *Journal of Physical Chemistry B* **2011,** *115* (13), 3681-3688.

56. Elber, R., A milestoning study of the kinetics of an allosteric transition: Atomically detailed simulations of deoxy Scapharca hemoglobin. *Biophys. J.* **2007,** *92* (9), L85-L87.

57. Majek, P.; Elber, R., Milestoning without a Reaction Coordinate. *Journal of Chemical Theory and Computation* **2010,** *6* (6), 1805-1817.

58. Viswanath, S.; Kreuzer, S. M.; Cardenas, A. E.; Elber, R., Analyzing milestoning networks for molecular kinetics: Definitions, algorithms, and examples. *The Journal of Chemical Physics* **2013,** *139* (17), 174105.

59. Lee, C. T.; Comer, J.; Herndon, C.; Leung, N.; Pavlova, A.; Swift, R. V.; Tung, C.; Rowley, C. N.; Amaro, R. E.; Chipot, C.; Wang, Y.; Gumbart, J. C., Simulation-Based Approaches for Determining

Membrane Permeability of Small Compounds. *Journal of Chemical Information and Modeling* **2016,** *56* (4), 721-733.

60. Comer, J.; Schulten, K.; Chipot, C., Diffusive Models of Membrane Permeation with Explicit Orientational Freedom. *Journal of Chemical Theory and Computation* **2014,** *10* (7), 2710-2718.

61. Cardenas, A. E.; Shrestha, R.; Webb, L. J.; Elber, R., Membrane Permeation of a Peptide: It Is Better to be Positive. *The Journal of Physical Chemistry B* **2015,** *119* (21), 6412-6420.

62. Parisio, G.; Stocchero, M.; Ferrarini, A., Passive Membrane Permeability: Beyond the Standard Solubility-Diffusion Model. *Journal of Chemical Theory and Computation* **2013,** *9* (12), 5236-5246.

63. Comer, J.; Schulten, K.; Chipot, C., Calculation of Lipid-Bilayer Permeabilities Using an Average Force. *Journal of Chemical Theory and Computation* **2014,** *10* (2), 554-564.

64. MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P., Transfer of Arginine into Lipid Bilayers Is Nonadditive. *Biophys. J.* **2011,** *101* (1), 110-117.

65. Wang, Y.; Hu, D.; Wei, D., Transmembrane Permeation Mechanism of Charged Methyl Guanidine. *Journal of Chemical Theory and Computation* **2014,** *10* (4), 1717-1726.

66. Cardenas, A. E.; Elber, R., Modeling kinetics and equilibrium of membranes with fields: Milestoning analysis and implication to permeation. *The Journal of Chemical Physics* **2014,** *141* (5), 054101.

67. Awoonor-Williams, E.; Rowley, C. N., Molecular simulation of nonfacilitated membrane permeation. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2015**.

68. Xiang, T. X.; Anderson, B. D., Liposomal drug transport: A molecular perspective from molecular dynamics simulations in lipid bilayers. *Advanced Drug Delivery Reviews* **2006,** *58* (12-13), 1357-1378.

69. Dickey, A. N.; Faller, R., Molecular Modeling of Biomembranes: A How-to Approach. *Biomedical Applications of Biophysics*, Jue, T., Ed. Springer: Totowa, NJ, 2010.

70. Tieleman, D. P., Chapter 1 Methods and Parameters for Membrane Simulations. In *Molecular Simulations and Biomembranes: From Biophysics to Function*, The Royal Society of Chemistry: 2010; pp 1-25.

71. Orsi, M.; Essex, J. W., Chapter 4 Passive Permeation Across Lipid Bilayers: a Literature Review. In *Molecular Simulations and Biomembranes: From Biophysics to Function*, The Royal Society of Chemistry: 2010; pp 76-90.

72. van der Spoel, D.; van Maaren, P. J.; Berendsen, H. J. C., A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *J. Chem. Phys.* **1998,** *108* (24), 10220-10230.

73. Tamai, Y.; Tanaka, H.; Nakanishi, K., Molecular Simulation of Permeation of Small Penetrants through Membranes .2. Solubilities. *Macromolecules* **1995,** *28* (7), 2544-2554.

74. Sugii, T.; Takagi, S.; Matsumoto, Y., A molecular-dynamics study of lipid bilayers: Effects of the hydrocarbon chain length on permeability. *The Journal of Chemical Physics* **2005,** *123* (18), 184714.

75. Shinoda, W.; Mikami, M.; Baba, T.; Hato, M., Molecular dynamics study on the effects of chain branching on the physical properties of lipid bilayers: 2. Permeability. *The Journal of Physical Chemistry B* **2004,** *108* (26), 9346-9356.

76. Riahi, S.; Rowley, C. N., Why Can Hydrogen Sulfide Permeate Cell Membranes? *Journal of the American Chemical Society* **2014,** *136* (43), 15111-15113.

77. Orsi, M.; Essex, J. W., Permeability of drugs and hormones through a lipid bilayer: insights from dual-resolution molecular dynamics. *Soft Matter* **2010,** *6* (16), 3797.

78. dos Santos, D. J. V. A.; Eriksson, L. A., Permeability of psoralen derivatives in lipid membranes. *Biophys. J.* **2006,** *91* (7), 2464-2474.

79. Bemporad, D.; Luttmann, C.; Essex, J. W., Behaviour of small solutes and large drugs in a lipid bilayer from computer simulations. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2005,** *1718* (1-2), 1-21.

80. Orsi, M.; Haubertin, D. Y.; Sanderson, W. E.; Essex, J. W., A Quantitative Coarse-Grain Model for Lipid Bilayers. *The Journal of Physical Chemistry B* **2008,** *112* (3), 802-815.

81. Liu, Y.; Ichiye, T., Soft Sticky Dipole Potential for Liquid Water: A New Model. *The Journal of Physical Chemistry* **1996,** *100* (7), 2723-2730.

82. Vorobyov, I.; Olson, Timothy E.; Kim, Jung H.; Koeppe, Roger E.; Andersen, Olaf S.; Allen, Toby W., Ion-Induced Defect Permeation of Lipid Membranes. *Biophys. J.* **2014,** *106* (3), 586-597.

83. Herce, H. D.; Garcia, A. E., Cell Penetrating Peptides: How Do They Do It? *Journal of Biological Physics* **2007,** *33* (5-6), 345-356.

84. Porasso, R. D.; Drew Bennett, W. F.; Oliveira-Costa, S. D.; López Cascales, J. J., Study of the Benzocaine Transfer from Aqueous Solution to the Interior of a Biological Membrane. *The Journal of Physical Chemistry B* **2009,** *113* (29), 9988-9994.

85. Fong, C. W., Permeability of the Blood–Brain Barrier: Molecular Mechanism of Transport of Drugs and Physiologically Important Compounds. *The Journal of Membrane Biology* **2015,** *248* (4), 651-669.

86. Nademi, Y.; Iranagh, S. A.; Yousefpour, A.; Mousavi, S. Z.; Modarress, H., Molecular dynamics simulations and free energy profile of Paracetamol in DPPC and DMPC lipid bilayers. *Journal of Chemical Sciences* **2014,** *126* (3), 637-647.

87. Filipe, H. A. L.; Moreno, M. J.; Róg, T.; Vattulainen, I.; Loura, L. M. S., How To Tackle the Issues in Free Energy Simulations of Long Amphiphiles Interacting with Lipid Membranes: Convergence and Local Membrane Deformations. *The Journal of Physical Chemistry B* **2014,** *118* (13), 3572-3581.

88. Khajeh, A.; Modarress, H., Effect of cholesterol on behavior of 5-fluorouracil (5-FU) in a DMPC lipid bilayer, a molecular dynamics study. *Biophysical Chemistry* **2014,** *187-188*, 43-50.

89. Genheden, S.; Essex, J. W., A Simple and Transferable All-Atom/Coarse-Grained Hybrid Model to Study Membrane Processes. *Journal of Chemical Theory and Computation* **2015,** *11* (10), 4749-4759.

90. Pastor, R. W.; MacKerell, A. D., Development of the CHARMM Force Field for Lipids. *The Journal of Physical Chemistry Letters* **2011,** *2* (13), 1526-1532.

91. Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E., Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *Journal of Chemical Theory and Computation* **2010,** *6* (2), 459-466.

92. Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M., Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **1983,** *4* (2), 187-217.

93. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **2009,** *30* (10), 1545-1614.

94. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013,** *29* (7), 845-854.

95. Nitschke, N.; Atkovska, K.; Hub, J. S., Accelerating potential of mean force calculations for lipid membrane permeation: System size, reaction coordinate, solute-solute distance, and cutoffs. *J. Chem. Phys.* **2016,** *145* (12), 9.

96. Neale, C.; Bennett, W. F. D.; Tieleman, D. P.; Pomes, R., Statistical Convergence of Equilibrium Properties in Simulations of Molecular Solutes Embedded in Lipid Bilayers. *Journal of Chemical Theory and Computation* **2011,** *7* (12), 4175-4188.

97. Berger, O.; Edholm, O.; Jahnig, F., Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* **1997,** *72* (5), 2002-2013.

98. Bochicchio, D.; Panizon, E.; Ferrando, R.; Monticelli, L.; Rossi, G., Calculating the free energy of transfer of small solutes into a model lipid membrane: Comparison between metadynamics and umbrella sampling. *The Journal of Chemical Physics* **2015,** *143* (14), 144108.

99. Jämbeck, J. P. M.; Lyubartsev, A. P., Exploring the Free Energy Landscape of Solutes Embedded in Lipid Bilayers. *The Journal of Physical Chemistry Letters* **2013,** *4* (11), 1781-1787.

100. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005,** *26* (16), 1781-1802.

101. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004,** *120* (20), 9665-9678.

102. Henin, J.; Chipot, C., Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.* **2004,** *121* (7), 2904-2914.

103. Elber, R.; Roitberg, A.; Simmerling, C.; Goldstein, R.; Li, H. Y.; Verkhivker, G.; Keasar, C.; Zhang, J.; Ulitsky, A., MOIL - a Program for Simulations of Macromolecules. *Computer Physics Communications* **1995,** *91* (1-3), 159-189.

104. Zhang, C.; Lai, C.-L.; Pettitt, B. M., Accelerating the weighted histogram analysis method by direct inversion in the iterative subspace. *Molecular Simulation* **2016,** *42* (13), 1079-1089.

105. Minoukadeh, K.; Chipot, C.; Lelièvre, T., Potential of Mean Force Calculations: A Multiple-Walker Adaptive Biasing Force Approach. *Journal of Chemical Theory and Computation* **2010,** *6* (4), 1008-1017.

106. Comer, J.; Phillips, J. C.; Schulten, K.; Chipot, C., Multiple-Replica Strategies for Free-Energy Calculations in NAMD: Multiple-Walker Adaptive Biasing Force and Walker Selection Rules. *Journal of Chemical Theory and Computation* **2014,** *10* (12), 5276-5285.

107. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999,** *314* (1-2), 141-151.

108. Dama, J. F.; Rotskoff, G.; Parrinello, M.; Voth, G. A., Transition-Tempered Metadynamics: Robust, Convergent Metadynamics via On-the-Fly Transition Barrier Estimation. *Journal of Chemical Theory and Computation* **2014,** *10* (9), 3626-3633.

109. Poger, D.; Van Gunsteren, W. F.; Mark, A. E., A New Force Field for Simulating Phosphatidylcholine Bilayers. *Journal of Computational Chemistry* **2010,** *31* (6), 1117-1125.

110. Poger, D.; Mark, A. E., On the Validation of Molecular Dynamics Simulations of Saturated and cis-Monounsaturated Phosphatidylcholine Lipid Bilayers: A Comparison with Experiment. *Journal of Chemical Theory and Computation* **2010,** *6* (1), 325-336.

111. Cardenas, A. E.; Elber, R., Markovian and Non-Markovian Modeling of Membrane Dynamics with Milestoning. *Journal of Physical Chemistry B* **2016,** *120* (33), 8208-8216.

112. Chipot, C.; Comer, J., Subdiffusion in Membrane Permeation of Small Molecules. *Scientific Reports* **2016,** *6*, 14.

113. Torrie, G. M.; Valleau, J. P., Monte-Carlo Study of a Phase-Separating Liquid-Mixture by Umbrella Sampling. *J. Chem. Phys.* **1977,** *66* (4), 1402-1408.

114. Bennett, C. H., Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. *Journal of Computational Physics* **1976,** *22* (2), 245-268.

115. Kästner, J., Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011,** *1* (6), 932-942.

116. Frenkel, D.; Smit, B., Chapter 7 - Free Energy Calculations. In *Understanding Molecular Simulation (Second Edition)*, Academic Press: San Diego, 2002; pp 167-200.

Chapter 2

Permeation of the Three Aromatic Dipeptides through Lipid Bilayers

Note: parts of this work have been reproduced and adapted from the following article with permission from AIP publishing.[1]

2.1. Overview:

This chapter explores the passive permeation of the three aromatic dipeptides through a lipid bilayer by both computational and experimental methods. The time-resolved parallel artificial membrane permeability assay with fluorescence detection and comprehensive computer simulations were used to study the passive permeation of three aromatic dipeptides: N-acetyl-phenylalanineamide (NAFA), N-acetyltyrosineamide (NAYA), and N-acetyl-tryptophanamide (NATA), through a 1,2-dioleoyl-*sn*-glycero-3-phospocholine (DOPC) lipid bilayer. Computationally, umbrella sampling simulations were performed to model the structure, dynamics, and interactions of the peptides as a function of the distance from the center of the lipid bilayer, denoted by the variable '*z*'. The calculated profiles of the potential of mean force show two primary effects: preferential binding of each of the three peptides to the lipid interface and large free energy barriers in the center of the membrane. We then used several approaches to calculate the position-dependent translational diffusion coefficients, *D(z)*, including one based on a numerical solution the Smoluchowski equation. Surprisingly, computed *D(z)* values change very little with reaction coordinate and are also quite similar for the three peptides studied. In contrast, calculated values of sidechain rotational correlation times, $\tau_{rot}(z)$, show extremely large changes with the insertion of the peptide into the membrane. Values become 100 times larger in the headgroup region and 10 times larger at the interface and in the membrane center, relative to the solution. The peptides' conformational freedom becomes systematically more restricted as they enter the membrane, sampling α and β and C_{7eq} conformations in the solution, α and C_{7eq} at the interface, and only C_{7eq} in the center. Residual waters of solvation remain around the peptides even in the center of the membrane.

Overall, this study provides an improved microscopic understanding of passive peptide permeation through membranes, especially with respect to rotational diffusion and the distance of the dipeptides from the center of the model membrane.

2.2. Introduction:

Biological membranes form the basis of all multicellular life. They regulate the intracellular and extracellular environment by serving as the gatekeepers for the passage of molecules through passive diffusion, facilitated diffusion, and active transport. Modern evolutionary theory predicts that the first cellular transport machinery developed from the passive diffusion of proteins across a simple lipid bilayer.[2-3] Roughly 30% of all proteins encoded by the human genome are membrane proteins,[4] and 70% of all modern pharmaceutical targets aim to influence and regulate membrane proteins.[5] Even drugs that do not interact with cell membranes must pass through this barrier in order to reach their intracellular targets.[6-7] As a result, a fundamental knowledge of the passive diffusion of small peptides is essential towards understanding all of these processes on both a theoretical and practical level. Rather than taking a macromolecular approach, we focus on the basic building blocks of these larger constructs – amino acids.

Unfortunately, little is known about the passive diffusion of many amino acids, which serve as the building blocks for proteins. Experimental methods that require researchers to use planar bilayer and liposome systems with detection through a wide range of approaches have been used extensively to study the interfacial regions of membranes[8-10] but tend to lack

atomistic detail of the processes involved within the membrane. The parallel artificial membrane permeation assay method uses polycarbonate microporous supports with a single lipid bilayer per pore and is one of the most commonly used experimental methods.[11-14] The permeation rates of samples also correspond to the Caco-2 model in intestinal absorption and clinical studies of the blood-brain barrier. The components of the lipid bilayer can also be separated and examined. This allows for the measurement of the translocation of an analyte through the constructed lipid bilayer.

As computational power and empirical force field parameterization have improved, atomistic level molecular dynamics (MD) simulations have gained greater prominence towards studying the diffusion of small molecules through a membrane.[15] Recent molecular dynamics studies have focused on a wide variety of molecules passively diffusing through membranes, such as: water,[16-18] small molecules,[19-24] model drug compounds,[7, 23, 25-26] analgesics,[27-29] drug delivery systems,[30-31] dyes,[32-33] other lipids,[34-35] nanoparticles,[36-38] toxins,[39] small peptides,[40-41] and even transmembrane proteins.[42-43] However, only a handful of MD studies have examined amino acid-related systems and are confined to tryptophan,[44-46] arginine,[47-48] lysine,[48] and amino acid analogues.[49] In terms of the potential of mean force (PMF), findings have consistently shown that small nonpolar molecules tend to be preferentially bound in the membrane center, while polar molecules tend to interact favorably with the lipid headgroups and experience a free energy barrier in the center. The PMFs for the sidechains of tryptophan, tyrosine, and phenylalanine determined by MacCallum *et al.* mostly fit this general picture, with tryptophan and tyrosine exhibiting PMF minima at the water-lipid interface. Both phenylalanine and tyrosine have lower free energies inside the lipid than in water, while tryptophan has to

overcome a significant free energy barrier in the membrane center. In studies of blocked tryptophan (N-acetyltryptophanamide, or NATA),[44] a similar behavior was found as for the tryptophan sidechain, with NATA exhibiting a significantly higher barrier in the central region.

The primary objective of this study is to expand the understanding of the permeation of small amino acids through lipid membranes. The results of computer simulations for blocked forms of the three aromatic dipeptides: NATA, N-acetyltyrosineamide (NAYA), and N-acetylphenylalanineamide (NAFA) in 1,2-dioleoyl-*sn*-glycero-3-phospocholine (DOPC) lipid bilayers (Figure 2.1) are presented. Computationally, molecular dynamics simulations with umbrella sampling were used to explore the structures, motions, and interactions of the systems as a function of the distance from the membrane center, including new approaches to calculating the position-dependent diffusion coefficients *D(z)* and an analysis of rotational diffusion. Overall, this computational study provides a new level of understanding of the mechanism of the passive permeation of aromatic peptides through lipid bilayers.

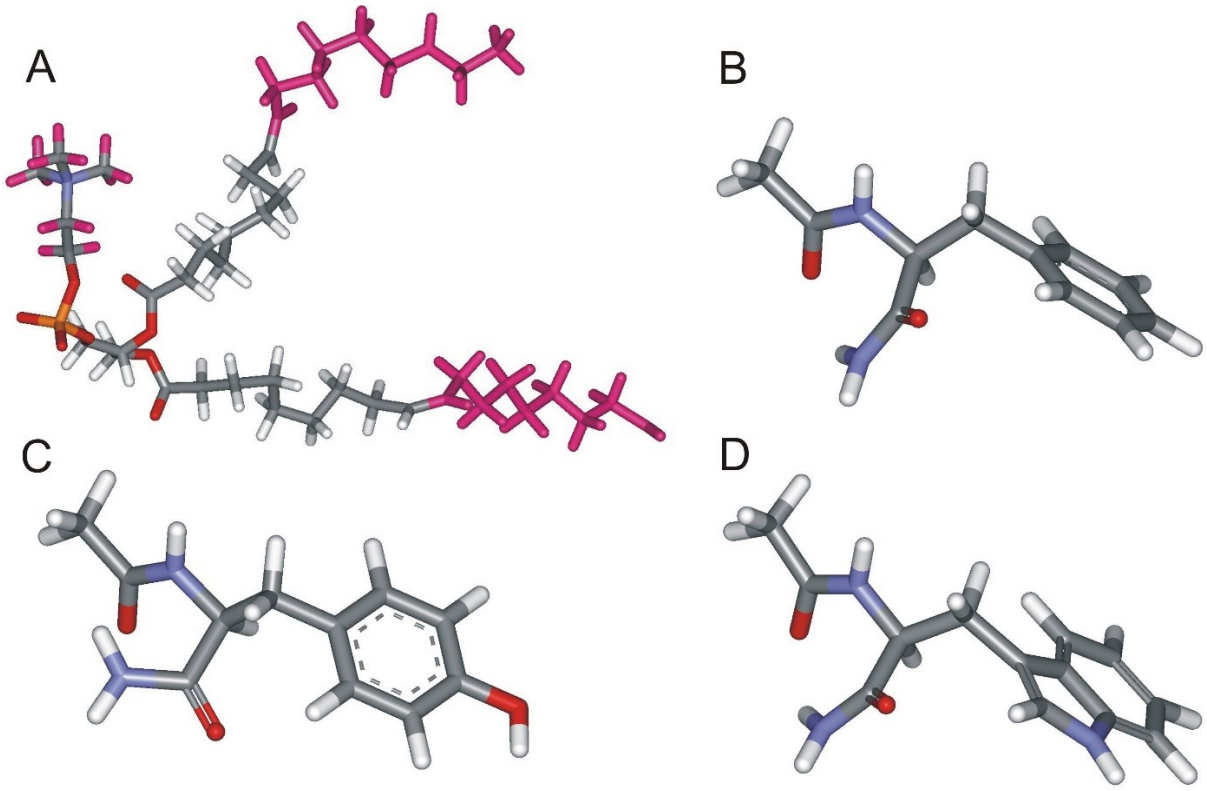


Figure 2.1. Chemical structures of studied systems. (A) 1,2-dioleoyl-sn-glycero-3-phospocholine (DOPC) (B) N-acetylphenylalanineamide (Ac-Phe-NH_2 or NAFA) (C) N-acetyltyrosineamide (Ac-Tyr-NH_2 or NAYA) (D) N-acetyltryptophanamide (Ac-Trp-NH_2 or NATA).

2.3. Methods:

The simulated peptides were N-acetyltryptophanamide (Ac-Trp-NH_2 or NATA), N-acetyltyrosineamide (Ac-Tyr-NH_2 or NAYA), and N-acetylphenylalanineamide (Ac-Phe-NH_2 or NAFA) (Fig. 1). Initial peptide structures were built with CHARMM[50-51] in extended conformations. The phospholipid bilayers with a single copy of peptide in aqueous phase were created by using CHARMM-GUI[52-54] and all molecular dynamics simulations were conducted with GROMACS 4.5.4 or 4.5.6.[55] The bilayer systems contained the following molecules: 50 (2×25) 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) molecules, one peptide, eight chloride, and eight sodium ions and TIP3P water. Sodium and chloride ions were added to maintain a

physiologically relevant ionic strength. Due to small variations in system size, the phenylalanine, tyrosine and tryptophan simulations contained 2939, 2949, and 2599 TIP3P water molecules, respectively, in tetragonal boxes with dimensions of 4.29 nm × 2.39 nm × 8.30 nm, 4.22 nm × 4.22 nm × 8.56 nm, and 4.19 nm × 4.19nm x 8.09 nm, respectively. These boxes produce DOPC headgroup areas of 0.7355 nm^2, 0.7122 nm^2, and 0.7017 nm^2, respectively, in good agreement with experimental averages of 0.723 nm^2.[56-57] The electron density profile of these membranes are also in agreement with experimental profiles (see Supplementary Information).[57] In our coordinate system, the *x* and *y* axes are in the plane of the membrane, and *z* is perpendicular to the plane of the membrane. DOPC and peptide molecular interactions were represented by the CHARMM v.36 force field[58-59] and water was described by the TIP3P model.[60]

Phenylalanine and tyrosine simulations were performed using GROMACS 4.5.4, and all tryptophan simulations were performed with GROMACS 4.5.6.[55] Periodic boundary conditions were used along all three coordinate axes. Direct electrostatic interactions were cut off at 0.13 nm, with long range effects calculated by using the particle mesh Ewald method with a mesh spacing of 0.12 nm. Van der Waals interactions were truncated at 1.2 nm and smoothed with a switching function between 1.0 and 1.2 nm. Newton's equations of motion were integrated by using the leap-frog algorithm[55, 61] with a time step of 2 fs and constraints on all bonds using the LINCS algorithm.[62] Temperature control was achieved by using velocity rescaling with an added stochastic term.[63] GROMACS implements a recent velocity rescaling algorithm by Bussi *et. al.* that improves upon the ergodicity of a traditional Nosé-Hoover thermostat while sampling a canonical ensemble. The position of the permeant was recorded every 0.1 ps, the permeant

pulling force every 0.1 ps, and complete structural information for the entire system every 1.0 ps. Initial velocities were determined from a Maxwell distribution at 300K. Simulations were first run with an NPT ensemble until pressure was equilibrated at one atmosphere of pressure. Stress profiles and pressure values were not monitored. All subsequent simulations were then run with an NVT ensemble. The initial simulation image was equilibrated over 500 ps intervals with increasingly more stringent restraints. An unrestrained molecular dynamics simulation was then run for 120 ns. A system image was then extracted where the peptide permeant was located 1.6 nm from the lipid bilayer center and was then used as the starting point for successive umbrella sampling windows.

Successive umbrella sampling windows were conducted at distances from 0 to 3.0 nm away from center of the lipid bilayer and along the z-axis. The phenylalanine and tryptophan windows were spaced in even 0.1 nm increments for a total of 31 different positions. Windows for tyrosine were run at 0.0 nm; from 0.09 to 0.9 nm in 0.09 nm increments; and then from 1.0 to 3.0 nm in 0.1 nm increments. The tighter increment spacing was used to improve statistical sampling and overlap between simulation windows. For phenylalanine and tryptophan, this was accomplished by gradually increasing the simulation length from 50 ns to 100 ns as the peptide neared the center of the lipid bilayer. All simulation windows were run for at least 50 ns. A restraining umbrella sampling potential of 3000 $kJmol^{-1}nm^{-1}$ was applied to peptide center of mass, with minimum at the center of each window.[64-65] The weighted histogram analysis method (WHAM) was then applied to the resulting data to obtain the potential of mean force (PMF).[66-68]

Position dependent translation diffusion coefficients *D(z*) were determined using three approaches – one based on the Smoluchowski equation and two on the Fluctuation-Dissipation theorem. The first approach uses a numerical solution to the Smoluchowski equation as described by Bicout and Szabo[69] and as discussed by Hummer.[70] Histograms were created from the center-of-mass distance between the peptide and the membrane *z(t)* within each umbrella sampling window and were used to determine the biased probability distribution, *p*(z)= p*(n)*, where *n* is the bin number. The transition rates between neighboring bins, $w_{n+1,n}$ were calculated from the number of transitions in *z(t)* and the bin residence times. Diffusion coefficients were then calculated as:

$$D_{n+1/2} = w_{n+1,n}\left[\frac{p^*(n)}{p^*(n+1)}\right]^{1/2} d^2 \qquad (2.1)$$

with *d* being the bin width (*d* = 0.02 nm was used). These coefficients correspond to motion within the biased potential, including the harmonic umbrella sampling restraint. However, from each simulation, we only took the value of *D(z)* at the center of the window, where the constraint potential is approximately 0. Standard errors were estimated by performing separate calculations over four quarters of the data and by multiplying the standard error of the mean by the appropriate Student's t-value at a 95% confidence level (3.182).

The two other methods of calculating *D(z)* were based on the Fluctuation-Dissipation Theorem:[71]

$$D(z) = \frac{(k_B T)^2}{\int_0^\infty \langle \Delta F_z(z,0) * \Delta F_z(z,t)\rangle dt} \qquad (2.2)$$

where k_B is the Boltzmann constant, *T* is the temperature, and $\Delta F_z(z, t)$ is the deviation of the *z*-component of the force experienced on the center of mass of the permeant from its average

value, as a function of the permeant depth, *z*, and time, *t*. These two approaches differed in terms of the forces used. In the first, forces corresponding to fixed values of *z* were calculated, using a custom modified version of CHARMM v.38, designed to enable fixing the difference in the center-of-mass *z*-coordinates between two sub-systems. For each umbrella sampling window, ten independent molecular dynamics trajectories were generated with the same CHARMM v. 36 protein and lipid parameters as were used in the GROMACS simulations.[51] The starting images were extracted from the umbrella sampling trajectories, and the CHARMM simulations were then run for 100 ps with a time step of 2 fs. Force data was recorded every 0.1 ps. Temperature was kept constant at 300K by using a Nosé-Hoover thermostat which is the constant temperature method available in CHARMM.[72-73] Force autocorrelation functions were then numerically integrated until temporal convergence was achieved, typically after 15 ps. This approach to *D(z)* estimation is analogous to that employed in the original constrained-*z* simulations of Marrink and Berendsen.[16] The final method for obtaining *D(z)* used autocorrelations of the umbrella restraining force, recorded for each US window. In this case, numerical integration typically converged after 1000 ps. The *D(z)* values from the Smoluchowski equation and the fixed-*z* forces method agreed throughout the simulation range, within the calculated confidence intervals. Values obtained in the aqueous phase with these methods also agreed with a separate molecular dynamics simulation of NATA in a TIP3P box (see Results and Discussion). The translational diffusion coefficients obtained from fluctuations of the restraint force were typically 5-6 times lower than those of the first two methods and did not agree with the free NATA MD in water (see Supplementary Information). Therefore, the last method of *D(z)* calculation was not used in further analysis.

The inhomogeneous solubility diffusion model was then used to calculate the permeability coefficient, *P*, and the mean first passage time, $\langle\tau\rangle$:[16, 20, 44, 74]

$$P = \left[\int_a^b \frac{e^{\beta w(z)}}{D(z)} dz\right]^{-1} \tag{2.3}$$

where *w(z)* is the potential of mean force at location *z*, *a* is the *z* location of the free energy minimum along the membrane interface, *b* is the opposite side of the membrane, and $\beta = (k_B T)^{-1}$. The mean free passage time, $\langle\tau\rangle$, can then be determined as:[75-76]

$$\langle\tau\rangle = \int_a^b \left\{\frac{e^{\beta\Delta G(z)}}{D(z)} \int_a^z e^{-\beta\Delta G(z')} dz'\right\} dz \tag{2.4}$$

Rotational motion was studied by following reorientations of two molecular axes for each peptide. For sidechains, the axes were related to the electronic transition dipoles: the in-plane axes perpendicular to the CG-CZ vector for NAFA and NAYA and the ^1L_b transition dipole axis for NATA.[77] For the overall reorientation, the axis was the vector connecting the center of mass of the backbone to the center of mass of the sidechain. For each case, the autocorrelation function $C_2(t) = \frac{1}{2}\langle(3\cos^2(\theta) - 1)\rangle$ was calculated, with θ being the angle of axis reorientation during time *t*. The rotational correlation time τ_{rot} was calculated as the integral of $C_2(t)$ over a time range where the function decays to zero, and the integral reaches a stable value. In the headgroup region, in several cases, the autocorrelation functions did not converge to zero, indicating that the sidechain reorientations were not completely sampled during the simulation period. As a result, the longest calculated correlation times are highly approximate, and this can be seen from the greater, calculated confidence intervals.

2.4. Results and Discussion

2.4.1. - Potentials of mean force.

The potential of mean force represents the relative free energies of a given permeant molecule at different *z*-distances from membrane center. As seen in Figure 2.2, all three peptides have qualitatively similar PMFs, exhibiting free energy minima at the lipid-water interface and maxima at the membrane center. The interfacial free energy minima are, respectively, for NAFA, NAYA, and NATA: -18, -14, and -12 $kJmol^{-1}$. The central free energy barriers are, respectively, for NAFA, NAYA, and NATA: +28, +41, and +44 $kJmol^{-1}$, relative to the respective minima (and +10, +27, and +32 $kJmol^{-1}$ relative to the solution, respectively). The peptide with the most hydrophobic sidechain, NAFA, exhibits the strongest preference for the interface and the lowest central free energy barrier. NAYA and NATA, with partly hydrophilic sidechains, exhibit both weaker binding to the interface and higher barriers in the membrane center. The permeant with the largest sidechain, NATA, has the weakest preference for the interface and the highest for barrier. As discussed further in the structural analysis section, the interfacial minima result from a tug of war between hydrophobic and hydrophilic interactions. The most hydrophobic system, NAFA, has the deepest minimum of -18 $kJmol^{-1}$ and is located farthest from the membrane center at 1.15 nm. This minimum is also quite broad. For the more polar NAYA and NATA, the minima are shallower (14 and -11 $kJmol^{-1}$, respectively) and are located closer to the membrane (at 1.54 and 1.45 nm, respectively).

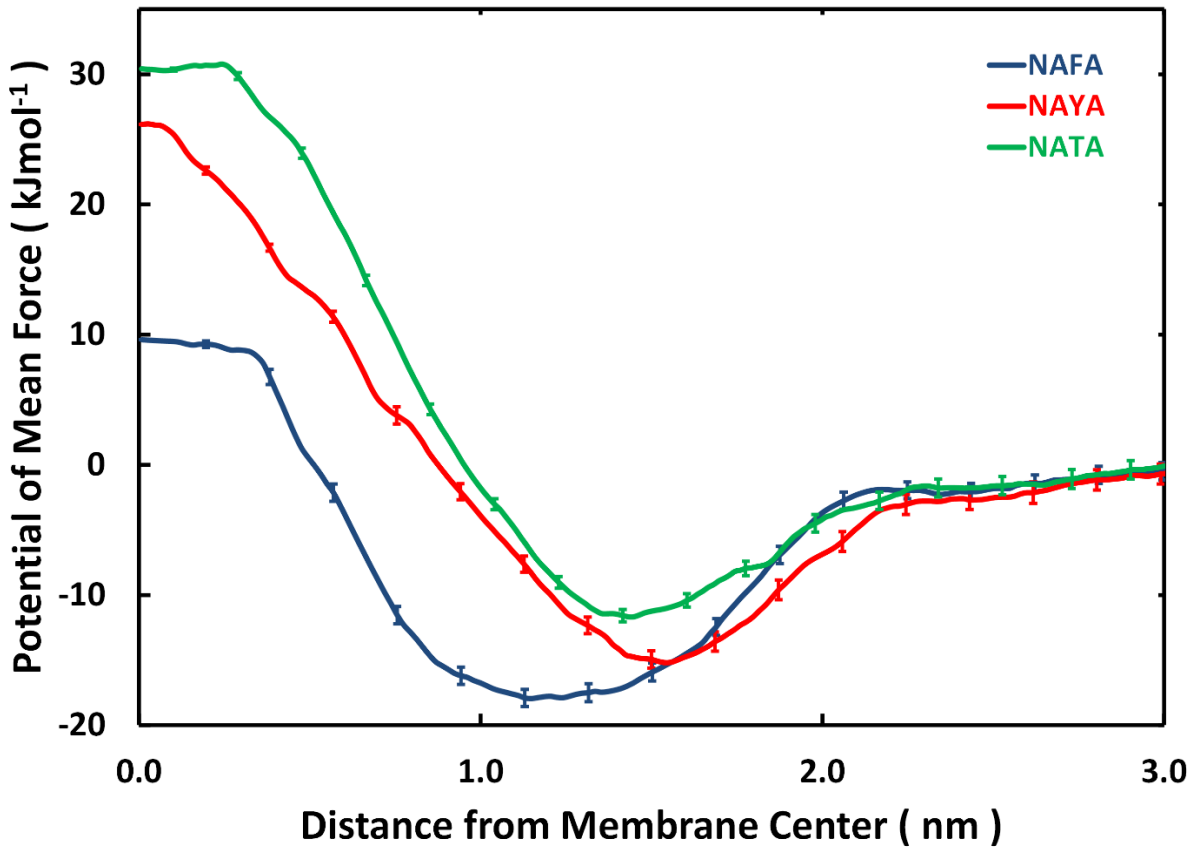


Figure 2.2. The potential of mean force is plotted for all three blocked amino acids. The free energy value was set to zero in the solvent region for each data set. The center of the lipid bilayer is located at a z = 0 nm. Error bars represent the standard error as calculated by the Bootstrap method.[78] The standard error was much larger when the calculations were conducted on four contiguous bins of the data: between 3 and 10 kJmol^{-1} for the central membrane barrier and between 1 and 4 kJmol^{-1} for the interfacial region (see Supplementary Information).

Previous simulations of amino acid sidechains in DOPC by MacCallum *et al*. predicted interfacial free energies of -13 kJmol^{-1} for phenylalanine and tyrosine, and -22 kJmol^{-1} for tryptophan; in the center of the membrane, negative free energies of -5 and -13 kJmol^{-1} were respectively predicted for phenylalanine and tryptophan, with a smaller barrier of 7 kJmol^{-1} for tyrosine. Except for tyrosine, these are qualitatively different from our results, due to the

presence of backbone residues in our systems.[49] Interestingly, our dipeptide PMFs are qualitatively similar to the MacCallum *et al*. results for polar sidechains – especially asparagine and glutamine.[49] Cardenas *et al.* have studied the permeation of NATA through a DOPC bilayer and found an interfacial minimum of about -24 $kJmol^{-1}$ relative to the solution and a barrier of about 75 $kJmol^{-1}$ relative to the minimum. The quantitative differences from our results are most likely due to their use of the Berger lipid and OPLS/AA protein force fields, as well as a slightly smaller number of DOPC lipids.[44, 49]

2.4.2. - Translational Diffusion.

Diffusion constants were calculated using three different approaches – using the numerical solution of the Smoluchowski equation[69-70] and autocorreration functions of force fluctuations from separate short trajectories with constrained *z* and from force fluctuations of the umrella restraint force. These methods are described in more detail within Section 2.3. The first two methods gave consistent results. For a 100 ns test simulation of NATA in a TIP3P water box with CHARMM36 parameters, the calculated one-dimensional NATA center-of-mass diffusion constant was $0.35 \pm 0.04 \times 10^{-9}$ m^2s^{-1}, which is in good agreement with the umbrella sampling results at 3.0 nm of $0.39 \pm 0.01 \times 10^{-9}$ m^2s^{-1} from the Smoluchowski equation and $0.26 \pm 0.11 \times 10^{-9}$ m^2s^{-1} from constrained MD at 2.0 nm. The translational diffusion estimates based on the umbrella constraint force fluctuations were roughly six times lower than from the first two approaches, and were not used in further analyses. The *D(z)* values based on the Smoluchowski equation are presented in Figure 2.3. Values obtained from constrained (fixed *z*)

simulations are given in the Supplementary Information. The diffusion coefficients demonstrate generally little variation with distance from membrane. For NAFA, *D(z)* changes from 0.44×10^{-9} m^2s^{-1} in the solvent region, to 0.40×10^{-9} m^2s^{-1} at the interface, and to 0.45×10^{-9} m^2s^{-1} at the membrane center (except for one outlier of 0.35×10^{-9} m^2s^{-1} at z = 1.8 nm). Similar effects are seen for NAYA: *D(z)* = 0.41×10^{-9} m^2s^{-1} in the solvent, 0.37×10^{-9} m^2s^{-1} at the interface, and 0.44×10^{-9} m^2s^{-1} in the center; and, for NATA: 0.39×10^{-9} m^2s^{-1} in the solvent, 0.35×10^{-9} m^2s^{-1} at the interface, and 0.43×10^{-9} m^2s^{-1} in the center. There is a trend for slower translational diffusion at the interface and for faster diffusion in the center of the lipid bilayer; however, it is weak and barely rises above the statistical uncertainties. There is also a weak systematic trend for diffusion rates between the three peptides, with NAFA > NAYA > NATA, consistent with increasing size. In general, the translational diffusion of the three dipeptides is quite similar, even though their sidechains differ significantly in terms of properties. Why does *D(z)* vary so little with respect to membrane insertion, in view of the strong preferential binding of the permeants at the interface and the well-documented lower molecular density in the center of the membrane? Few membrane translocation studies report their diffusion constant data. Several studies for small molecules do report increased diffusion rates at the center of the lipid bilayer.[16, 19-22] In contrast, most studies of larger molecules that are at least the size of the amino acids studied in this work show relatively flat diffusion profiles in homogenous lipid bilayers.[26, 28, 31] Therefore, our translational diffusion results are consistent with other reported values for larger molecules.

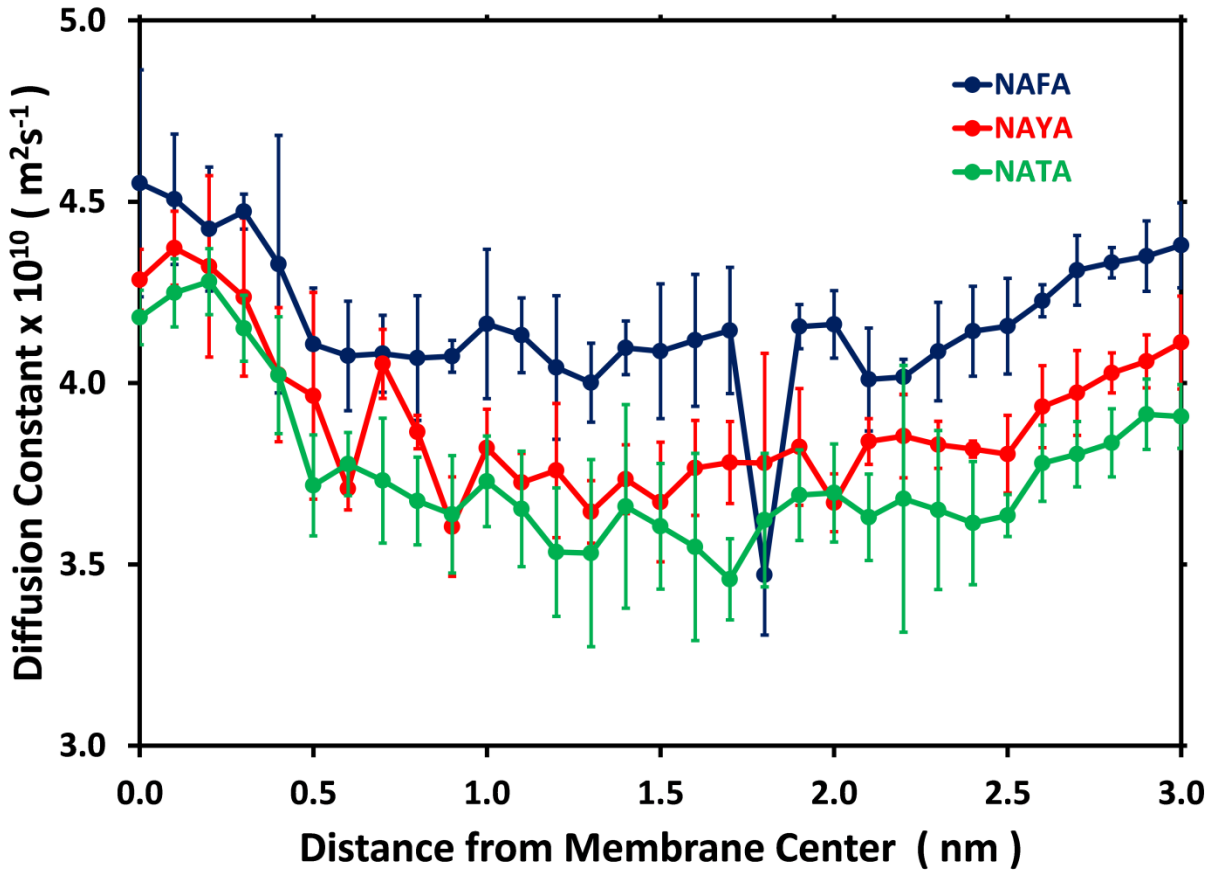


Figure 2.3. Translational diffusion constants are plotted as a function of the permeant distance from the center of the lipid bilayer, calculated by numerical solution of the Smoluchowski equation. Error bars were obtained from separate calculations by dividing the data into four contiguous bins.

2.4.3. - **Permeability Measures**.

By combining the diffusion coefficients with the PMF profile, quantitative measures of the permeation time scale and rate may be obtained – the mean first passage time (MFPT) and the permeation coefficient, *P* (see Methods in section 2.3). Results are reported in Table 2.1. Our simulations predict that NAFA passes through the membrane on a microsecond time scale, whereas NAYA and NATA pass through on a millisecond time scale. The corresponding permeation coefficients are 1×10^{-4}, 2×10^{-6}, and 3×10^{-7} cms^{-1} for NAFA, NAYA, and NATA,

respectively. The results for tryptophan are quite similar to those for tyrosine, which is to be expected considering their similar diffusion and potential of mean force profiles. Due to the relatively flat diffusion coefficient profiles and the similarity of the *D(z)* values for the three peptides, the differences in passive diffusion rates are mostly determined by the free energy profiles.[40]

Table 2.1. The permeation coefficients and mean passage time for NAFA, NAYA and NATA. The averages correspond to integration over PMF and *D(z)* over the second halves of the umbrella sampling trajectories, while the ranges correspond to results from dividing the data into contiguous quarters. Experimental data was obtained courtesy of Dr. Gouri Jas.[1]

	Permeation coefficient, P (cm s^{-1})			Mean passage time, τ (μs)	
		Computational		Computational	
Molecule	Experimental	Average	Range	Average	Range
NAFA	$(56 \pm 5) \times 10^{-7}$	1×10^{-4}	$2 \times 10^{-5} - 2 \times 10^{-3}$	50	30-2 600
NAYA	$(6.2 \pm 1.1) \times 10^{-7}$	2×10^{-6}	$1 \times 10^{-6} - 1 \times 10^{-5}$	3 000	600-3 000
NATA	$(26 \pm 2) \times 10^{-7}$	3×10^{-7}	$1 \times 10^{-7} - 2 \times 10^{-6}$	15 000	3000-30 000

The calculated passage times are systematically smaller, and the permeability coefficients are systematically higher than our experimentally measured values for all three peptides. The calculated permeation times differ by many orders of magnitude. Qualitatively, the simulations correctly predict that the permeation of NAFA should be the fastest of the three systems. Quantitatively, the calculated P values for NAYA and NATA of 3×10^{-6} and 2×10^{-6} cms^{-1}, respectively, are comparable to the corresponding experimentally determined results of 6×10^{-7} and 26×10^{-7} cms^{-1}, respectively. However, this order is reversed in the simulations. Although a direct comparison does not exist for phenylalanine and tyrosine, our

results are roughly the same order of magnitude as those obtained using the same model for small molecules.[26] Previous studies of tryptophan, using the more advanced method of milestoning, report permeation times on the time scale of hours, which are also in agreement with experimental results.[44-45, 79-80] Most other researchers have also reported much larger and faster permeability coefficients using the inhomogeneous solubility-diffusion model, which has been discussed in great detail in other works.[19, 26, 44-45, 81-83] Several assumptions underpin the solubility-diffusion model, including memoryless, diffusive-type motion along the reaction coordinate and the presence of only one slow variable describing the motion of the permeant — namely, the translocation along the membrane normal.[83-84] Many recent studies suggest that an additional rotational barrier exists that slows down the movement of the permeant.[19, 44-45, 81, 83, 85] Others hypothesize that membrane and solvent structural fluctuations play an important role as well.[40, 79] Use of the CHARMM lipid force field also overestimates the electric field strength within the membrane by a factor of 3, which may be lowering the value of the PMF within the membrane interior and subsequently accelerating the mean passage time.[58]

The discrepancy between our experimental and theoretical results can be attributed to some of the above concerns. Additional slow variables such as rotational barriers or membrane fluctuations can couple with the longitudinal translocation of the permeant and slow down the process. For example, Figure 2.5B demonstrates that NAYA adopts preferential rotational orientations based upon its depth within the membrane. The correlation of these motions could not be calculated within the time frame of each umbrella window. These motions may be essential for the longitudinal diffusion of the permeant, but they are not included within our permeability model. This phenomenon has been well documented in assisted diffusion and

passive diffusion through channels formed by porins.[86] For NATA, rotational barriers have also been found to reduce the permeability coefficient through more advanced models such as milestoning.[79, 83] Similarly, membrane fluctuations within the lipid bilayer also play a role in the permeation process. For small molecules, these fluctuations are very fast, with timescales less than a nanosecond.[87] For larger molecules, membrane effects become increasingly more important. As Neale *et al.* discovered, there are rare sampling barriers at the lipid interface with timescales on the order of 10 μs — far too long for our simulations to detect.[88] Whether our amino acids fall into the small or large category remains unknown and would require much longer simulation times. Finally, the artificially high electric field strength within the simulation may be assisting our amino acids to adopt favorable orientations on a much faster timescale due to the increased forces on the molecular dipole. All of these concerns could cause our theoretical permeability coefficients to be far larger than those determined experimentally.

2.4.4. - Rotational Diffusion.

Rotational correlation times τ_{rot} for the peptide sidechains are presented in Figure 2.4. Unlike the translational diffusion rates, rotational diffusion speeds change very strongly upon membrane insertion. Rotations are fastest in the aqueous region, with average sidechain τ_{rot} values of 17, 33, and 27 ps for NAFA, NAYA and NATA, respectively. Reorientational motions slow down dramatically as the peptide approaches the lipid-water inteface, with average τ_{rot} values of over 0.3 to 1.5 ns for all three peptides at z = 2 nm. The reorientation rates are even slower in the headgroup region, with τ_{rot} in the 1 to 4 ns range for z = 0.5 - 1.3 nm. Motions in

the bilayer center occur at rates intermediate between water and headgrups, with τ_{rot} = 300 ps, 800 and 200 ps for NAFA, NAYA and NATA, respectively. If local viscosity changes were the dominant effect, we would expect a similar variation of translation diffusion, *D(z)*, and rotational correlation, $\tau_{rot}(z)$, with membrane insertion. However, because the translational diffusion profile is quite flat, the variations of τ_{rot} by factors 10 to 100 must result from very strong specific interactions with the lipid environment. Clear evidence for strong preferential binding of the peptides with the interface is seen in Figure 2.2, which depicts the potential of mean force. What is unexpected is the slowing down of reorientations by as much as a factor of 100 in the headgroup region and by 10 in the membrance center, compared to solution. Previous studies, which focused on translational dynamics, typically found faster motions in the central region of lower relative density. [16, 19-22] The presence of slow reorientations in the course of membrane transport has been previously explored.[26, 28, 31]

Our simulations suggest that translational diffusion in the *z* direction, along the elongated lipid molecules, is relatively easy; while rotational diffusion of the peptide sidechains, which are tethered to their backbones and must move in directions perpendicular to the tightly packed lipids, is unexpectedly slow. Our results present a new look at peptide dynamics in membranes, indicating that computational modeling and experimental measurement of reorientational motions should be a sensitive probe of peptide-membrane interactions. A more detailed analysis of peptide-lipid interactions is given below (see Specific interactions).

Computer simulations of NAFA, NAYA and NATA reorientations using the CHARMM force field and TIP3P water have reported values of of τ_{rot} = 20, 27, and 30 ps, respectively, at

298 K.[77] Experimental values previously reported were 40 ± 5 ps for NAYA, 48 ± 5 ps for NATA at 298 K, and 98 ± 30 ps for NAFA at 278 K.[77] Values for τ_{rot} within the aqueous region are in very good agreement with the previous calculations and are also in good agreement with the experimental estimates at 298 K, given that the TIP3P water model systematically underestimates the viscosity of water.[60]

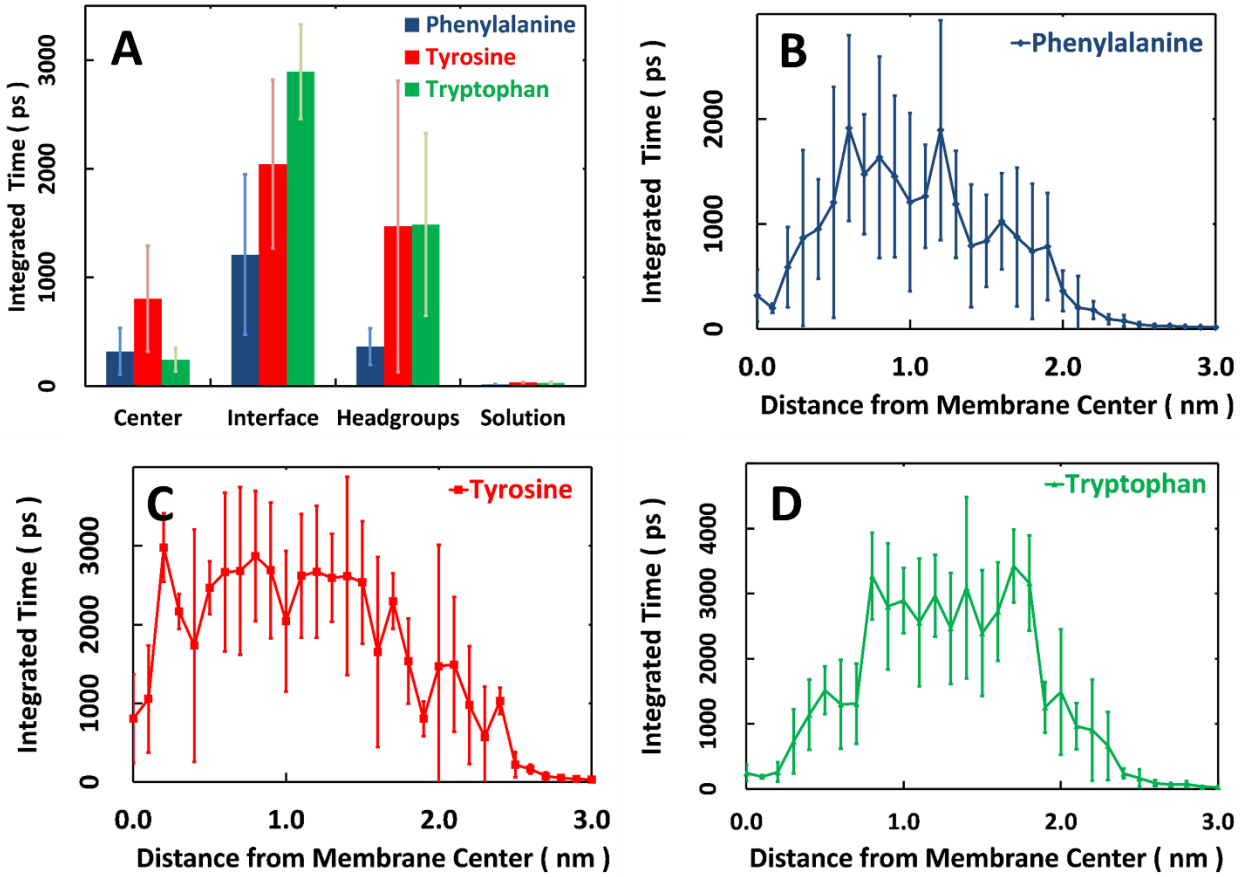


Figure 2.4. Sidechain rotational correlation times as a function of the membrane insertion distance, *z*. (A) Comparison of τ_{rot} values at *z* = 0.0 nm, *z* = 1.5 nm, and *z* = 3.0 nm for the three peptides and values of $\tau_{rot}(z)$ for NAFA (B), NAYA (C) and NATA (D). Values obtained by integrating the autocorrelation function for the transition axis of each of the sidechains (see Methods). In the headgroup region, from 0.5 to 1.5 nm, the estimated correlation times are approximate because in many cases the autocorrelation functions do not decay to zero, denoting incomplete sampling in the trajectories.

2.4.5. - Insertion Angle.

The peptide insertion angle, θ, is defined as the angle between the *z* axis and the vector pointing from the center-of-mass (COM) of the backbone to the center of mass of the sidechain. The distribution of these angles is plotted in Figure 2.5. All three simulated peptides exhibit a wide range of allowed angles in the aqueous phase, corresponding to free reorientations. [49] At *z* = 2.2 nm for tryptophan and 2.0 nm for phenylalanine and tyrosine, reorientations become resticted. Nearing the headgroup region in the *z* = 1.5 to 2.0 nm range, all three peptides insert at an angle of around 150^{o}, corresponding to the backbone pointing into the solvent and to the sidechain pointing into the lipid headgroups. For NAFA, this insertion angle remains stable at the preferred value until *z* = 0.5 nm, after which a broad distribution of insertion angles reappears in the membrane center in the *z* = 0.0-0.5 nm range. NAYA undergoes a systematic change in the preferred insertion angle while permeating the membrane. The angle changes to 90^{o} at *z* = 1.5 nm, 60^{o} at *z* =0.8 nm, 30^{o} for *z* in the 0.1-0.6 nm range, and finally rotates freely at the center. For NATA, the insertion angle changes in a manner somewhere between the NAFA and NAYA cases. In NATA, the preferred insertion angle changes to 120^{o} for *z* between 1.0 and 2.0 nm and to about 90^{o} for *z* between 0.5 and 0.8 nm, with mostly unrestricted orientations within the *z* = 0.0 to 0.5 nm range. NATA's ability to reorient in the central membrane region, as found in our simulations, agrees with the results obtained previously by umbrella sampling and milestoning from other groups.[44, 83]

In general, all three aromatic dipeptides tend to initiate interactions with the mebrane by orienting their sidechains toward the lipid headgroups and backbones towards the solution.

The most hydrophobic peptide, NAFA, retains this orientation throughout the headgroup region, while NAYA and NATA assume an orientation parallel to the membrane surface at the interface. Finally, at the center of the membrane, the most hydrophilic peptide, NAYA, reorients to a direction that is almost opposite of its value in solution, before again assuming free rotation at z = 0.0 nm. In contrast, the orientation of the more hydrophobic peptides, NAFA and NATA, is unrestricted over a wider range of positions, 0.0 to 0.5 nm. This difference in behavior may be the result of water molecules being pulled into the membrane interior by the permeating peptides (see Special Interactions below in Section 2.4.10).

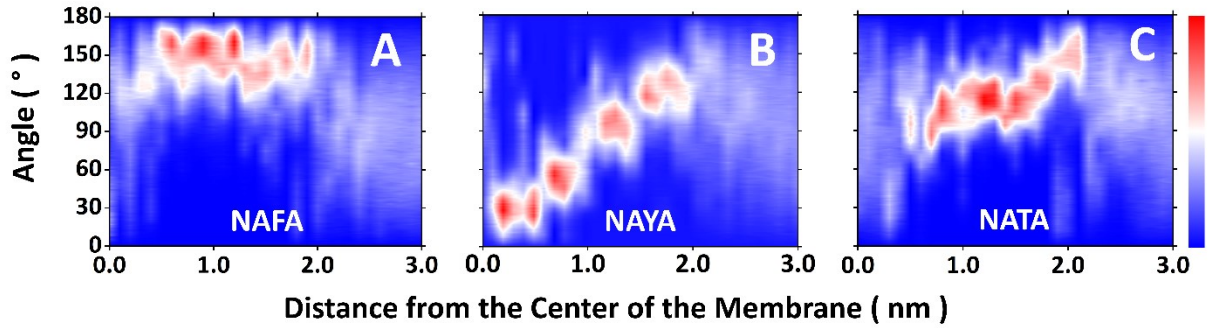


Figure 2.5. Insertion angles, θ, for NAFA (A), NAYA (B), and NATA (C). A vector is defined from the center of mass of the backbone to the center of mass of the sidechain. The angle θ is between this vector and the positive z-axis of the simulation box. The normalized probability density, $P(\theta,z)$, is plotted with blue representing little or no probability and with red representing the highest probability. The insertion vector is plotted with respect to the distance from the center of the membrane.

2.4.6. - Backbone Conformations.

Backbone conformations are represented by two dihedral angles, φ and ψ. φ is the dihedral angle formed by the carbon atom of the acetyl blocking group, the terminal nitrogen atom, the peptide C_α carbon, and the terminal carbon atom. ψ is the dihedral angle formed by

the nitrogen of the amidated C-terminus, the terminal carbon atom, the peptide C_α carbon, and the terminal nitrogen atom. A plot of φ and ψ angles is called a Ramachandran[89] plot, as depicted in Figure 2.6. In the aqueous region of the simulation box, the φ and ψ angles of NAFA, NAYA and NATA are grouped into three areas: the large region with negative ψ angles represents α-helical type conformations, the region in the top left corner represents β-sheet and extended structures, and the small region in between to the C_{7eq} conformer characteristic of dipeptides.[90] As the dipeptides move from solution to interface, the probabilities in the α and β regions decrease, while that of C_{7eq} increases. For all three dipeptides, C_{7eq} becomes the dominant structure at the center of the membrane. The three peptides exhibit a clear structural response to the different chemical environments in solution, at the memberane interface, and in the center. The role of such conformational change in more complicated structures is worthy of further investigation.

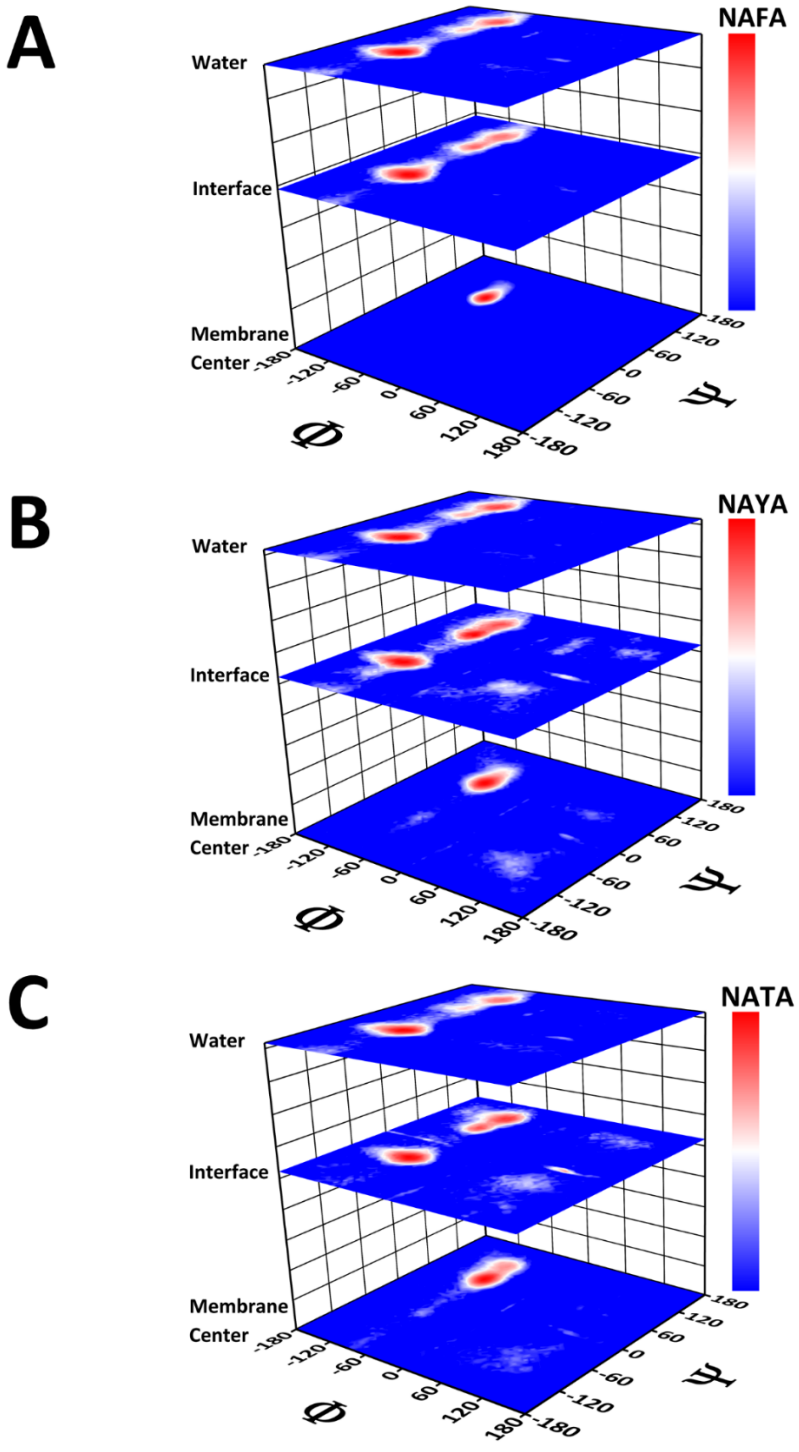


Figure 2.6. Ramachandran plots, in terms of the natural log of the probability density, are presented above for (A) NAFA, (B) NAYA, and (C) NATA at the center of the lipid bilayer, z = 0 (bottom), the interfacial region at z = 1.5 nm (middle), and the aqueous region at z = 3.0 nm. Phi and psi angles are plotted from -180° to +180°.

2.4.7. - Sidechain Conformations.

Distributions of the sidechain angles (χ_1 , χ_2) for each peptide were also examined (Figure 2.7). The nonpolar phenyl ring of NAFA sampled the same four main conformers along the whole permetaion path (tg-, tg+, g-g-, g-g+). The phenol ring of NAYA sampled two main sidechain conformers along the whole path (tt, tg-). The indole ring of NATA explored two main conformers in solution (tt, tg-), one at the interface (tt), and two in the bilayer center (tg-, tg+). Therefore, the more polar phenol and indole groups exhibited different conformational preferences than the nonpolar phenyl.

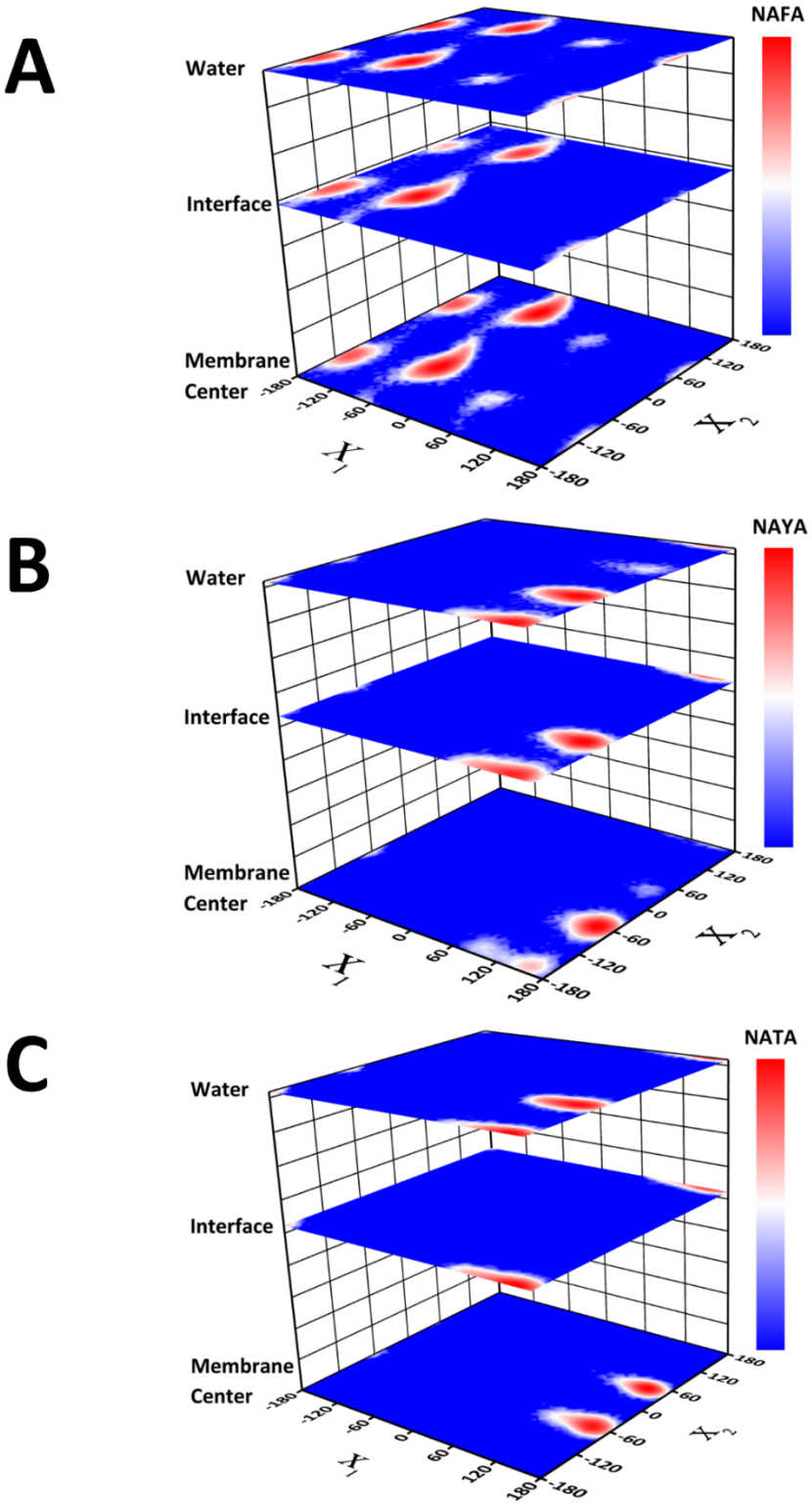


Figure 2.7. Peptide sidechain conformations as a function of membrane insertion *z*. Natural log of population for probability of (χ_1 , χ_2) distributions for (A) NAFA, (B) NAYA and (C) NATA.

2.4.8. - Peptide Structures - Clustering.

Clustering of the trajectories in dihedral angle space confirmed that the studied peptides undergo definite structural changes during mebrane permeation. Examples of main sampled conformations are shown in Figure 2.8. For NAFA, seven clusters were found in the solution region, with 4 corresponding to α, 1 to β, and 3 to C_{7eq} structures; at the interface, four clusters were identified, 2 α and 2 C_{7eq} and in the membrane center four clusters of C_{7eq} type were sampled, with different sidechain combinations. For NAYA, 37 total clusters were found in solution, of which 8 had populations above 0.5% - 4 α, 1 β and 3 C_{7eq} . At the interface, NAYA sampled only two clusters: one in the α region and one in the C_{7eq} region. At the center of the membrane, NAYA sampled three clusters, all of C_{7eq} type. For NATA, the situation was the simplest – there were only two clusters in each environment – 1 α and 1 C_{7eq} in solution and at interface, and two C_{7eq} at the center of the membrane. The growing population of the C_{7eq} conformation upon membrane insertion may be illustrated by the changes in the O to N distance between the carbonyl oxygen of the N-terminal blocking group and the nitrogen of the C-terminal blocking group shown in Figure 2.9. In solution, this distance fluctuates over a relatively wide range of values, 3 to 5 nm, while a narrow range of 2.8 to 3.1 nm is sampled inside of the membrane. Therefore, in the hydrophobic environment of the acyl lipid chains, the peptide interacts with itself, partially shielding the polar backbone atoms from the external environment.

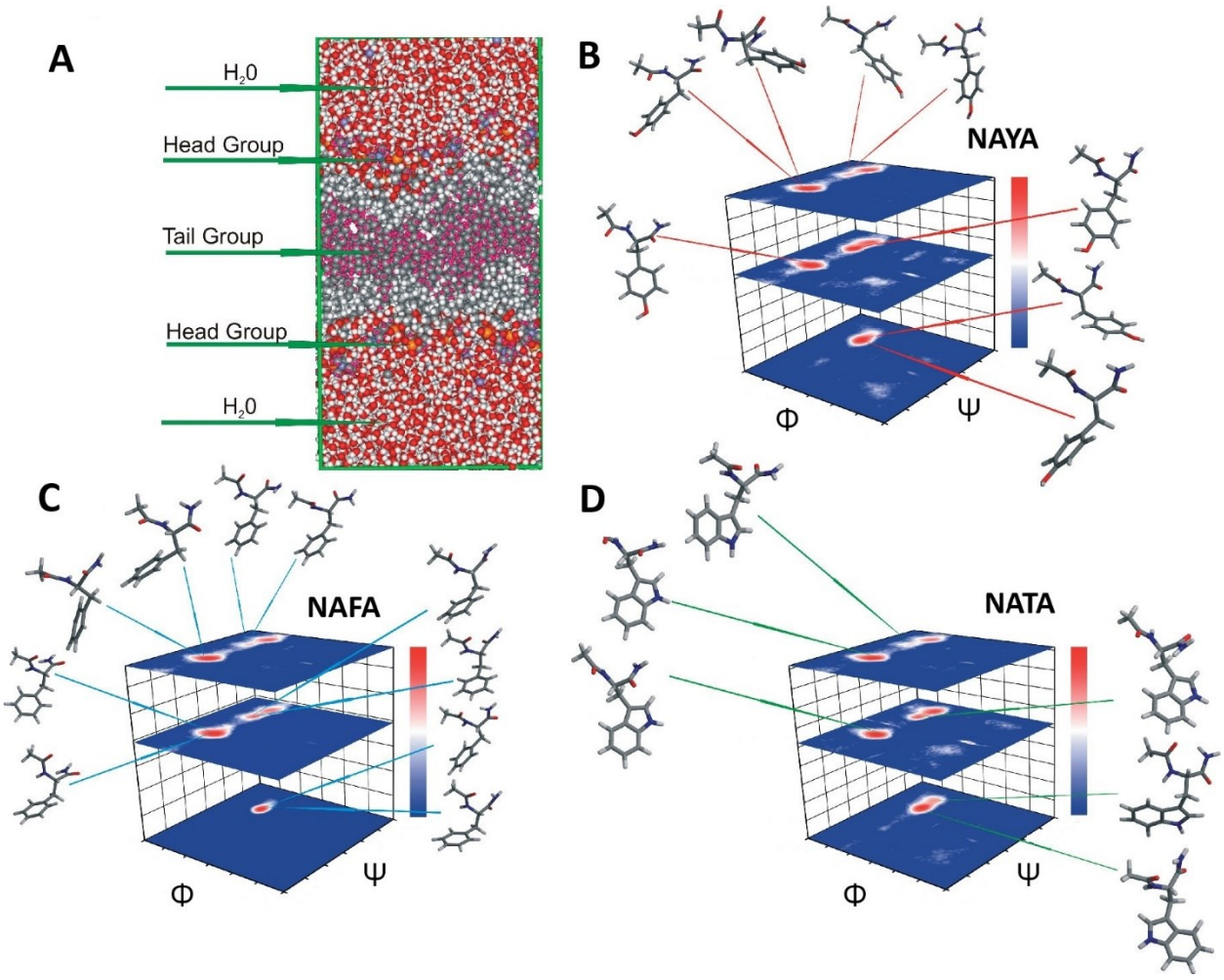


Figure 2.8. Representative structures from trajectory clustering are shown. Lipid bilayer with water molecules, hyrophilic head groups, and hydrophobic tail groups are shown in (A). Central structures of structural clusters corresponding to the alpha and C_{7eq} free energy minima are shown in NAYA (B), in NAFA (C), and in NATA (D). Representative structures are indicated on the same Ramachandran plots depicted in Figure 2.6.

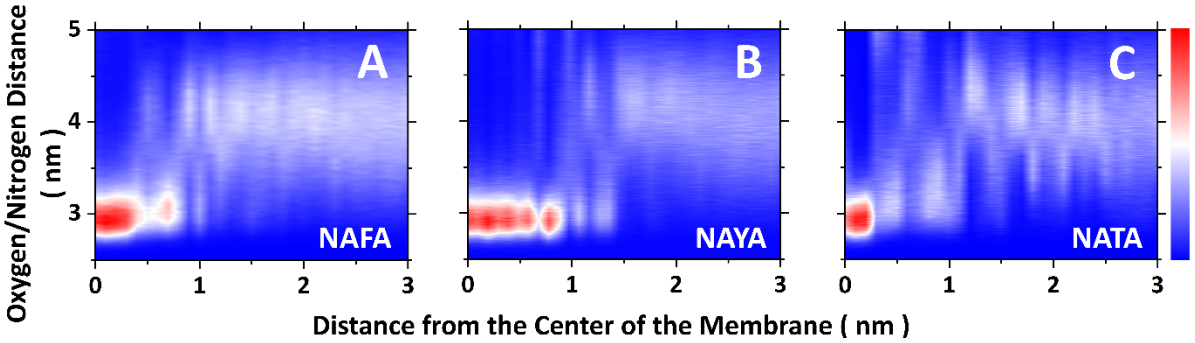


Figure 2.9. Distributions of O to N distances between blocking groups in (A) NAFA, (B) NAYA and (C) NATA peptides as function of distance from membrane center. Figures display probability distributions $P(r_{O...N}, z)$.

2.4.9 - Molecule Shape and Size.

Changes of the radius of gyration (R_g) which is the mass weighted root mean square atomic distance from the center of mass of the molecule, and solvent accessible surface area (SASA) with membrane insertion are shown in the Supplementary Information. The peptide size, measured by R_g, and the solvent exposure, measured by SASA, do not exhibit significant variation as a function of membrane insertion. Only in the case of NATA, a small effect of slight lowering of R_g and SASA is found in the tail region, z = 0.0 to 0.5 nm. Interestingly, the conformational changes observed for the peptides as they translocate across the mebrane have little effect on SASA and R_g for these relatively small systems.

2.4.10. - Specific Interactions.

Interactions between the peptides and their environment are analyzed below. Figure 2.10 shows the average number of water molecules within 0.3 nm the peptide as a function of

z, representing roughly the first solvation shell. Peptides were solvated by 45 - 50 waters in the aqueous region. The number of waters decreased systematically with insertion depth, reaching around 20 at the interface (*z* = 1.5 nm), around 15 at *z* = 1.0 nm, and around 5 at *z* = 0.5 nm. All three peptides are surrounded by a large number of water molecules as they approach the membrane. Even at the center, *z* = 0.0 nm, residual solvation remains: 0.7, 0.1, and 3.4, on average, for NAFA, NAYA and NATA, respectively. Previous simulations of NATA in DOPC found 4 waters solvating the peptide in membrane center, very close to our result.[44]

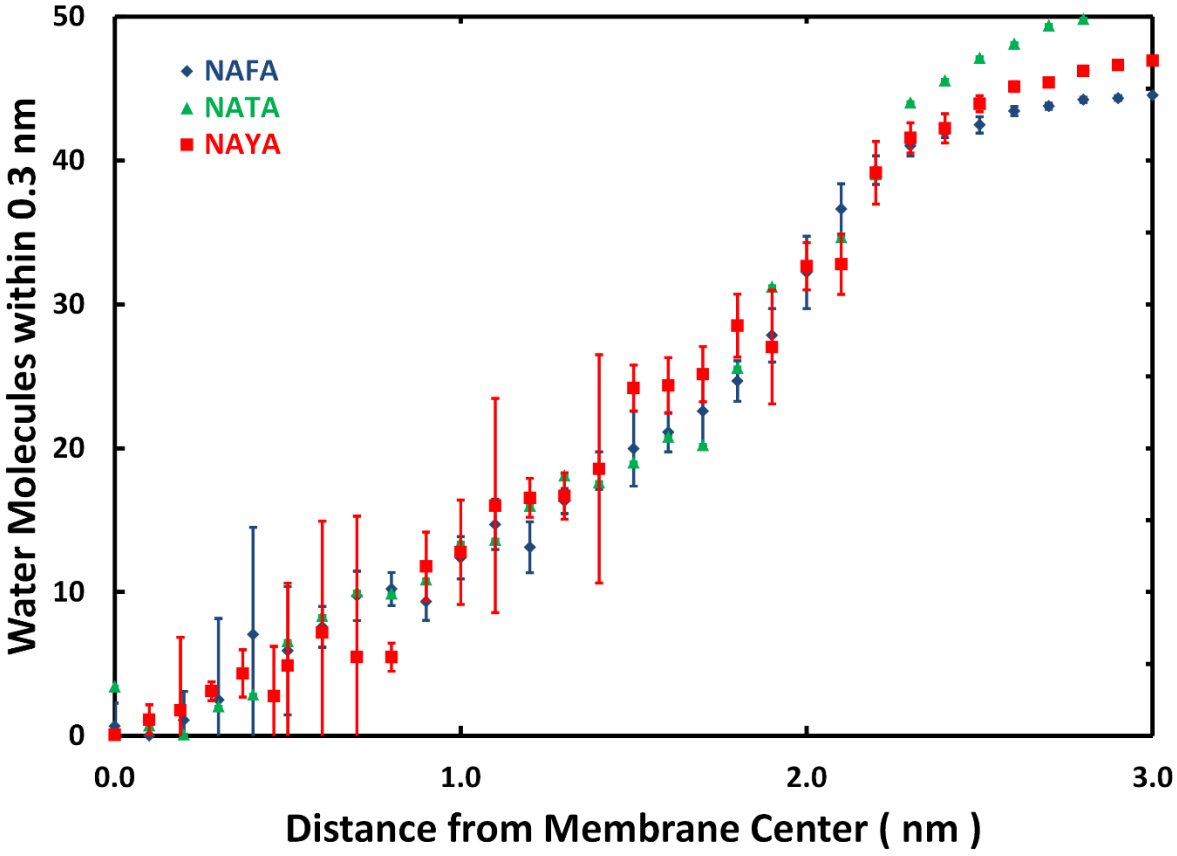


Figure 2.10. Average number of water molecules within 0.3 nm of the peptides. Error bars represent a 95% confidence interval for each trajectory. Note: the 95% confidence interval for the sample mean for each trajectory is too small to be seen when plotted.

Figure 2.11 displays the details on specific interactions of the peptides and the lipids. The average coordination numbers show the expected trends – upon insertion into the membrane, the number of solvating waters systematically decreases, the number of acyl carbons increases, and the number of headgroup atoms first increases to a maximum in the interfacial region and then drops off. For the backbone (Figures 2.12A-C), there are some interesting differences between the three peptides. The highest backbone coordination by headgroups occurs at z = 1.2 nm for NAFA, 2.0 nm for NAYA, and 1.6 nm for NATA. For NAYA, the coordination between the backbone atoms and the lipid tails is generally greater than for the other peptides. Similarly, coordination between the backbone of NAYA and water is generally lower than for the other peptides. The switchover distance, at which the coordination by water and the acyl tails becomes equal, is at 1.2, 1.6 and 1.4 nm for the NAFA, NAYA, and NATA backbone, respectively. The behavior of the peptide sidechains (Figures 3.12D-F) is similar to the backbone. The highest sidechain coordination by headgroups occurs at z = 2.0 nm for NAFA, 2.2 nm for NAYA, and 2.0 nm for NATA. This is consistent with the membrane insertion angles discussed above: with NAFA inserting mostly sidechain-first; NAYA mostly at a 90^{o} "sideways" angle, with both the sidechain and backbone parallel to the membrane surface; and, NATA intermediate between the other two peptides. All three peptides appear to be able to interact well with the lipid tails, reaching tail coordination numbers at z = 0 of similar value to those by water at z = 3.0 nm.

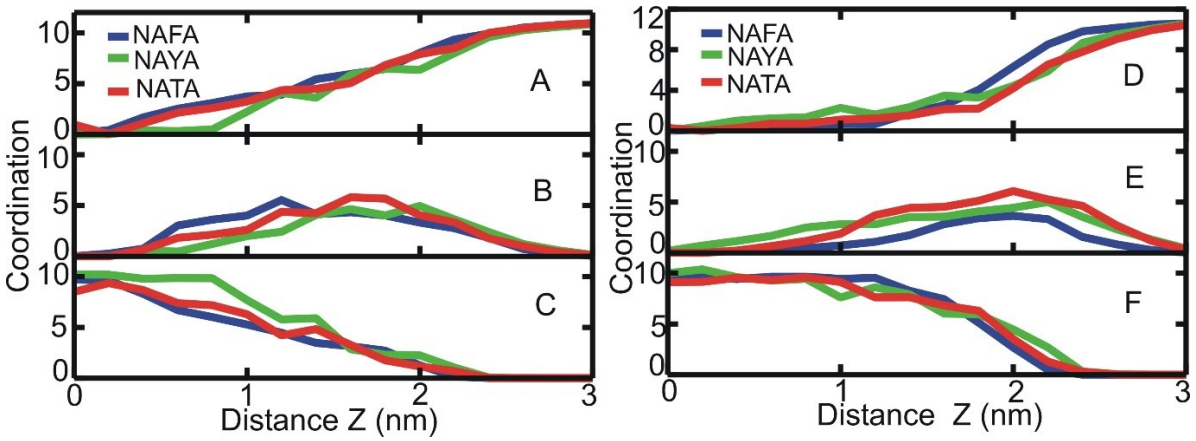


Figure 2.11. Average coordination numbers of the three peptides as function of insertion depth *z*. Coordination of backbone atoms of peptides by (A) water, (B) headgroups and (C) lipid tails. Coordination of sidechain atoms of peptides by (D) water, (E) headgroups and (F) lipid tails. The presented X – Y values are the average number of atoms of species Y within 0.5 nm of an atom of species X. Species are: backbone – non-hydrogen atoms of peptide backbone and blocking groups; sidechains – non-hydrogen atoms of peptide sidechains; water - water oxygens; tails – lipid acyl tail carbon atoms; heads – headgroups, non-hydrogen atoms of lipids, excluding tails.

2.5. Conclusions

The results of a computational study of the passive permeation of three aromatic dipeptides – NAFA, NAYA and NATA - through DOPC lipid bilayers were presented. Umbrella sampling simulations were performed for the three systems, using at least 30 windows of 50-100 ns length for each peptide to model the thermodynamics, dynamics, and microscopic interactions along the chosen one-dimensional reaction path, *z*, the center-of-mass distance between the peptide and the lipid bilayer. The calculated profiles of the potential of mean force show two strong effects: preferential binding of each of the three peptides to the lipid interface, with free energies of -18 $kJmol^{-1}$ of NAFA, -14 $kJmol^{-1}$ for NAYA , and -12 $kJmol^{-1}$ for NATA and large free energy barriers in the membrane center of +28 $kJmol^{-1}$ for NAFA, +41 $kJmol^{-1}$ for NAYA, and +44 $kJmol^{-1}$ for NATA, relative to the respective minima. We used three

approaches to calculate the position-dependent translational diffusion coefficients *D(z)*. Two methods agree with each other and agree with independent estimates from the pure solution phase: one based on the numerical solution of the Smoluchowski equation and the other based on force autocorrelations from short trajectories with constrained values of *z*. Unfortunately, use of autocorrelations from the restraining force significantly overestimates values for the diffusion constants. Surprisingly, computed *D(z)* values change very little with the reaction coordinate and are also quite similar for the three peptides studied. In contrast, calculated values of the sidechain rotational correlation times, $\tau_{rot}(z)$, show extremely large changes with peptide membrane insertion: values become 100 times larger in headgroup region and 10 times larger in membrane center, relative to solution. Therefore, it appears that these small peptides can relatively easily undergo translational diffusion along the *z* axis, parallel to the lipid molecules, while reorientations, involving motion perpendicular to the lipids, is strongly hindered, especially in the tightly packed headgroup region.

Analysis of the insertion angle shows the peptides inserting initially with the sidechain, pointing into the membrane and backbone into solution. While NAFA retains this preferred orientation through the interfacial and headgroup regions, NAYA systematically switches to a backbone-first orientation as its insertion progresses, while NATA behaves in an intermediate fashion, changing to a perpendicular orientation. In the central region of the membrane, all three peptides sample the full range of insertion angles. Peptide conformational freedom becomes systematically restricted as they enter the membrane: α, β and C_{7eq} regions are explored in solution, α and C_{7eq} at the interface, and only C_{7eq} in the center. The C_{7eq} structures are characterized by a short contact between the polar atoms of the blocking groups. In the

permeation process, the peptide interactions with water are replaced first by the lipid headgroups and then by the lipid sidechains. The passage of the sidechain and backbone through the different regions is consistent with the insertion angle analysis. Some residual water molecules of solvation remain even in the membrane center, deforming the membrane structure, as previously noted.[44]

The experiments described in this work provide useful baseline information for aromatic peptide membrane permeation processes, namely passage times and permeation coefficients. Our calculated values of passage times are several orders of magnitude smaller than experimental data, while permeation coefficients for NATA and NAYA are in reasonable agreement. This effect has been found in previous studies and has been attributed primarily to the presence of more than one slow variable characterizing membrane permeation, among which peptide orientation and large-scale membrane structural fluctuations have been proposed.[26, 28, 31, 91] The accommodation of additional slow variables is an ongoing topic of current inquiry.

Excluding the passage times and permeation coefficients, our simulation results are in reasonable agreement with most studies on similar systems, including computer simulations with different force fields and experimental measurements. Our PMF profile for NATA qualitatively agrees with results of Cardenas *et al.*,[44] and the calculated translational and rotational diffusion rates in the solution region agree with both computational results and experimental data.[16, 19-22, 26, 28, 31] As a result, this computational study provides an improved understanding of the process of transmembrane permeation of small aromatic peptides. The

microscopic insights from our simulations are particularly valuable, including the large difference between translational and rotational diffusion rates and changes in peptide structure as a function of membrane insertion depth. Membrane permeation by flexible amphiphilic molecules remains a fruitful area for further studies.

2.6. Supplementary Information

2.6.1. - Initial Membrane Electron Density.

The initial simulation box is very similar to the experimentally determined electron densities by Liu and Nagle for a series of stacked lipid bilayers, as seen in Figure 2.12.[57] Electron densities are often compared and used in the validation and verification of various force fields. The experimental results are "squashed" together a bit due to the use of several stacked layers, as opposed to the computation results which represent only a single lipid bilayer.

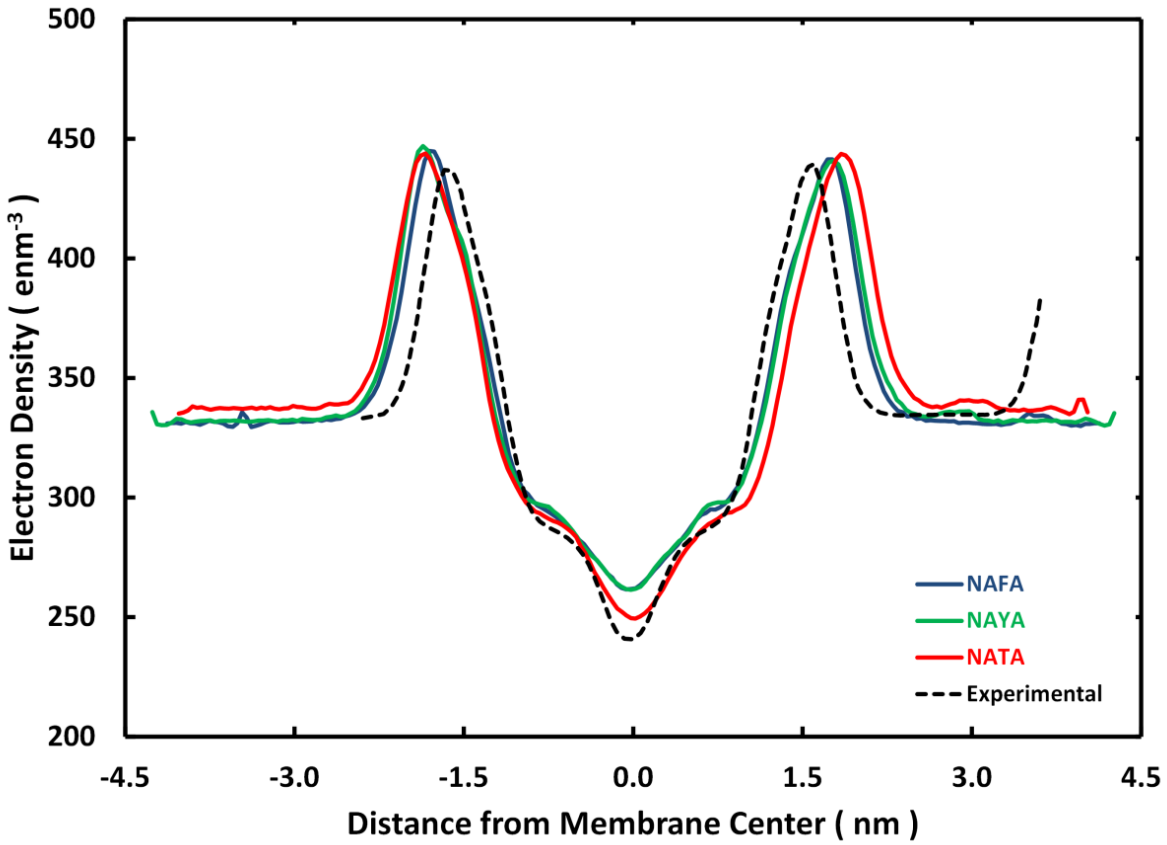


Figure 2.12. Comparison of the total electron density for each aromatic dipeptide simulation and experimental results.

2.6.2. - PMF Quartiles.

In the following figures (Figures 2.13, 2.14, and 2.15), the simulation data were split into quartiles, and the resulting potentials of mean force were plotted. These variations were much larger than those found by using the latter half of the simulation data with the bootstrap method.

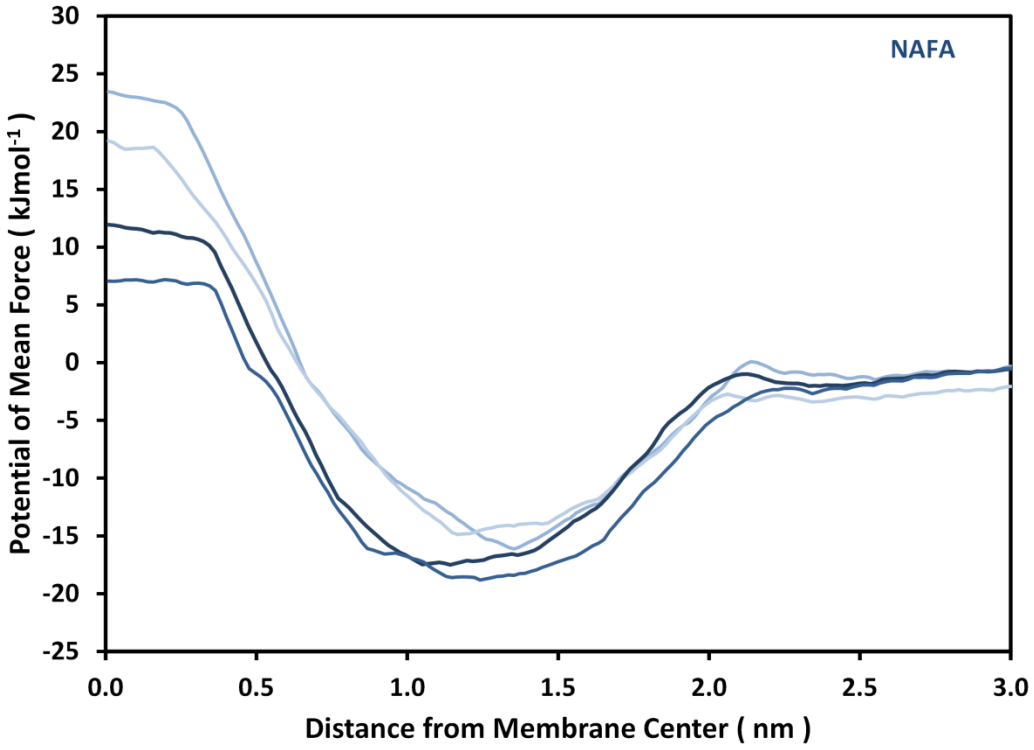


Figure 2.13. NAFA potential of mean force quartiles.

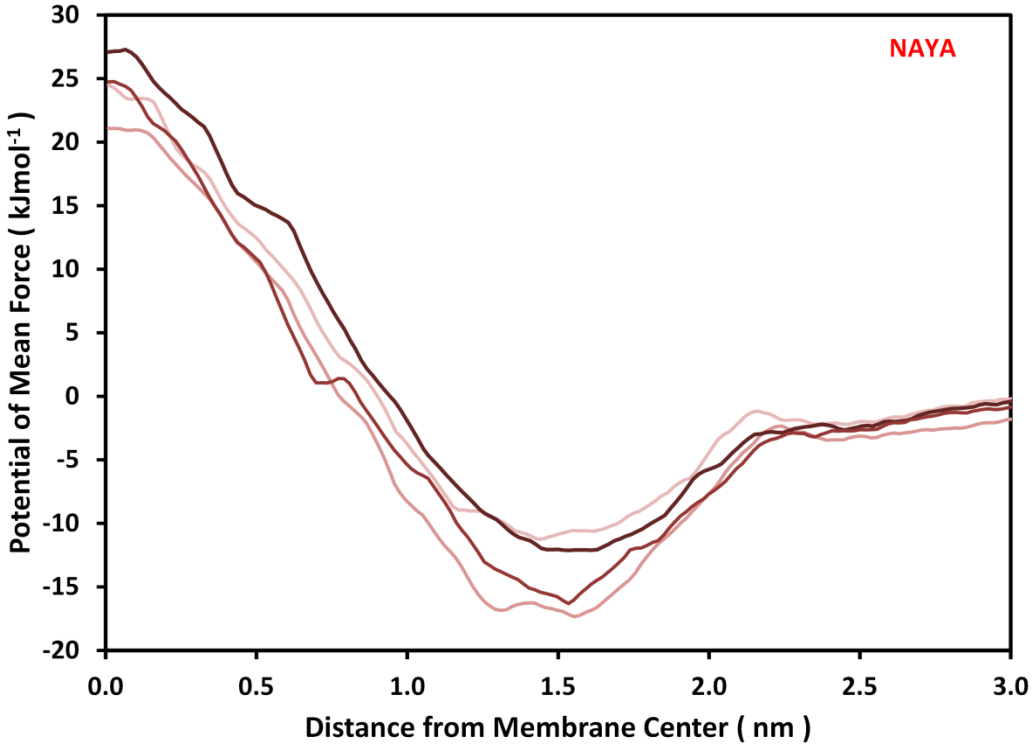


Figure 2.14. NAYA potential of mean force quartiles.

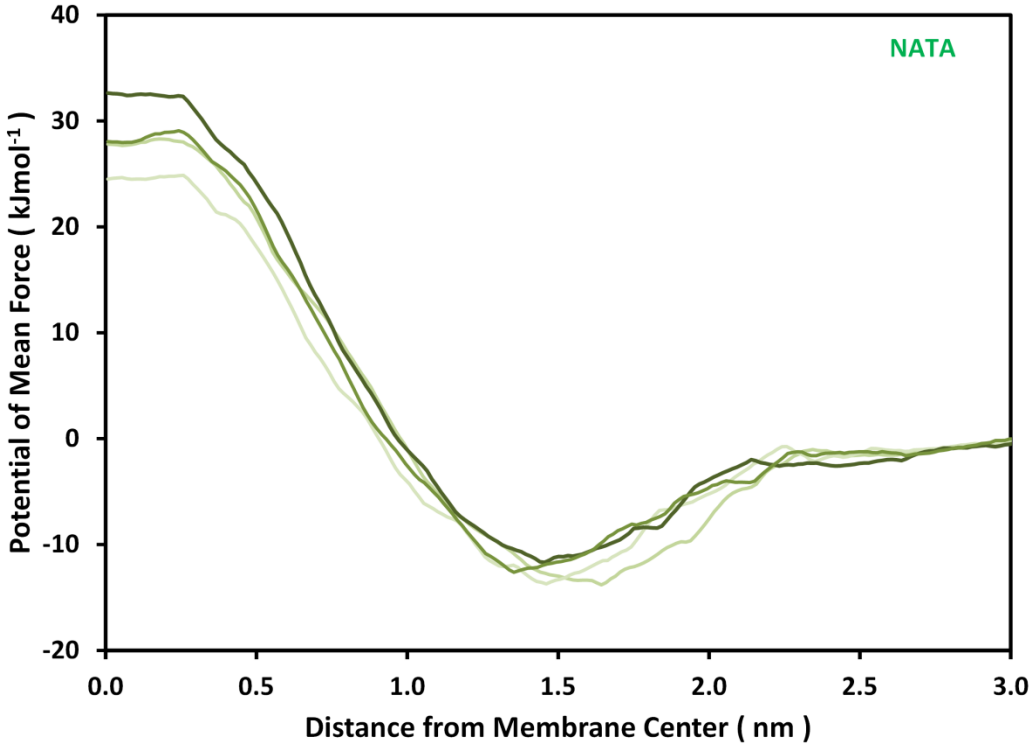


Figure 2.15. NATA potential of mean force quartiles.

2.6.3. - Alternative Diffusion Constant Methods.

Figure 2.16 depicts the diffusion constants, as calculated from fixed-*z* simulations with a custom-made version of CHARMM 36.

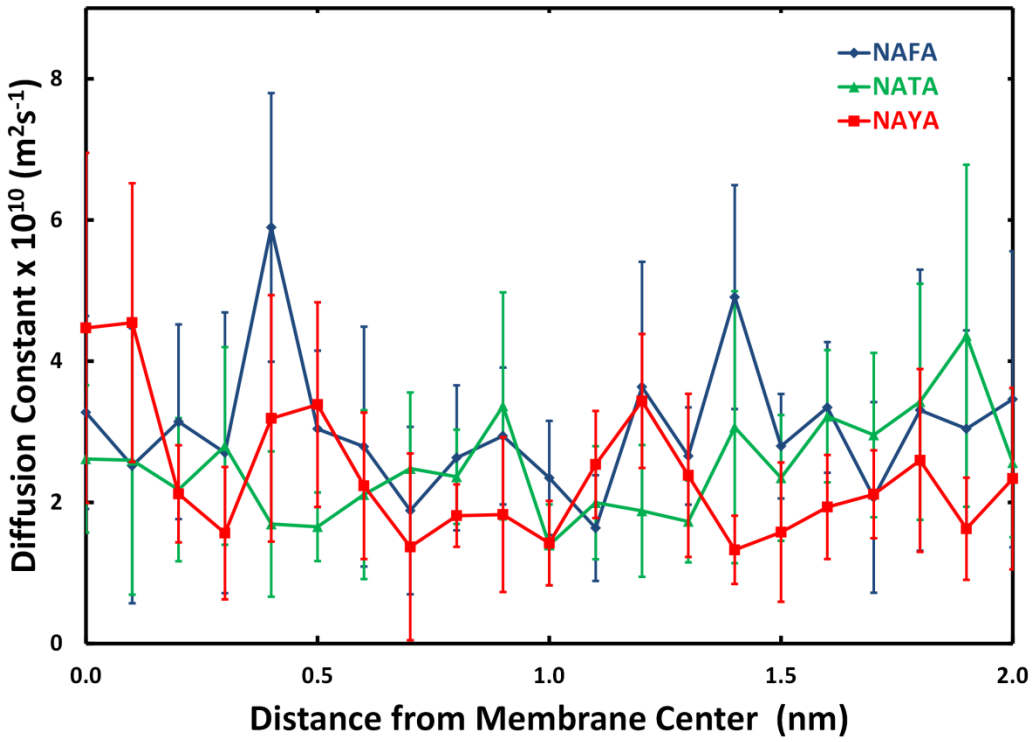


Figure 2.16. Diffusion constants for the aromatic dipeptides from unconstrained force fluctuations.

Figure 2.17 depicts the diffusion constants, as calculated from the pulling forces of the constrained umbrella sampling window. The smallest diffusion constants are on the order of 10^{-11} m^2s^{-1}.

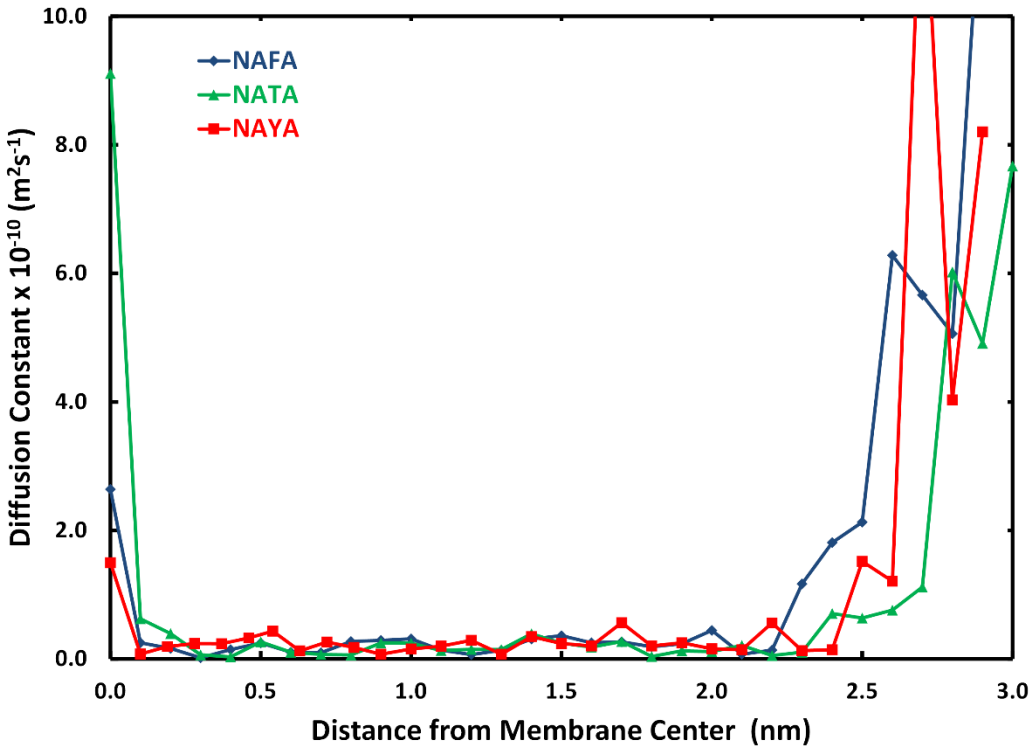


Figure 2.17. Diffusion constants from the fluctuation of the umbrella pulling forces.

2.6.4. - Radius of Gyration.

The radius of gyration for each dipeptide varies little in Figure 2.18; however, NATA shows a small decrease near the center of the membrane. Few of the observed differences between the aromatic dipeptides are statistically significant, however; and, the displayed error bars represent the standard error for the sample means of each trajectory at a given distances. Surprisingly, these fluctuations are rather large.

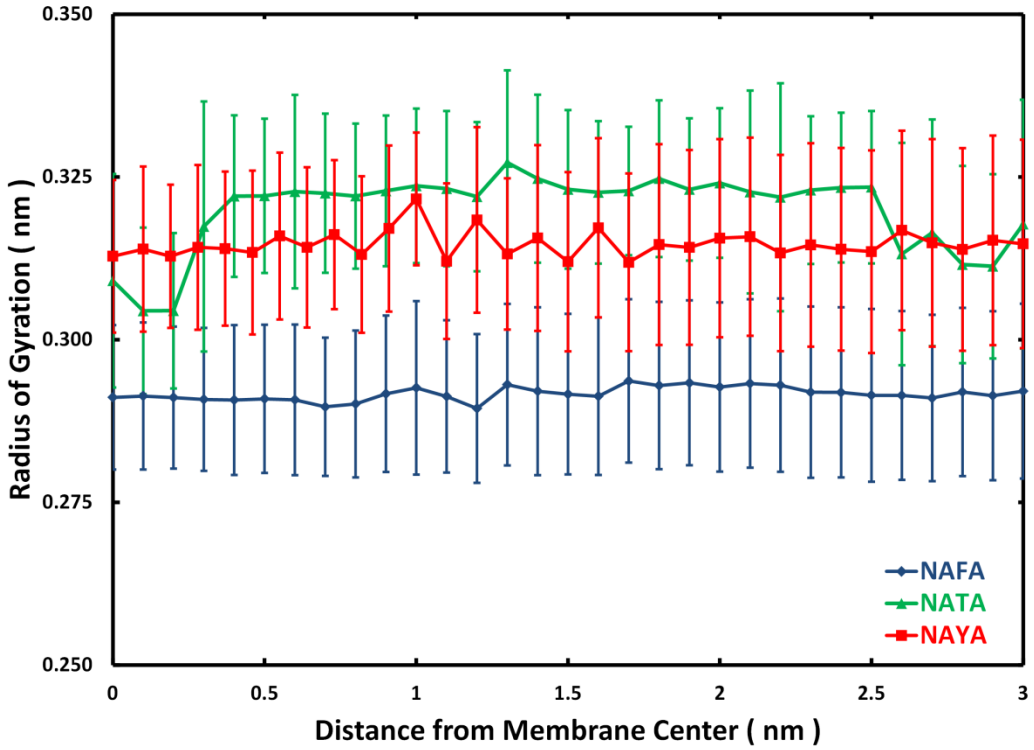


Figure 2.18. Radius of gyration for the three aromatic dipeptides.

2.6.5 - Solvent Accessible Surface Area.

The solvent accessible surface area changes little; however, the hydrophilic accessible surface area for NATA decreases slightly as the hydrophobic surface area increase similarly. In general, the accessible surface area in Figures 2.19 and Figures 2.20 generally follows the trend in dipeptide size: NAFA is smaller than NAYA which is smaller than NATA. It is interesting to note that the total accessible surface area in Figure 2.21 contains similar trends to those found in the radius of gyration in Figure 2.18; unfortunately, neither trend is statistically significant.

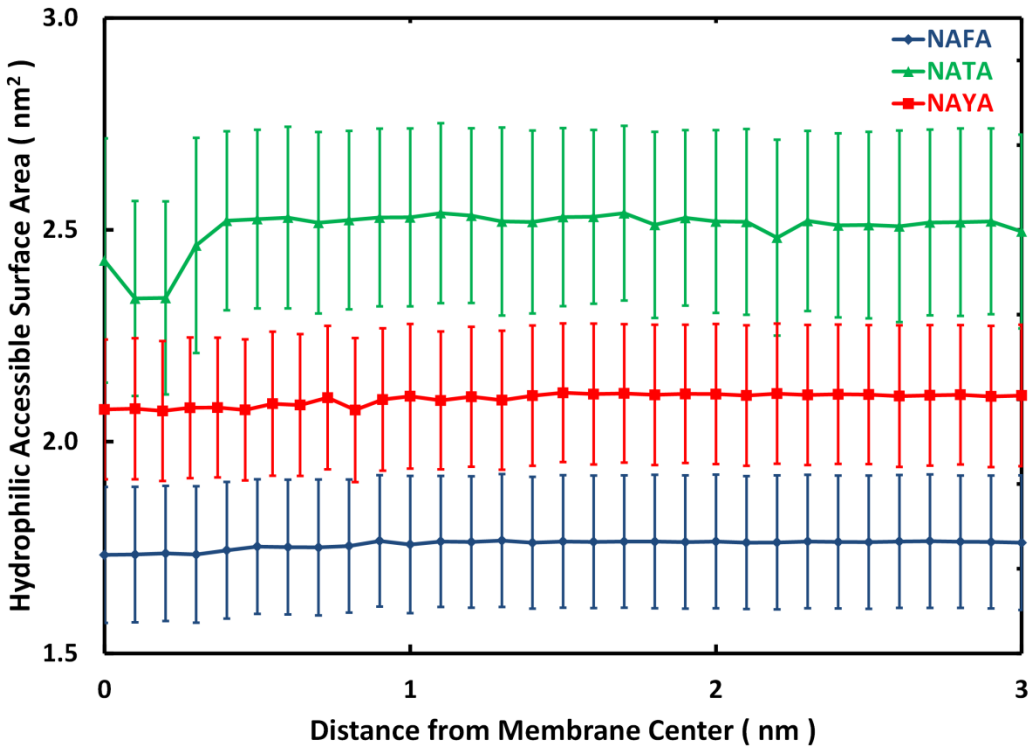


Figure 2.19. Hydrophilic accessible surface area for the three aromatic dipeptides.

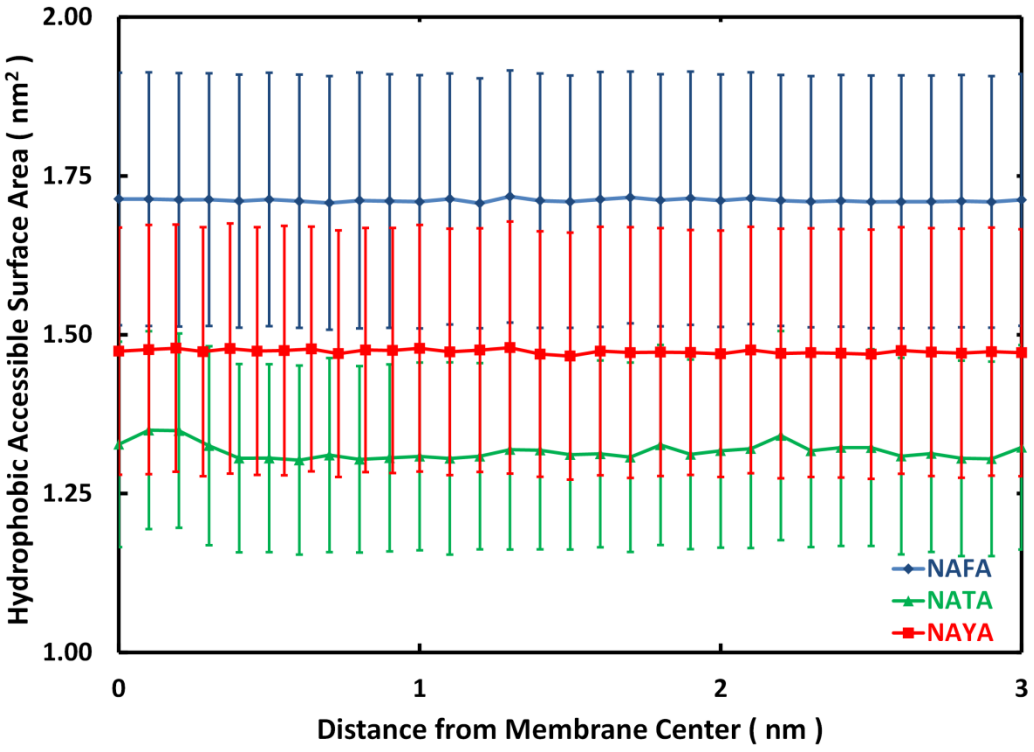


Figure 2.20. Hydrophobic accessible surface area for the three aromatic dipeptides.

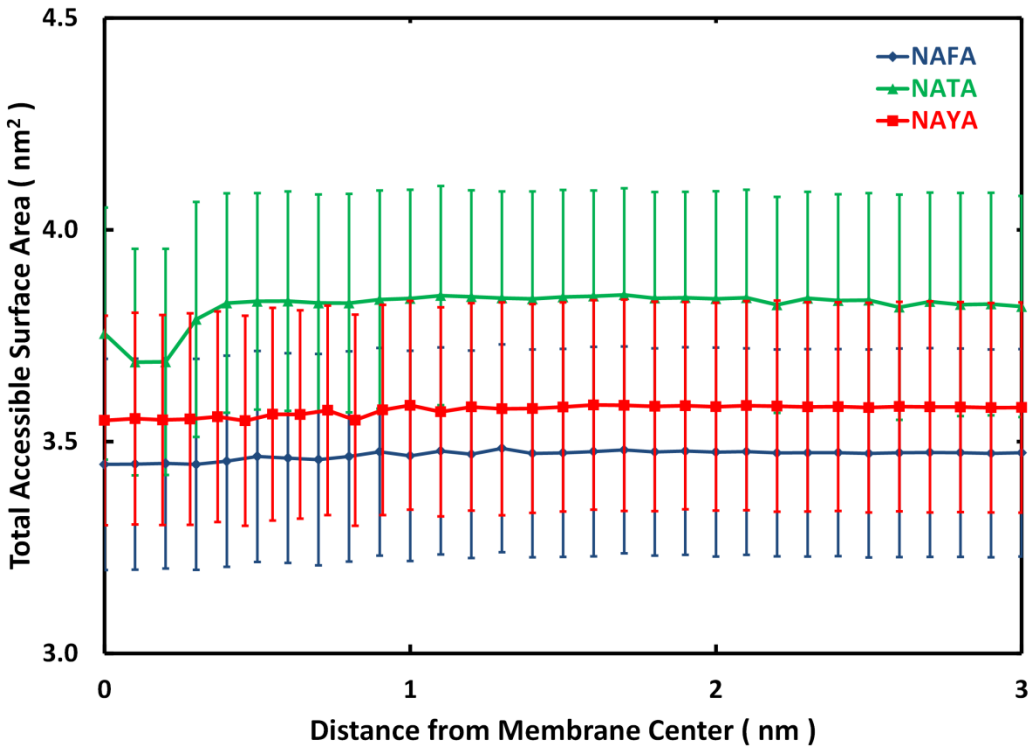


Figure 2.21. Total accessible surface area for the three aromatic dipeptides.

2.7. References

1. Lee, B. L.; Kuczera, K.; Middaugh, C. R.; Jas, G. S., Permeation of the three aromatic dipeptides through lipid bilayers: Experimental and computational study. *J. Chem. Phys.* **2016,** *144* (24), 14.

2. Deamer, D. W., Origins of life - How leaky were primitive cells? *Nature* **2008,** *454* (7200), 37-38.

3. Adamala, K.; Szostak, J. W., Competition between model protocells driven by an encapsulated catalyst. *Nature Chemistry* **2013,** *5* (6), 495-501.

4. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L., Opinion - How many drug targets are there? *Nature Reviews Drug Discovery* **2006,** *5* (12), 993-996.

5. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L., Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology* **2001,** *305* (3), 567-580.

6. Fong, C. W., Permeability of the Blood–Brain Barrier: Molecular Mechanism of Transport of Drugs and Physiologically Important Compounds. *The Journal of Membrane Biology* **2015,** *248* (4), 651-669.

7. Loverde, S. M., Molecular Simulation of the Transport of Drugs across Model Membranes. *The Journal of Physical Chemistry Letters* **2014,** *5* (10), 1659-1665.

8. Müller, D. J.; Engel, A.; Matthey, U.; Meier, T.; Dimroth, P.; Suda, K., Observing Membrane Protein Diffusion at Subnanometer Resolution. *Journal of Molecular Biology* **2003,** *327* (5), 925-930.

9. Sarangi, N. K.; Ramesh, N.; Patnaik, A., Structure and dynamics of H2O vis-á-vis phenylalanine recognition at a DPPC lipid membrane via interfacial H-bond types: Insights from polarized FT-IRRAS and ADMP simulations. *The Journal of Chemical Physics* **2015,** *142* (2), 024702.

10. Nakagawa, K., Diffusion coefficient and relaxation time of aliphatic spin probes in a unique triglyceride membrane. *Langmuir* **2003,** *19* (12), 5078-5082.

11. Kansy, M.; Senner, F.; Gubernator, K., Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. *Journal of medicinal chemistry* **1998,** *41* (7), 1007-1010.

12. Thompson, M.; Krull, U. J.; Worsfold, P. J., The Structure and Electrochemical Properties of a Polymer-Supported Lipid Biosensor. *Analytica Chimica Acta* **1980,** *117* (Jun), 133-145.

13. Thompson, M.; Krull, U. J.; Worsfold, P. J., The Analytical Potential of Chemoreception at Bilayer Lipid-Membranes. *Analytica Chimica Acta* **1980,** *117* (Jun), 121-132.

14. Thompson, M.; Krull, U. J., The Electroanalytical Response of the Bilayer Lipid-Membrane to Valinomycin - an Empirical-Treatment. *Analytica Chimica Acta* **1982,** *141* (Sep), 49-56.

15. Tieleman, D. P., COMPUTER SIMULATIONS OF TRANSPORT THROUGH MEMBRANES: PASSIVE DIFFUSION, PORES, CHANNELS AND TRANSPORTERS. *Clinical and Experimental Pharmacology and Physiology* **2006,** *33* (10), 893-903.

16. Marrink, S.; Berendsen, H., Simulation of Water Transport Through a Lipid-Membrane. *Journal of Physical Chemistry* **1994,** *98* (15), 4155-4168.

17. Hansen, F. Y.; Peters, G. H.; Taub, H.; Miskowiec, A., Diffusion of water and selected atoms in DMPC lipid bilayer membranes. *The Journal of Chemical Physics* **2012,** *137* (20), 204910.

18. Wei, T.; Huang, T.; Qiao, B.; Zhang, M.; Ma, H.; Zhang, L., Structures, Dynamics, and Water Permeation Free Energy across Bilayers of Lipid A and Its Analog Studied with Molecular Dynamics Simulation. *The Journal of Physical Chemistry B* **2014,** *118* (46), 13202-13209.

19. Comer, J.; Schulten, K.; Chipot, C., Diffusive Models of Membrane Permeation with Explicit Orientational Freedom. *Journal of Chemical Theory and Computation* **2014,** *10* (7), 2710-2718.

20. Marrink, S. J.; Berendsen, H. J. C., Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations. *The Journal of Physical Chemistry* **1996,** *100* (41), 16729-16738.

21. Orsi, M.; Sanderson, W. E.; Essex, J. W., Permeability of Small Molecules through a Lipid Bilayer: A Multiscale Simulation Study. *The Journal of Physical Chemistry B* **2009,** *113* (35), 12019-12029.

22. Riahi, S.; Rowley, C. N., Why Can Hydrogen Sulfide Permeate Cell Membranes? *Journal of the American Chemical Society* **2014,** *136* (43), 15111-15113.

23. Bemporad, D.; Luttmann, C.; Essex, J. W., Behaviour of small solutes and large drugs in a lipid bilayer from computer simulations. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2005,** *1718* (1-2), 1-21.

24. Wang, Y.; Hu, D.; Wei, D., Transmembrane Permeation Mechanism of Charged Methyl Guanidine. *Journal of Chemical Theory and Computation* **2014,** *10* (4), 1717-1726.

25. Alper, H. E.; Stouch, T. R., Orientation and diffusion of a drug analog in biomembranes: Molecular dynamics simulations. *The Journal of Physical Chemistry* **1995,** *99* (15), 5724-5731.

26. Orsi, M.; Essex, J. W., Permeability of drugs and hormones through a lipid bilayer: insights from dual-resolution molecular dynamics. *Soft Matter* **2010,** *6* (16), 3797.

27. Martin, L. J.; Chao, R.; Corry, B., Molecular dynamics simulation of the partitioning of benzocaine and phenytoin into a lipid bilayer. *Biophysical Chemistry* **2014,** *185*, 98-107.

28. Porasso, R. D.; Drew Bennett, W. F.; Oliveira-Costa, S. D.; López Cascales, J. J., Study of the Benzocaine Transfer from Aqueous Solution to the Interior of a Biological Membrane. *The Journal of Physical Chemistry B* **2009,** *113* (29), 9988-9994.

29. Nademi, Y.; Iranagh, S. A.; Yousefpour, A.; Mousavi, S. Z.; Modarress, H., Molecular dynamics simulations and free energy profile of Paracetamol in DPPC and DMPC lipid bilayers. *Journal of Chemical Sciences* **2014,** *126* (3), 637-647.

30. Khuntawee, W.; Wolschann, P.; Rungrotmongkol, T.; Wong-ekkabut, J.; Hannongbua, S., Molecular Dynamics Simulations of the Interaction of Beta Cyclodextrin with a Lipid Bilayer. *Journal of Chemical Information and Modeling* **2015,** *55* (9), 1894-1902.

31. dos Santos, D. J. V. A.; Eriksson, L. A., Permeability of psoralen derivatives in lipid membranes. *Biophys. J.* **2006,** *91* (7), 2464-2474.

32. Cascales, J. J. L.; Huertas, M. L.; de la Torre, J. G., Molecular dynamics simulation of a dye molecule in the interior of a bilayer: 1,6-diphenyl-1,3,5-hexatriene in dipalmitoylphosphatidylcholine. *Biophysical Chemistry* **1997,** *69* (1), 1-8.

33. Kyrychenko, A.; Sevriukov, I. Y.; Syzova, Z. A.; Ladokhin, A. S.; Doroshenko, A. O., Partitioning of 2,6-Bis(1H-Benzimidazol-2-yl)pyridine fluorophore into a phospholipid bilayer: Complementary use of fluorescence quenching studies and molecular dynamics simulations. *Biophysical Chemistry* **2011,** *154* (1), 8-17.

34. Wei, C.; Pohorille, A., Flip-Flop of Oleic Acid in a Phospholipid Membrane: Rate and Mechanism. *The Journal of Physical Chemistry B* **2014,** *118* (45), 12919-12926.

35. Xiang; Jiang, Z.-Q.; Song, L.; Anderson, B. D., Molecular Dynamics Simulations and Experimental Studies of Binding and Mobility of 7- tert -Butyldimethylsilyl-10-hydroxycamptothecin and Its 20(S)-4-Aminobutyrate Ester in DMPC Membranes. *Molecular Pharmaceutics* **2006,** *3* (5), 589-600.

36. Nolte, T. M.; Kettler, K.; Meesters, J. A. J.; Hendriks, A. J.; van de Meent, D., A semi-empirical model for transport of inorganic nanoparticles across a lipid bilayer: Implications for uptake by living cells: Passive uptake of nanoparticles. *Environmental Toxicology and Chemistry* **2015,** *34* (3), 488-496.

37. Oroskar, P. A.; Jameson, C. J.; Murad, S., Surface-Functionalized Nanoparticle Permeation Triggers Lipid Displacement and Water and Ion Leakage. *Langmuir* **2015,** *31* (3), 1074-1085.

38. Fiedler, S. L.; Violi, A., Simulation of Nanoparticle Permeation through a Lipid Membrane. *Biophys. J.* **2010,** *99* (1), 144-152.

39. Casalegno, M.; Raos, G.; Sello, G., Hydrophobic aggregation and collective absorption of dioxin into lipid membranes: insights from atomistic simulations. *Phys. Chem. Chem. Phys.* **2015,** *17* (4), 2344-2348.

40. Herce, H. D.; Garcia, A. E., Cell Penetrating Peptides: How Do They Do It? *Journal of Biological Physics* **2007,** *33* (5-6), 345-356.

41. Herce, H. D.; Garcia, A. E., Molecular dynamics simulations suggest a mechanism for translocation of the HIV-1 TAT peptide across lipid membranes. *Proceedings of the National Academy of Sciences* **2007,** *104* (52), 20805-20810.

42. Khajeh, A.; Modarress, H., Effect of cholesterol on behavior of 5-fluorouracil (5-FU) in a DMPC lipid bilayer, a molecular dynamics study. *Biophysical Chemistry* **2014,** *187-188*, 43-50.

43. Söderhäll, J. A.; Laaksonen, A., Molecular Dynamics Simulations of Ubiquinone inside a Lipid Bilayer. *The Journal of Physical Chemistry B* **2001,** *105* (38), 9308-9315.

44. Cardenas, A. E.; Jas, G. S.; DeLeon, K. Y.; Hegefeld, W. A.; Kuczera, K.; Elber, R., Unassisted Transport of N-Acetyl-L-tryptophanamide through Membrane: Experiment and Simulation of Kinetics. *The Journal of Physical Chemistry B* **2012,** *116* (9), 2739-2750.

45. Cardenas, A. E.; Shrestha, R.; Webb, L. J.; Elber, R., Membrane Permeation of a Peptide: It Is Better to be Positive. *The Journal of Physical Chemistry B* **2015,** *119* (21), 6412-6420.

46. Wood, I.; Martini, M. F.; Pickholz, M., Similarities and differences of serotonin and its precursors in their interactions with model membranes studied by molecular dynamics simulation. *Journal of Molecular Structure* **2013,** *1045*, 124-130.

47. MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P., Transfer of Arginine into Lipid Bilayers Is Nonadditive. *Biophys. J.* **2011,** *101* (1), 110-117.

48. Bonhenry, D.; Tarek, M.; Dehez, F., Effects of Phospholipid Composition on the Transfer of a Small Cationic Peptide Across a Model Biological Membrane. *Journal of Chemical Theory and Computation* **2013,** *9* (12), 5675-5684.

49. MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P., Distribution of Amino Acids in a Lipid Bilayer from Computer Simulations. *Biophys. J.* **2008,** *94* (9), 3393-3404.

50. Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M., Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **1983,** *4* (2), 187-217.

51. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **2009,** *30* (10), 1545-1614.

52. Jo, S.; Kim, T.; Im, W., Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS ONE* **2007,** *2* (9), e880.

53. Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W., CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *Journal of Computational Chemistry* **2014,** *35* (27), 1997-2004.

54. Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W., CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **2009,** *97* (1), 50-58.

55. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013,** *29* (7), 845-854.

56. Petrache, H. I.; Tristram-Nagle, S.; Gawrisch, K.; Harries, D.; Parsegian, V. A.; Nagle, J. F., Structure and fluctuations of charged phosphatidylserine bilayers in the absence of salt. *Biophys. J.* **2004,** *86* (3), 1574-1586.

57. Liu, Y.; Nagle, J. F., Diffuse scattering provides material parameters and electron density profiles of biomembranes. *Physical Review E* **2004,** *69* (4).

58. Pastor, R. W.; MacKerell, A. D., Development of the CHARMM Force Field for Lipids. *The Journal of Physical Chemistry Letters* **2011,** *2* (13), 1526-1532.

59. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **2009**, NA-NA.

60. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926.

61. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008,** *4* (3), 435-447.

62. Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*. Oxford University Press: New York, 1990; p 385.

63. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007,** *126* (1), 014101.

64. Torrie, G.; Valleau, J., Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *Journal of Computational Physics* **1977,** *23* (2), 187-199.

65. Kästner, J., Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011,** *1* (6), 932-942.

66. Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. the Method. *Journal of Computational Chemistry* **1992,** *13* (8), 1011-1021.

67. Souaille, M.; Roux, B., Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer Physics Communications* **2001,** *135* (1), 40-57.

68. Roux, B., The calculation of the potential of mean force using computer-simulations. *Computer Physics Communications* **1995,** *91* (1-3), 275-282.

69. Bicout, D. J.; Szabo, A., Electron transfer reaction dynamics in non-Debye solvents. *J. Chem. Phys.* **1998,** *109* (6), 2325-2338.

70. Hummer, G., Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New Journal of Physics* **2005,** *7*, 34-34.

71. Kubo, R., Fluctuation-Dissipation Theorem. *Reports on Progress in Physics* **1966,** *29*, 255-&.

72. Nosé, S., A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **1984,** *81* (1), 511.

73. Hoover, W., Canonical Dynamics - Equilibrium Phase-Space Distributions. *Physical Review A* **1985,** *31* (3), 1695-1697.

74. Diamond, J. M.; Katz, Y., Interpretation of Nonelectrolyte Partition-Coefficients between Dimyristoyl Lecithin and Water. *Journal of Membrane Biology* **1974,** *17* (2), 121-154.

75. Schulten, K., Dynamics of reactions involving diffusive barrier crossing. *The Journal of Chemical Physics* **1981,** *74* (8), 4426.

76. Ulander, J.; Haymet, A. D. J., Permeation across hydrated DPPC lipid bilayers: Simulation of the titrable amphiphilic drug valproic acid. *Biophys. J.* **2003,** *85* (6), 3475-3484.

77. Kuczera, K.; Unruh, J.; Johnson, C. K.; Jas, G. S., Reorientations of Aromatic Amino Acids and Their Side Chain Models: Anisotropy Measurements and Molecular Dynamics Simulations. *The Journal of Physical Chemistry A* **2010,** *114* (1), 133-142.

78. Hub, J. S.; de Groot, B. L.; van der Spoel, D., g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *Journal of Chemical Theory and Computation* **2010,** *6* (12), 3713-3720.

79. Cardenas, A. E.; Elber, R., Modeling kinetics and equilibrium of membranes with fields: Milestoning analysis and implication to permeation. *The Journal of Chemical Physics* **2014,** *141* (5), 054101.

80. Viswanath, S.; Kreuzer, S. M.; Cardenas, A. E.; Elber, R., Analyzing milestoning networks for molecular kinetics: Definitions, algorithms, and examples. *The Journal of Chemical Physics* **2013,** *139* (17), 174105.

81. Parisio, G.; Stocchero, M.; Ferrarini, A., Passive Membrane Permeability: Beyond the Standard Solubility-Diffusion Model. *Journal of Chemical Theory and Computation* **2013,** *9* (12), 5236-5246.

82. Jämbeck, J. P. M.; Lyubartsev, A. P., Exploring the Free Energy Landscape of Solutes Embedded in Lipid Bilayers. *The Journal of Physical Chemistry Letters* **2013,** *4* (11), 1781-1787.

83. Cardenas, A. E.; Elber, R., Computational study of peptide permeation through membrane: searching for hidden slow variables. *Molecular Physics* **2013,** *111* (22-23), 3565-3578.

84. Marrink, S.-J.; Berendsen, H. J. C., Simulation of water transport through a lipid membrane. *The Journal of Physical Chemistry* **1994,** *98* (15), 4155-4168.

85. Comer, J.; Schulten, K.; Chipot, C., Calculation of Lipid-Bilayer Permeabilities Using an Average Force. *Journal of Chemical Theory and Computation* **2014,** *10* (2), 554-564.

86. Henin, J.; Tajkhorshid, E.; Schulten, K.; Chipot, C., Diffusion of glycerol through Escherichia coli aquaglyceroporin GlpF. *Biophys. J.* **2008,** *94* (3), 832-839.

87. Shinoda, W.; Okazaki, S., A Voronoi analysis of lipid area fluctuation in a bilayer. *The Journal of Chemical Physics* **1998,** *109* (4), 1517.

88. Neale, C.; Madill, C.; Rauscher, S.; Pomès, R., Accelerating Convergence in Molecular Dynamics Simulations of Solutes in Lipid Membranes by Conducting a Random Walk along the Bilayer Normal. *Journal of Chemical Theory and Computation* **2013,** *9* (8), 3686-3703.

89. Pal, D.; Chakrabarti, P., On residues in the disallowed region of the Ramachandran map. *Biopolymers* **2002,** *63* (3), 195-206.

90. Tobias, D. J.; Brooks, C. L., Conformational Equilibrium in the Alanine Dipeptide in the Gas-Phase and Aqueous-Solution - a Comparison of Theoretical Results. *Journal of Physical Chemistry* **1992,** *96* (9), 3864-3870.

91. Filipe, H. A. L.; Moreno, M. J.; Róg, T.; Vattulainen, I.; Loura, L. M. S., How To Tackle the Issues in Free Energy Simulations of Long Amphiphiles Interacting with Lipid Membranes: Convergence and Local Membrane Deformations. *The Journal of Physical Chemistry B* **2014,** *118* (13), 3572-3581.

Chapter 3

Effect of System Size and Lipid Type on the Passive Permeation of the Phenylalanine Dipeptide

3.1. Overview:

In this chapter, the effect of lipid type and system size are examined on the passive permeation of the phenylalanine dipeptide through a lipid bilayer. Specifically, umbrella sampling was used to calculate the potential of mean force for the passive permeation of the phenyalalanine dipeptide – also known as N-acetyl-phenylalanineamide and NAFA – through a lipid bilayer consisting of either 50 1,2-dioleoyl-*sn*-glycero-3-phospocholine (DOPC) lipids, 50 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) lipids, or 40 POPC lipid molecules. Diffusion constants were then calculated by numerically solving the Smoluchowski equation. Permeability coefficients and mean first passage times where then calculated. All of this and other structural properties – such as Ramachandran plots, backbone angles, clustering, peptide insertion angles, radial distribution functions, and proximal peptide water molecules – were also examined in order to determine the effect of system size and lipid type. In terms of system size, we observed a small decrease in the highest barrier of the potential of mean force and fewer sampled sidechain dihedral angle conformations with 40 versus 50 POPC lipids due to weaker membrane deformations within a smaller lipid bilayer. In terms of lipid type, DOPC contains two monounsaturated acyl chains compared to only one such acyl chain in POPC; therefore, DOPC bilayers are less ordered and more easily deformed, as seen by a much broader potential of mean force profile. The DOPC lipid also transitioned to a C_{7eq} backbone conformation at lower membrane depths than POPC. Finally, we also confirmed many of the general conclusions that were made in Chapter 2, such as the insertion angle analysis indicating that the phenylalanine dipeptide enters the membrane side-chain first, then followed by the backbone, for all three lipid systems.

3.2. Introduction:

Phenylalanine is a resonant amino acid with a benzene ring as an R-group. It was first discovered by Schulze and Barbieri in 1879 from germinated and etiolated lupine seedlings and can be readily extracted by the hydrolysis of any protein rich compound, such as cheeses.[1] In 1882, phenylalanine was first synthesized by Erlenymeyer and Lipp from acetaldehyde, hydrogen cyanide, and ammonia.[1] From these humble beginnings, phenylalanine has proven to be an important molecule for many biological processes. Phenylalanine is an essential amino acid and cannot be synthesized by the human body, where it serves as either a direct precursor or an indirect substrate for the creation of many amino acids, neurotransmitters, and catecholamines, such as: tyrosine, dopamine, adrenaline, and epinephrine.[2-3] These hormones and neurotransmitters then regulate mood, alertness, and motor control. Phenylalanine has even been used as a treatment for Parkinson's disease, depression, and other psychological disorders.[4-6] On a biochemical level, phenylalanine can easily be converted into an amidated and acetylated form by common posttranslational modifications or other chemical reactions. This neutral form more easily traverses cell membranes, unlike polar and charged amino acids.[7] Physically, the electronic properties of the aromatic sidechain allow the behavior of phenylalanine to be examined through fluorescence anisotropy and other experimental methods.[8-9] These traits make phenylalanine an excellent molecule to study in a passive membrane permeation process.

The computational modeling of passive membrane processes can first be traced back to Marrink and Berendsen in 1994 and 1996 when they examined the permeation of water and other small molecules through a model 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine

(DPPC) membrane by using molecular mechanics.[10-12] These articles form the foundation for many later studies and data analyses. Since then, a wide variety of models and methods have been used to study membrane permeation.[13-15] Water and other small molecules are naturally an ideal system to study due to their ability to passively permeate a lipid bilayer.[10-11, 16-21] Quantifying and classifying permeability is also very important in pharmacological studies and drug development.[21-29] Recently, larger and more novel systems such as molecular dyes,[30-31] nanoparticles,[32-34] small proteins,[35-36] and transmembrane proteins have been studied.[37-38] However, only several of these journal articles pertain to amino acids[7, 39-43] – and only one contains references to phenylalanine.[44] Similarly, many research journal articles chronicle the properties of various lipid membrane systems, typically with the addition of cholesterol, [45-47] but only three contain discussions of the effects of these properties on the permeation of molecules[27, 42, 48] and only two contain discussions of the effect of system size.[43, 49] None of these journal articles contain discussions of the aromatic amino acids or contain discussions concerning the effect of system size and lipid type on rotational motion. Therefore, the effect of system size and lipid type is examined in greater detail for our phenylalanine system in the following pages. For further information and background reading, we direct the reader to a wide array of excellent review articles and guides that have been published in recent years.[14, 22, 47, 50-55]

As previously mentioned, we have discovered five different computational studies that examine the effects of either lipid type or system size on simple lipid bilayers. The earliest computational lipid variety study that we found was conducted by Sugii *et. al.* in 2005 and examined the permeation of H_20, O_2, CO, and NO through lipid bilayers with unsaturated, linear

acyl chains of carbon length 12, 14, and 16.[48] Longer lipids undergo slower rotational motion and demonstrate larger and broader free energy profiles, but do not statistically influence local diffusion constants. Permeability coefficients, however, decrease slightly with increasing acyl chain length. In 2014, Nademi *et. al.* studied the permeation of paracetamol, a small analgesic that shares many similarities with phenylalanine, across lipid bilayers with unsaturated, linear acyl chains with 14 and 16 carbon atoms.[27] They found no effect on the free energy profile and a similar trend in hydrogen bonding between paracetamol and water in the presence of both lipids. Finally, Bonhenry *et. al.* studied the permeation of neutral and cationic lysine through a branched lipid with either ether or ester linking groups and compared these results to a linear lipid with a single double bond.[42] They found that the free energy profile is shifted between the ether and ester lipids due to a greater distance between the ether linked headgroups. Neutral lysine traverses both lipids easily well; however, cationic lysine more easily traverses the ester, rather than the ether linked lipids. The free energy minimum in the linear lipid is shifted to the center of the bilayer with a smaller central maximum, as compared to the branched lipid, likely caused by the reduced stability of the unsaturated lipid bilayer.[42] In terms of system size, Hu *et. al.* examined the effect of 64 versus 288 lipid molecules with linear, unsaturated tails on the permeation of charged arginine. The free energy barrier of the larger system was reduced by several $kcalmol^{-1}$ due to broader and deeper membrane deformations, which also influenced the orientation of the arginine sidechain. Finally, Nietschke *et. al.* examined the effect of 18, 32, 50, 72, 98, or 128 lipid molecules with linear, monounsaturated acyl chains on a wide variety of parameters in 2016. For example, maximum differences of 5 $kJmol^{-1}$ at the free

energy maximum for methanol and maximum differences of 10 $kJmol^{-1}$ at the free energy minimum for ibuprofen were observed.

Our primary objective is to further our understanding of the permeation of the phenylalanine dipeptide -- and, by corollary, the other resonant and relatively nonpolar amino acids -- through model cell membranes. Molecular dynamics simulations were run on the passive permeation of the phenylalanine dipeptide by using umbrella sampling with lipid bilayers consisting of 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) lipid bilayers. As seen in Figure 3.1, DOPC contains acyl chains with 18 carbon atoms and one double bond near the center. POPC contains one acyl chain that is identical to DOPC; however, the second acyl chain is only 16 carbons long and does not contain a double bond. Simulations were run with 50 DOPC, 50 POPC, or 40 POPC lipid molecules in order to examine the effect of both lipid type and system size on various permeability measures, the potential of mean force, position dependent diffusion constants, and a wide variety of structural information.

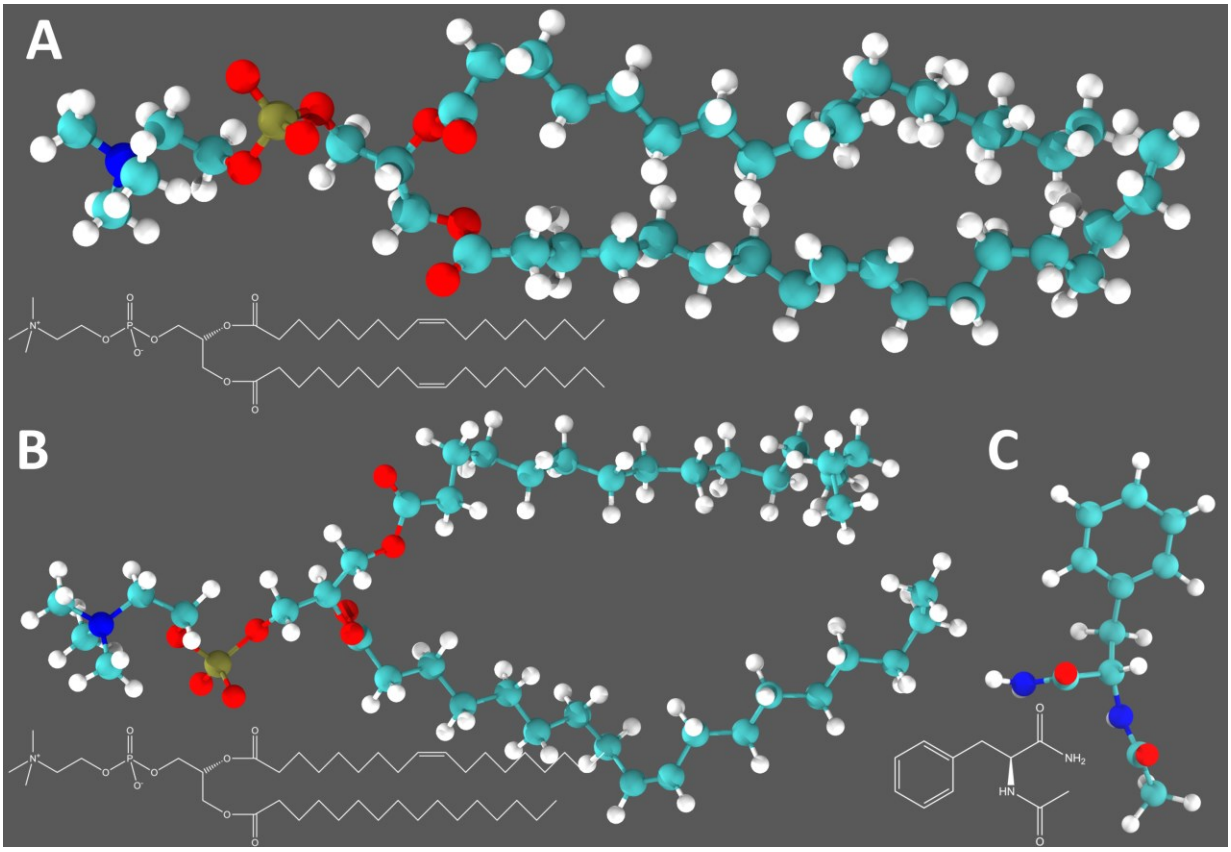


Figure 3.1. Molecules involved with this study are plotted above. 3.1A depicts DOPC with its associated line structure immediately beneath the artistic rendering. 3.1B depicts POPC, and 3.1C depicts the phenylalanine dipeptide, which is the same molecule used in the analyses in Chapter 2.

3.3. Methods:

The physical setup of our simulation window followed a procedure similar to our previous work, as seen in Chapter 2.[56] Our simulation windows consisted of one phenylalanine dipeptide, also known as N-acetylphenylalanineamide, Ac-Phe-NH2, or NAFA; water; and a phospholipid bilayer consisting of either 1,3-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) or 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC). Specifically, three systems were studied and were classified based upon the number and type of lipid molecules comprising the bilayer. The DOPC 50 system contained the following molecules: 1 phenylalanine dipeptide, 50

DOPC lipids, 2991 TIP3P waters, 8 sodium atoms, and 8 chlorine atoms, for a grand total of 15,764 atoms. This is the same system that was studied in Chapter 2 as 'NAFA.' The POPC 50 system contained 1 phenylalanine dipeptide, 50 POPC lipids, 2996 TIP3P waters, 6 sodium atoms, and 6 chlorine atoms, for a grand total of 15,578 atoms. Finally, the POPC 40 system contained 1 phenylalanine dipeptide, 40 POPC lipids, 2483 TIP3P waters, 7 sodium atoms, and 7 chlorine atoms, for a grand total of 15,578 atoms. Sodium and chlorine atoms were added to maintain a physiologically relevant ionic strength. All systems were confined to tetragonal boxes of dimensions: 4.29 × 4.29 × 8.30 nm for DOPC 50, 4.15 × 4.15 × 8.74 nm for POPC 50, 3.70 × 3.70 × 8.92 nm for POPC 40, corresponding to respective lipid headgroup areas of 0.736 nm^2, 0.689 nm^2, and 0.685 nm^2. For the DOPC 50 system, 0.736 nm^2 is in close agreement with an experimental average of 0.723 nm^2,[57-58] and the electron density profile is similar to one obtained from X-ray scattering experiments.[56, 58] Experimentally, POPC headgroup areas are 70.5 nm^2, which are close to our simulated values.[59] Unfortunately, experimental data for a POPC electron density profile does not exist; however, our data favorably compares to an experimentally determined DOPC electron density profile,[58] and a plot of this information (Figure 3.13) can be found in the Supplementary Information. The *x* and *y* axes of the simulation box are parallel to the plane of the lipid bilayer, and the *z* axis is defined as being perpendicular to the plane of the lipid bilayer. The initial coordinates for all systems, including the phospholipid bilayer, were determined by using CHARMM-GUI.[60-62]

Molecular dynamics simulations for the DOPC 50 system were carried out with GROMACS 4.5.4; simulations for the POPC 50 and POPC 40 systems were carried out with GROMACS 4.5.6.[63] The following options were chosen within GROMACS for all simulations.

Forces for DOPC, POPC, and the phenylalanine dipeptide were all represented by the version 36 CHARMM force field.[64-65] Water was represented by using the TIP3P model.[66] Direct electrostatic interactions were truncated at 0.13 nm and a long range-correction using the Particle Mesh Ewald method was used.[67] Van der Waals interactions were similarly truncated at 1.2 nm; however, a force switching function was used to smooth the transition between 1.0 and 1.2 nm. A canonical ensemble was used; and, therefore, the number of system particles, the system volume, and temperature were constant. Temperature was held constant at 300K by using velocity rescaling.[68] Initial velocities were sampled from a Maxwell distribution at 300K. The default leap-frog algorithm was used to integrate the relevant equations of motion with a time step of 2 fs. Periodic boundary conditions in the *x*, *y*, and *z* simulation box directions were used. Constraints on all bonds involving hydrogen atoms were enforced by using the LINCS algorithm.[69] Position, energy, and other simulation data were saved every 500 time steps and once every picosecond.

Simulations began by equilibrating each system, as setup by CHARMM-GUI, over 500 ps intervals with the parameters described in the previous paragraph. Once equilibrated, an unrestrained simulation was run for at least 50 ns, and a system image was obtained at z = 1.6 nm. This image was then used as the starting point for further umbrella sampling windows in both positive and negative z directions. Umbrella sampling windows for all systems were run from z = 0.0 nm to z = 3.0 nm in 0.1 nm increments. A harmonic restraining potential of 3000 kJ $mol^{-1}nm^{-1}$, located at the center of each umbrella sampling window, was applied to the center of mass of the phenylalanine dipeptide.[70-71] Each window ran from 50 to 100 ns; as needed, simulation windows near the center of the membrane were extended up to 100 ns in order to

improve statistical overlap for potential of mean force calculations. The potential of mean force (PMF) was then calculated by using the weighted histogram analysis method (WHAM).[72-74]

Diffusion constant data were obtained by numerically integrating the Smoluchowski equation. Our approach uses the method described by Hummer[75] that is, in turn, based on ideas developed by Zusman[76] and Bicout and Szabo.[77] Specifically,

$$D_{n+1/2} = w_{n+1,n}\left[\frac{p^*(n)}{p^*(n+1)}\right]^{1/2} d^2 \qquad (3.1)$$

where *n* is the bin number, $w_{n+1,n}$ is the transition rate between neighboring bins, *p** is the biased probability distribution, and *d* is the width of each bin. This approach begins by binning *z(t)*, the distance between the center of mass of the phenylalanine dipeptide and the center of mass of the lipid bilayer, over the course of the entire umbrella sampling simulation window. The biased probability distribution, *p*(z)*, is then approximated by the resulting, normalized histogram as *p*(n)*. The average number of transitions between neighboring bins is then calculated from the position time series *z(t)*. Next, $w_{n+1,n}$ is calculated from the average number of transitions between neighboring bins divided by the bin residence time. We chose a bin width of 0.025 nm so that a bin transition is located at the center of any given umbrella window and to minimize error. Under these parameters, the position dependent translational diffusion coefficient, *D(z)*, for each of our umbrella windows is then found when *n* = 1. Please note: although we are using our harmonically restrained umbrella sampling windows, the diffusion coefficient is calculated at the center of each window where the restraining potential is equal to zero, thereby minimizing the impact of the harmonic umbrella sampling restraint.

Permeation kinetics were then assessed by calculating the permeability coefficient, *P*, and the mean first passage time (MFPT), $\langle\tau\rangle$, by using the inhomogeneous solubility diffusion model.[10-11, 39, 78] The permeability coefficient is calculated by:

$$P = 1/\int_a^b e^{\beta w(z)} D(z)^{-1} dz \tag{3.2}$$

where *a* is a location in the aqueous region on one side of the membrane along the *z* axis, *b* is a location in the aqueous region on the opposite side of the membrane, $\beta = (k_B T)^{-1}$, *w(z)* is the potential of mean force at a given location along the *z*-axis, and *D(z)* is the diffusion constant at a given value of *z*. The mean first passage time is calculated by:[79-80]

$$\langle\tau\rangle = \int_a^b \left[e^{\beta w(z)} D(z)^{-1} \int_a^z e^{-\beta w(z')} dz' \right] dz \tag{3.3}$$

where all variables are defined in the same manner as for the calculation of P.

A wide variety of additional and smaller data analyses were also conducted. The timescale of molecular motion was investigated at select distances of *z* = 0.0, 0.5, 1.0, 1.5, 2.0 nm by extending the corresponding umbrella windows to 250 ns. Sidechain anisotropy was investigated by defining a vector in the plane of the phenyl group in the phenylalanine dipeptide that is perpendicular to a vector formed by the CG and CZ carbon atoms. A C_2 autocorrelation function for this vector was then calculated. C_2 is a second order associated Legendre polynomial, defined below, and is related to simple rotational diffusion in 3D space.

$$C_2(t) = \frac{1}{2}\langle(3cos^2\theta(t) - 1)\rangle \tag{3.4}$$

where $\theta(t)$ is the angle formed between the investigated vector at two different points in time. The rotational correlation time, τ_{rot}, was then determined by fitting the C_2 autocorrelation function to a double exponential decay function. In a similar manner, the autocorrelation time for the projection of the insertion vector along the *z*-axis was determined, as well. Backbone and sidechain conformational angles, radial distribution functions, solvent accessible surface areas, and the radius of gyration were calculated by using the relevant GROMACS 4.5.6 tools.[63] The dihedral angle clustering analysis was done with CHARMM.[81]

3.4. Results and Discussion

3.4.1. - Potential of Mean Force.

The potential of mean force represents the free energy required to reversibly pull the dipeptide along the *z*-axis and is relative to the free energy in the aqueous region. The potential of mean force in the aqueous region at 3.0 nm for each system was defined as 0.0 $kJmol^{-1}$ and all other values are relative to this reference. The phenylalanine dipeptide exhibits very similar behavior in all three systems, as seen in Figure 3.1: a free energy minimum around the lipid-water interface and a free energy barrier at the center of the membrane. The minima are -16, -13, and -18 $kJmol^{-1}$, respectively, for the POPC 40, POPC 50, and DOPC 50 systems at 1.3, 1.2, and 1.2 nm. The free energy barriers are 14, 12, and 10 $kJmol^{-1}$, in the same order as before. Qualitatively, these results are very similar to findings for other amino acids, and agree with our previous studies concerning aromatic dipeptides.[39, 44, 56] Blocked phenylalanine contains a hydrophilic backbone and a hydrophobic side chain which results in a smaller free

energy near the membrane interface where the backbone can interact with the aqueous region at larger *z*-values and where the sidechain can interact with the lipid environment at smaller *z*-values.

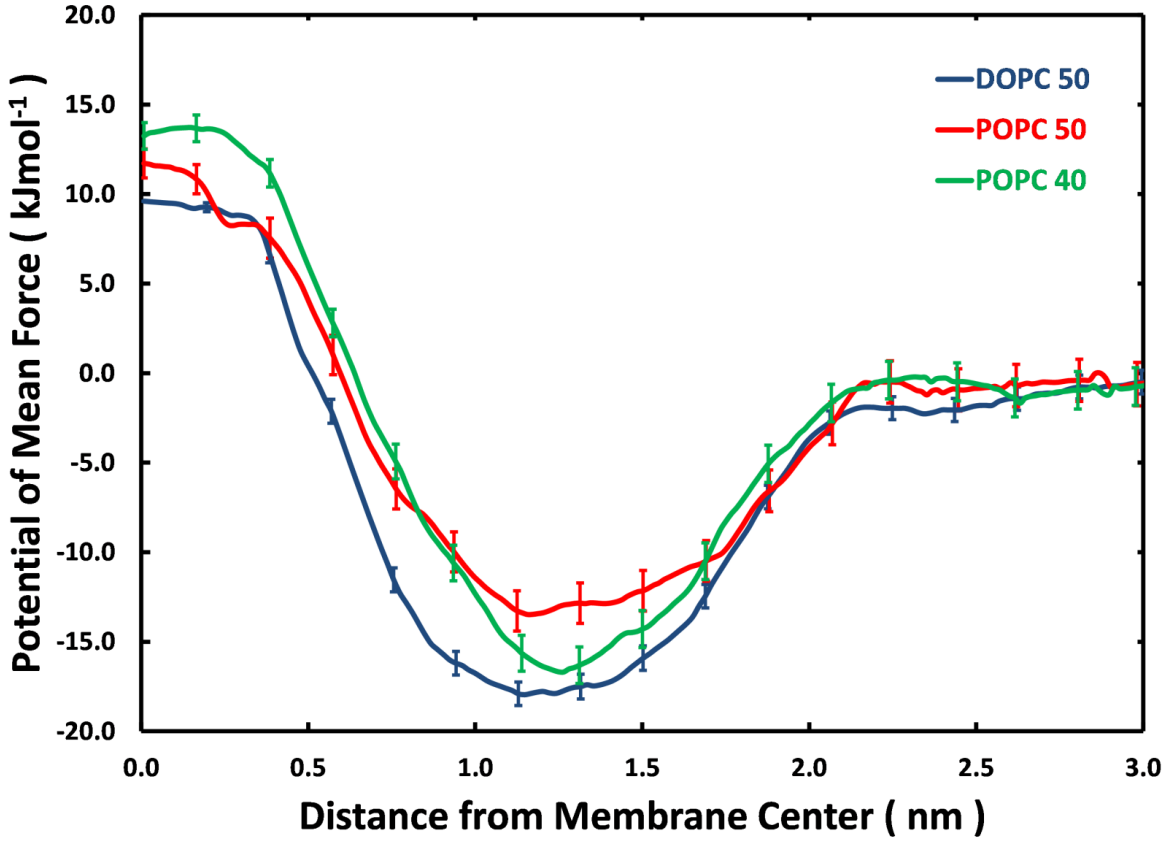


Figure 3.2. The free energy required to pull the phenylalanine dipeptide along the *z*-axis of the simulation box is plotted below as the potential of mean force. The center of the lipid bilayer is represented by $z = 0$. A 95% confidence interval representing the standard error is also plotted for select points as determined by the Bootstrap method.[82] Alternate error estimates are described in the Supplementary Information (Figures 3.14 to 3.16.)

The PMFs for POPC 40 and POPC 50 demonstrated little difference at the center of the membrane; however, POPC 40 demonstrates a small, but statistically significant difference at the interface. Although the effect of lipid bilayer size on aromatic dipeptide PMFs has not been

examined before, Nitschke *et. al.* have conducted membrane size studies on methanol and ibuprofen with POPC bilayers.[49] Even though ibuprofen is less polar and is, therefore, more likely to pass through a lipid bilayer[83], it still contains a similar combination of polar and nonpolar moieties and produces the same qualitative PMF shape as the phenylalanine dipeptide. In Nitschke's analysis, lipid size may influence the PMF by three key mechanisms: 1.) averaging out the effects of membrane fluctuations, 2.) finite size effects altering the lipid structure, and 3.) solute stabilization through the interaction of periodic images. They determined that only membrane fluctuations influence the PMF for lipid bilayers as small as 32 molecules. For example, the central barrier for ibuprofen is lowered by 3 $kJmol^{-1}$ when the system size is increased from 32 to 50 lipids. In our case, the PMF is lowered 2 $kJmol^{-1}$ when the system size is increased from 40 to 50 lipid molecules, in good agreement with Nitschke's results. In the interfacial region, the PMF is lowered by 5 $kJmol^{-1}$ for ibuprofen and by 3 $kJmol^{-1}$ for the phenylalanine dipeptide. Interestingly, these system size effects can be removed by utilizing a cylinder based center of mass coordinate, rather than the all lipid center of mass coordinate that we use.[83-84] Alternatively, the small variation of the PMF due to system size may be due to the relatively short simulation length: 20 ns per window for Nitschke and 50 ns per window for our results. The passive transport of a peptide through a lipid bilayer heavily depends on membrane fluctuations at the interface, which may not be adequately sampled in short simulations of small numbers of lipids, producing a systematic bias towards smaller, less negative, interfacial energies.[84-86] For example, lipid molecules can flip from one side to another on the scale of hundreds of nanoseconds[87]; lipid defects and void spaces form in the center of

the membrane on the order of tens of microseconds[85]; and, rotational energy barriers can further complicate matters by occurring on the order of tens of nanoseconds.[86, 88]

The DOPC 50 PMF exhibits a significantly smaller central barrier and a broader interfacial minimum than the POPC 50 curve, with particularly large deviations within the region between 0.6 and 1.0 nm, as shown in Figure 3.1. Traditionally, the width of the PMF near the interfacial minimum has been associated with the ability of the lipid membrane to either exclude polar solutes or to enfold apolar solutes due to membrane fluctuations.[49, 89-90] Because our simulations use the same solute for each lipid bilayer system, the calculated PMF differences must be due to differences in bilayer properties. DOPC contains an additional double bond in its aliphatic tail as compared to POPC. This greater lack of saturation causes the DOPC bilayer to be less ordered than POPC and other saturated phospholipids.[91] Our results suggest that the less ordered DOPC bilayer is more easily capable of undergoing distortions that better accommodate the amphiphilic nature of the phenylalanine dipeptide, thereby producing a broader PMF curve near the interfacial free energy minimum. These membrane distortions either allow the hydrophilic moiety to remain in contact with the bulk aqueous region[85-87, 92] or allow the partial retention of the solvation shell at a greater insertion depth, thereby lowering the free energy. These effects are discussed in greater detail in some of the following sections.

3.4.2. - Translational Diffusion.

The translational diffusion in all three systems closely follows our previous results for aromatic amino acids in DOPC 50, as seen in Figure 3.2.[56] In the aqueous region, the

phenylalanine dipeptide exhibits diffusion constants of $5.20 \pm 0.10 \times 10^{-10}$ m^2s^{-1} in DOPC 50, $5.40 \pm 0.15 \times 10^{-10}$ m^2s^{-1} in POPC 50, and $5.23 \pm 0.15 \times 10^{-10}$ m^2s^{-1} in POPC 40. All of these values possess overlapping confidence intervals, and any differences are statistically insignificant. At the membrane interface at z = 1.5 nm, diffusion constants of $4.82 \pm 0.17 \times 10^{-10}$ m^2s^{-1} in DOPC 50, $4.86 \pm 0.089 \times 10^{-10}$ m^2s^{-1} in POPC 50, and $4.83 \pm 0.15 \times 10^{-10}$ m^2s^{-1} in POPC 40 were determined. Again, these values are not statistically different from each other. However, they are slightly smaller than the values determined within the aqueous region. Finally, at the center of the membrane, the phenylalanine dipeptide diffusion constants are $5.63 \pm 0.26 \times 10^{-10}$ m^2s^{-1} for DOPC 50, $5.46 \pm 0.17 \times 10^{-10}$ m^2s^{-1} for POPC 50, and $5.48 \pm 0.15 \times 10^{-10}$ m^2s^{-1} for POPC 40. Here, as well, the results for the three systems are not statistically different from each other, and they are similar to the values within the aqueous region. This pattern of slightly smaller diffusion constants near the membrane interface as compared to either the membrane interior or aqueous regions has also been witnessed in many different molecules from water to organic solvents to small pharmaceutical compounds.[11, 19-20, 26, 29, 89] The observed differences are most pronounced for polar and amphiphilic molecules and negligible for larger, more nonpolar molecules.[21, 23-24, 54] While a majority of the *D(z)* values follow relatively smooth variation with insertion depth *z*, there are two data points with significant deviations, for POPC 40 at 1.6 nm and DOPC 50 at 1.7 nm. These outliers may represent the sampling of a rare event, such as one of the rotational motions explored in Chapter 2. In summary, switching from DOPC to POPC or from 50 to 40 lipid molecules does not have a measurable impact on the position-dependent translational diffusion constants.

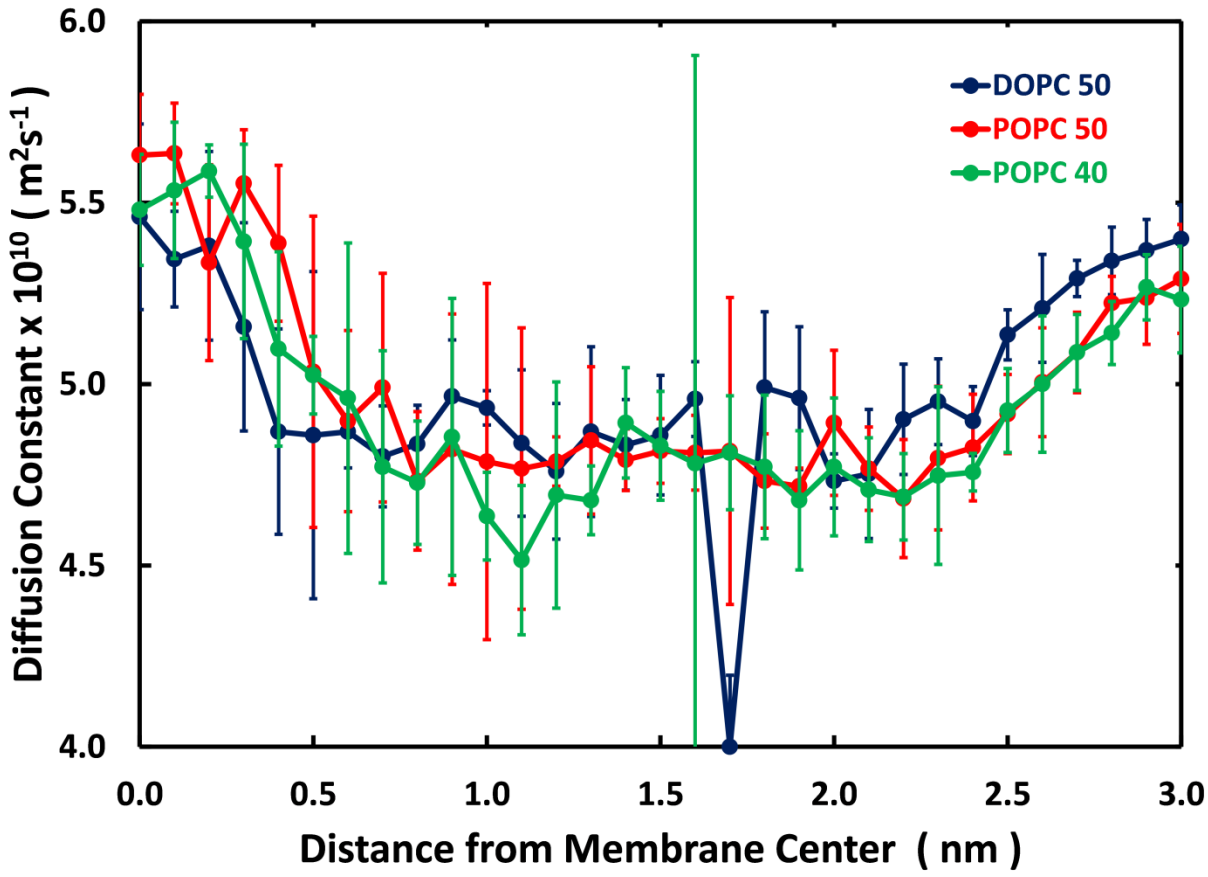


Figure 3.3. Translational diffusion constants of the dipeptide with respect to the center of mass of the lipid bilayer are depicted. These constants were determined by a numerical solution of the Smoluchowski equation. The error bars represent the standard error at a 95% confidence level, as calculated by separating the data for each simulation window up into four consecutive bins.

3.4.3. - Permeability Measures.

The permeation coefficient -- an important quantity used in pharmacokinetic analyses -- and the mean first passage time (MFPT) are discussed in this section. Both quantities represent a more physically relevant interpretation of our diffusion constant and potential of mean force data. The MFPT for the phenylalanine dipeptide through a POPC 50, POPC 40, and DOPC 50

membrane system are, respectively: $9.1 \times 10^{-6} \pm 5.0 \times 10^{-7}$ s, $1.2 \times 10^{-6} \pm 8.4 \times 10^{-7}$ s, and $4.1 \times 10^{-6} \pm 2.7 \times 10^{-7}$s by using Eq. 3.3. The given 95% confidence intervals were calculated by propagating error from the diffusion and potential of mean force data displayed in Figures 3.2 and 3.3. All of the passage times were on the order of microseconds, which is in agreement with similar computational studies on molecules around the same size as the phenylalanine dipeptide.[39, 80] Of the three systems studied, POPC 40 has the largest passage time, POPC 50 the smallest, and DOPC 50 in between the two. These calculations are most sensitive to the permeant environment near the center of the membrane and therefore reflect the potential of mean force and diffusion constant around z = 0 nm. The diffusion constants near the center of the membrane are largely the same for all three systems; however, the PMF is greatest for POPC 40, followed by POPC 50, and then DOPC 50, which matches the qualitative order of the MFPT. The MFPT is a function of the PMF and the diffusion coefficients; however, the diffusion coefficients are nearly identical, indicating that the MFPT should follow trends in the PMF, as seen above. It is interesting to note that the width and depth of the PMF did not have a substantial impact on this ordering because the PMF for DOPC 50 is much deeper and broader than for POPC 50.

Our calculated permeation coefficients for the POPC 50, POPC 40, and DOPC 50 systems are, respectively: 0.82 ± 0.03 cms^{-1}, 0.24 ± 0.003 cms^{-1}, and 0.66 ± 0.03 cms^{-1} by using Eq. 3.2. Again, the reported uncertainties are 95% confidence intervals determined through the propagation of error in the PMF and diffusion constant data. It should also be noted that these calculations are highly sensitive to the choice of the free energy in the aqueous region; for example, differences of only a few kJmol^{-1} resulted in order of magnitude differences in the

permeation constant. Therefore, actual uncertainties are likely far larger and care should be taken in terms of converging the aqueous PMF. The MFPT calculations are not sensitive to this variation and may be a more accurate reflection of the permeation kinetics. As a result, the relatively small differences in the permeation coefficients for each system are not significant. In general, these permeation coefficients are in agreement with other simulations with small molecules.[19] We previously reported experimental coefficients for the phenylalanine dipeptide in DOPC 50 of 5.6×10^{-8} cms^{-1},[56] as compared to our computational value in this work of 0.66 cms^{-1}. As expected, the computational value is far larger than the experimental result due to any of the following factors which have been well documented in the literature: the use of the inhomogenous solubility-diffusion model,[7, 24, 39, 56, 88, 93-95] hidden rotational barriers,[7, 39, 88-89, 93-96] membrane and solvent fluctuations,[35, 96-98] and the choice of force field.[56, 64] We have discussed each of these effects in greater detail within our earlier work and within Chapter 2.[56]

3.4.4. - Rotational Sidechain Diffusion.

The rotation of the phenylalanine dipeptide sidechain demonstrates a dependence on insertion depth, as seen in Figure 3.4. Please note that many autocorrelation functions did not decay to zero within our extended trajectories of 250 ns. Therefore, reported numerical time constants depicted in Figure 3.4 that are beyond this length should be treated in a purely qualitative manner. However, there is clearly a position dependence unlike the largely flat translational diffusion profile in Figure 3.3. Specifically, rotational correlation times for the POPC 40 system vary from 24 ns at z = 0.0 nm, 160 ns at 1.5 nm, to 190 ns at 2.0 nm. Values at

z = 0.5 nm and *z* = 1.0 nm are either undefined or unreliable. Rotational correlation times for the POPC 50 system vary from 75 ns at *z* = 0.0 nm, 150 ns at 1.5 nm, to 1.8 ns at 2.0 nm. Again, values at *z* = 0.5 nm and *z* = 1.0 nm are either undefined or unreliable. Correlation times in the center of the membrane (*z* = 0.0 nm) and at the interface (*z* = 1.5 and 2.0 nm) were determined and are likely smaller than in the interfacial region. All of these values are substantially larger than previously reported computational and experimental values for the phenylalanine dipeptide in water, respectively: 20 ps with the CHARMM force field and TIP3P water[8] and 98 ± 30 ps at 278 K.[8] However, the correlation time when the phenylalanine dipeptide is located at *z* = 3.0 is 70 ps for POPC 50 and 64 ps for POPC 40, both of which are close to the previous experimental and computational values. This indicates that the rotational motion of the sidechain plays an important role in the permeation process.

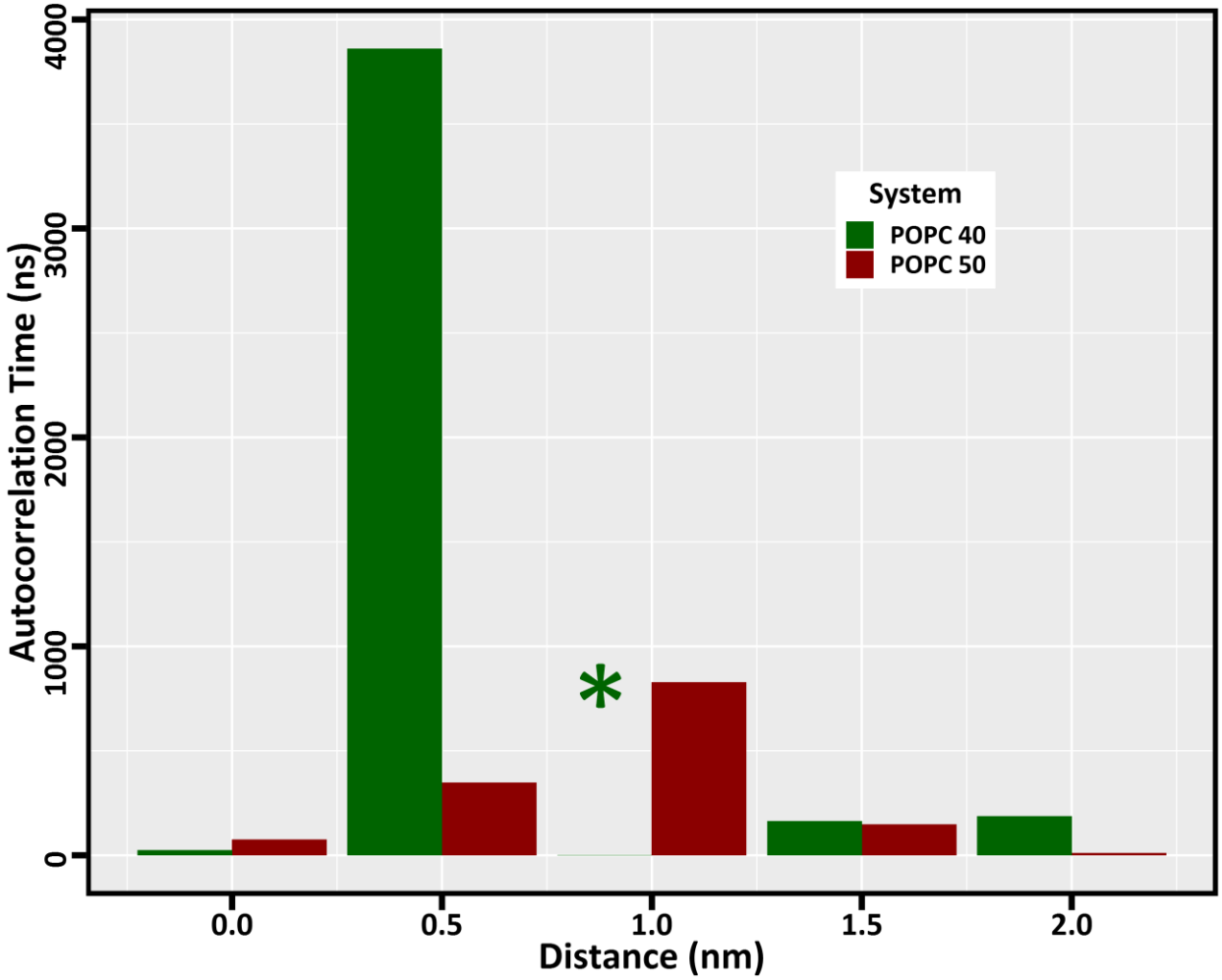


Figure 3.4. The autocorrelation time for the rotation of the phenylalanine sidechain is plotted for both the POPC 40 and POPC 50 systems. Note: * indicates that an autocorrelation time could not be determined through the following fitting process. Autocorrelation functions were fitted to an exponential decay function over a 250 ns trajectory at insertion distances of 0.0, 0.5, 1.0, 1.5, and 2.0 nm. The C_2 correlation function is defined as: $C_2(t) = \frac{1}{2}\langle(3cos^2\theta - 1)\rangle$, with the reorientation angle, θ, defined as the angle between a vector perpendicular to the CG and CZ carbon atoms of the sidechain phenyl group and the simulation *z*-axis. Note that many autocorrelation functions did not decay within the length of the simulation, and the reported, numerical results beyond 250 ns should not be quantitatively compared.

3.4.5. - Insertion Angle.

The insertion vector of each peptide is defined by an angle, θ, between two vectors: one vector points along the positive *z*-axis with respect to the lipid membrane, and the other vector starts from the center of mass of the backbone and ends at the center of mass of the sidechain. In other words, an angle of 0° indicates that the sidechain is pointing in the positive *z*-direction and the backbone in the negative *z*-direction. Alternatively, an angle of 180° represents the reverse: the peptide backbone is pointing in the positive *z*-direction, and the sidechain is pointing in the negative *z*-direction. From Figure 3.5, peptide orientations are largely homogenous in the aqueous region beyond 2.0 nm. However, as the peptide nears the membrane interface, the insertion angle prefers large values between 120° and 180°, indicating a tendency for the sidechain to enter the membrane first. The sidechain, particularly for phenylalanine, is largely nonpolar and interacts with the non-polar aliphatic groups within the lipid membrane, while the polar amide and acetate groups within the peptide backbone prefer interactions with the solvent. This preference continues until around 0.3 to 0.4 nm from the center of the membrane, where the insertion angle can, once again, adopt a wide range of different values. At the center of the membrane, the local environment is similar in both the positive or negative *z*-direction, and a preference for one angle or direction is unlikely to be seen in this homogenous region. The ability of the phenylalanine dipeptide to change its orientation in either the positive or negative *z*-direction is very similar to the results found in previous studies of aromatic dipeptides.[39, 56, 88] Both lipid types, POPC and DOPC, and system sizes, POPC 40 and POPC 50, exhibit very similar results.

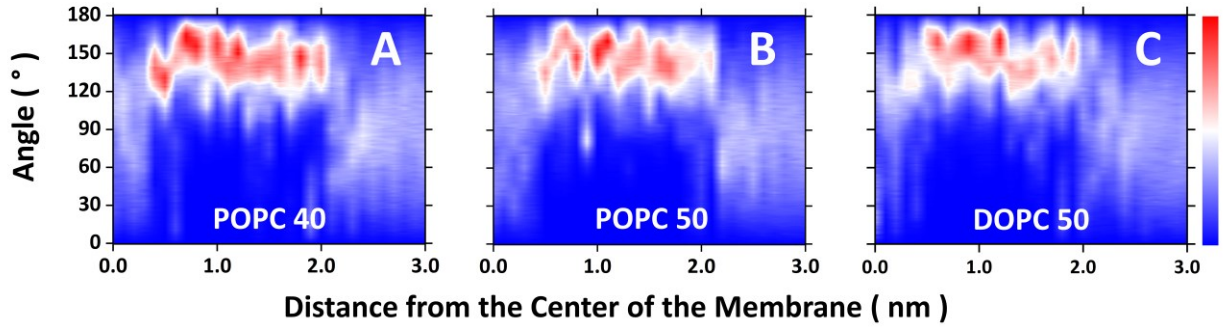


Figure 3.5. The insertion angle of the dipeptide relative to the distance from the center of the lipid membrane can be seen below. The backbone center of mass to the sidechain center of mass defines the insertion vector. The insertion angle, θ, is then the angle inbetween the insertion vector and the positive *z*-axis. Blue regions represent areas of low, normalized probability, and red regions represent areas of high, normalized probability as a function of θ and the distance from membrane center, *z*.

The timescale of the reorientation of the insertion vector was also examined by calculating the autocorrelation function for the projection of this vector along the *z*-axis, also known as the cosine of the insertion angle. Unfortunately, this analysis suffers from many of the same problems as the rotational sidechain analysis. Most of the autocorrelation functions did not decay to zero, and the fitted time constants were much larger than the 250 ns trajectories. These time constants are plotted in Figure 3.6 below in a purely qualitative manner to indicate that there is some relationship between the reorientation of the insertion vector and the position of the permeant. However, in order to determine what that relationship is, longer simulation trajectories need to be run and analyzed.

We believe that this relationship might increase in the interfacial region before then decreasing in the center of the membrane and within the aqueous regions. Longer potential timescales are likely due to the large scale fluctuations of the lipid molecules within the membrane, which are absent at the center of the membrane.[85] Conversely, the smaller

autocorrelation times at the center are due to the creation of transient void spaces as a result of the greater disorder of the lipid tails.[23, 54, 92] These void spaces then allow the phenylalanine dipeptide to rotate with greater freedom.

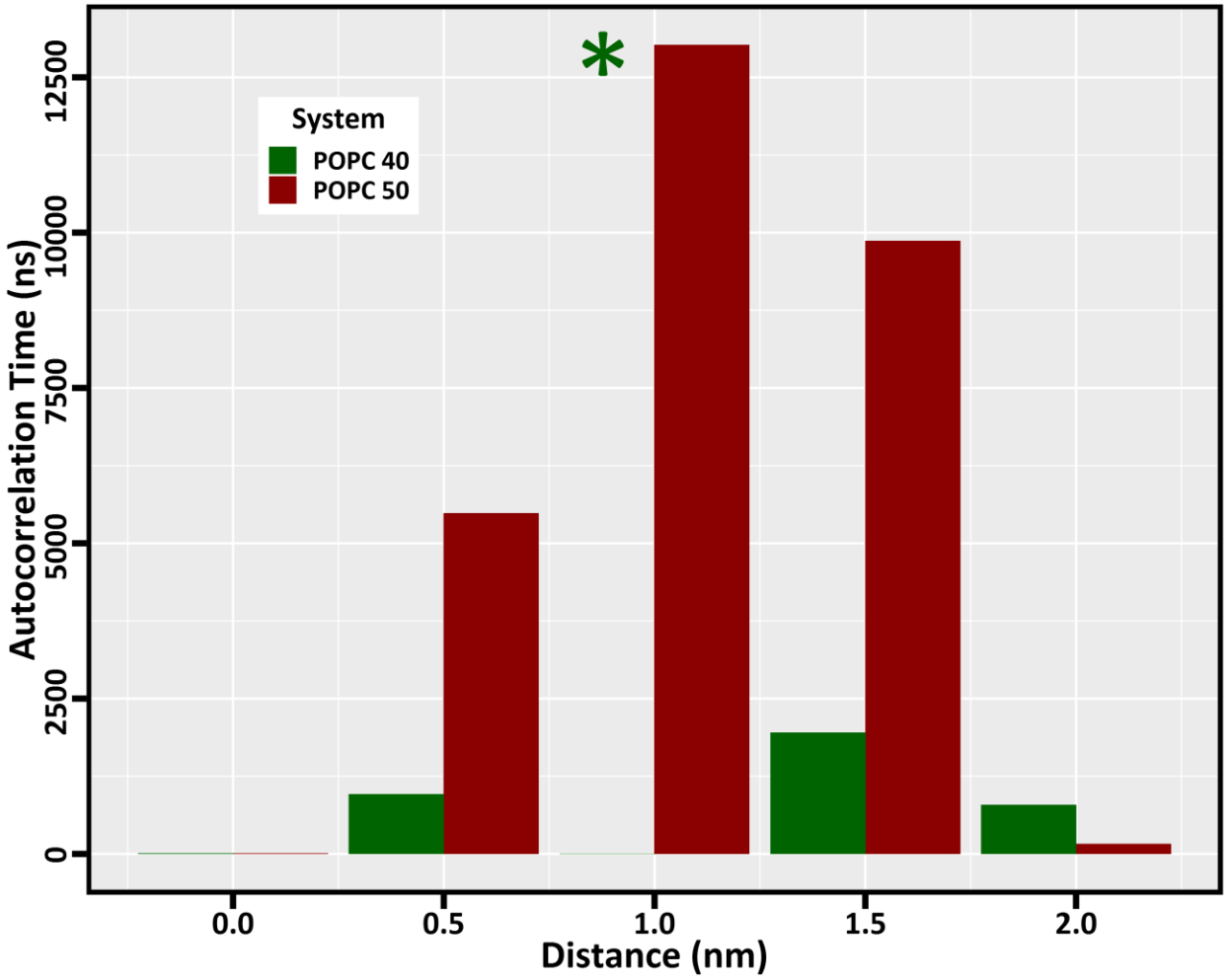


Figure 3.6. The correlation times for the reorientation of the insertion angle vector are plotted here. Note: * indicates that an autocorrelation time could not be determined through the following fitting process. Longer times indicate a slower rotational motion; whereas, shorter times indicate a faster rotational motion. The correlation times were determined by fitting the autocorrelation function for the projection of the insertion vector along the *z*-axis to an exponential decay function of the form: $C(t) = ae^{bt} + (1-a)e^{ct}$, where *C(t)* is the autocorrelation function, and *a*, *b*, and *c* are parameters determined by the fitting routine. The correlation times reported in this figure are equal to the inverse of parameter '*c*'. Note that these times exceed the length of the simulation, 250 ns, and numerical values beyond this point should not be treated in a quantitative manner.

3.4.6. - Backbone Conformations.

Figure 3.7 contains Ramachandran plots for each of the systems studied.[99-100] Simulations were split up into three regions: the water region at the far edge of the simulation box, the interface where the lipid headgroups interact with the aqueous region, and the center of the membrane where the lipid tails interact, as seen in 3.7A. 3.7B-D contain plots of all three slices for the POPC 40, POPC 50, and DOPC 50 systems. The sample structures around the Ramachandran plot are depictions of the two most prevalent conformations, representing around 80% or greater of all observed clusters. In the aqueous region, all three systems sample three conformational areas: the α-helical region with negative Φ and Ψ angles, the β-sheet area with negative Φ and positive Ψ angles in the top corner, and the small region sandwiched in-between the α-helices and β-sheets, representing a C_{7eq} conformation, a free energy minimum that is commonly found in dipeptides.[101] These findings are similar to what we found for the other aromatic dipeptides of tyrosine and tryptophan in the aqueous region.[56] At the interface, the α-helix and β-sheet regions lose probability density to the C_{7eq} conformation until, eventually, at the center of the membrane, only the C_{7eq} conformation remains to a significant degree. This demonstrates that the phenylalanine dipeptide adopts different conformations as it travels through the membrane. This behavior is consistent regardless of system size or lipid type, as explored in greater detail throughout the clustering analysis in the next section.

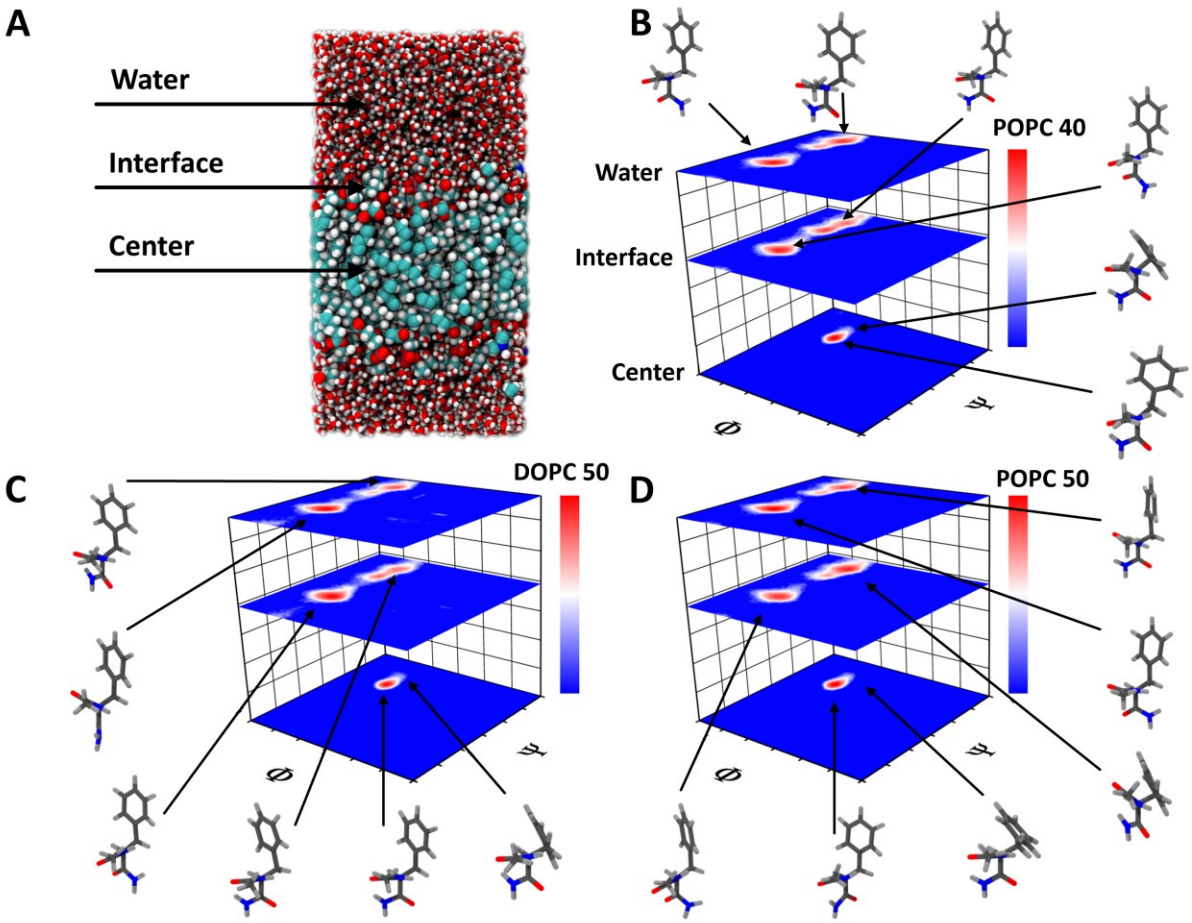
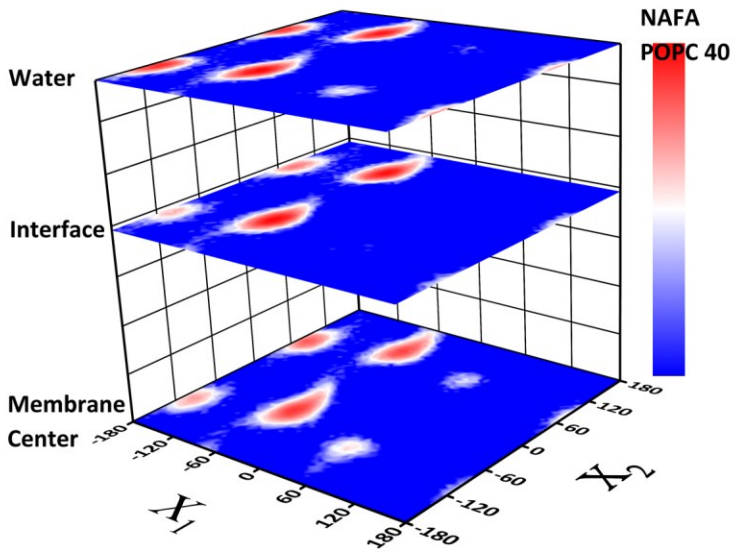


Figure 3.7. Ramachandran plots and representative structures from a clustering analysis of the phenylalanine dipeptide in lipid bilayers are depicted. The Ramachandran plots are slices taken at the center, middle, and edge of the simulation box, as shown in 3.7A. 3.7B-D show the actual Ramachandran plots with red representing areas with a high probability density: (B) DOPC 50, (C) POPC40, (D) POPC50. Blue regions exhibit either zero or little likelihood of the dipeptide adopting such a conformation. The surrounding structures display the two most prevalent conformations in each region, as determined by the clustering analysis. In the Ramachandran plots, the bottom left regions of high (red) probability represent α-helical conformations, the top left red regions represent β-sheets and the regions in between the two, as seen in the center of the membrane, indicate C_{7eq} conformations. Phi and psi angles are plotted from -180° to 180°.

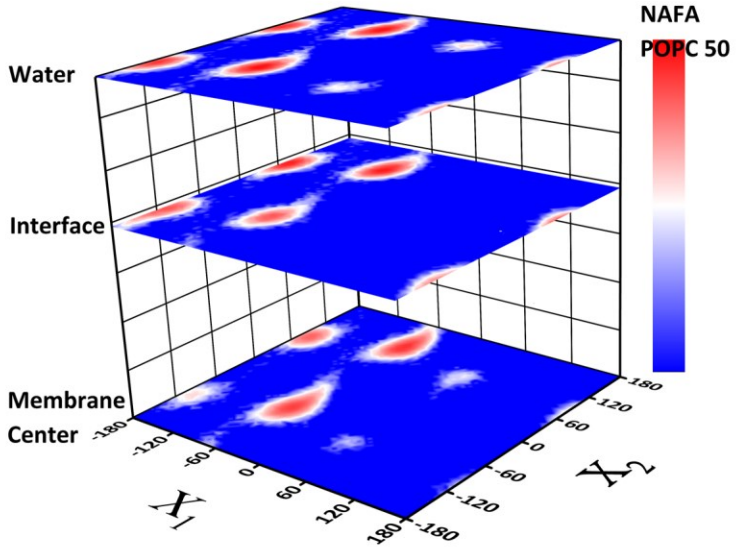
3.4.7. - Sidechain Conformations.

The two sidechain dihedral angles, X_1 and X_2, were plotted in Figure 3.8 in the same manner as the Ramachandran plots in Figure 3.7. The following nomenclature conforms to the standard convention for the naming of amino acid sidechain conformations. For example, tg^+ refers to a *trans* X_1 dihedral angle around 180° and a *gauche(+)* X_2 dihedral angle around +60°. By looking at the top X_1- X_2 plot in 3.8A for the aqueous region, the phenyl ring of the phenylalanine dipeptide in DOPC 50 samples six different conformers: tg^+, g^-g^+, and g^+g^+ in the top row; and tg^-, g^-g^-, and g^+g^- in the bottom row. However, only tg^+, g^-g^+, tg^-, and g^-g^- are heavily sampled, as indicated by the large, red regions. g^+g^+ and g^+g^- are only weakly sampled due to the much smaller, less intense, and whiter regions on the right-hand side of the panel. In the interfacial region of 3.8A, the four heavily sampled conformations are still present; however, g^+g^+ and g^+g^- vanish completely before reappearing within the center of the membrane. This phenomenon was also witnessed in POPC 40 and POPC 50, as seen in 3.8B and 3.8C. Although the dihedral sidechain angles do not play as important a role in membrane permeation as seen in the Ramachandran plots of Figure 3.7, there is a slight conformational preference in the interfacial region against g^+g^+ and g^+g^-. In terms of membrane size, the g^+g^+ conformer is largely absent in the aqueous region of the POPC 40 system and slightly more prevalent in POPC 50. Finally, there is little difference between the DOPC 50 and POPC 40 systems.

A


NAFA
POPC 40
Water
Interface
Membrane
Center
-180
-120
-60
0
60
120
180
X1
X2

B


NAFA
POPC 50
Water
Interface
Membrane
Center
-180
-120
-60
0
60
120
180
X1
X2

C

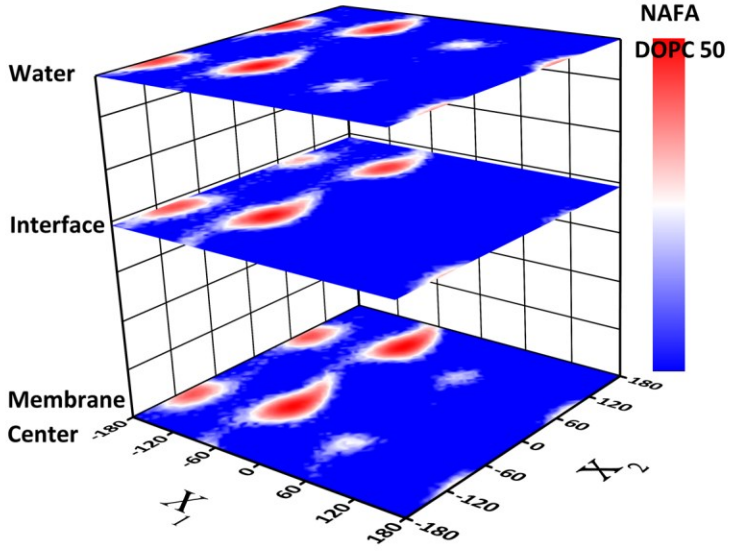

NAFA
DOPC 50
Water
Interface
Membrane
Center
-180
-120
-60
0
60
120
180
X1
X2

Figure 3.8. X_1 and X_2 sidechain angles for the phenylalanine dipeptide are displayed in the same manner as the previously examined Ramachandran plots. Specifically, 3.8A represents three such plots at the center, interfacial, and aqueous regions, in ascending order, for the simulations of the phenylalanine dipeptide with 50 DOPC lipids. Blue represents regions where the binned probability distribution is very small or zero; whereas, red indicates a large likelihood that the sidechain conformations will adopt that particular conformation. 3.8B is the same as 3.8A, except a membrane consisting of 40 POPC lipids was used; and, finally, 3.8C represents the same type of plot where 50 POPC lipids were used within the membrane.

3.4.8. - Peptide Structures and Clustering.

A clustering analysis of the Φ, Ψ, X_1, and X_2 dihedral angles was conducted for structures at the center of the membrane, interface, and water regions. As we previously reported,[56] the phenylalanine dipeptide in a 50 DOPC membrane system adopts seven clusters in the water region: 4 α and 3 β; four at the interface: 2 α and 2 C_{7eq}-β hybrids; and four at the center of the membrane: 2 C_{7eq} and 2 β. Please note that clusters were classified based on those identified by Tobias and Brooks to within a tolerance of around 10° in the Φ dihedral angle and around 20° in the Ψ angle.[101] Also, the two most prevalent structures within each clustering region are depicted in the margins of panels B,C, and D within Figure 3.7. In the 50 POPC lipid membrane system, seven clusters were identified in the water region: 3 α and 4 β; four at the interface: 2 α and 2 β; and five at the center: 2 C_{7eq}, 2 β, and 1 C_{7eq}-β hybrid. In the 40 POPC lipid membrane system, ten clusters were identified in the water region: 5 α and 5 β; four at the interface: 2 α and 2 β; and five at the center: 2 C_{7eq}, 2 β, and 1 distorted C_{7eq} cluster. The larger number of clusters in the water region for this system was surprising; however, all of the additional clusters beyond the original seven identified for the DOPC 50 and POPC 50 systems

represented fewer than 0.5% of the total cluster representation. More interestingly, the interfacial region of the POPC systems clearly had only α and β clusters; however, the DOPC interfacial region contains two β clusters whose Ψ angles are halfway between a typical β or C_{7eq} cluster, hence their designation as C_{7eq}-β hybrids. These hybrid clusters may indicate that the phenylalanine dipeptide is being forced into lower energy conformations sooner.

Indeed, the oxygen to nitrogen atom distances on the terminal blocking groups, as plotted in Figure 3.9, indicate that these two atoms are coming closer together to form a lower energy C_{7eq} cluster at a *z*-distance of 0.7 nm, whereas the POPC systems do not begin this process until around 0.5 or 0.6 nm. In the water region, all systems adopt an oxygen to nitrogen distance between 3 to 5 nm and do not exhibit a strong preference. However, at the center of the membrane, all lipids and sizes clearly indicate a strong preference for a very small oxygen and nitrogen distance around 2.8 to 3.1 nm as the polar dipeptide backbone attempts to either minimize its interfacial free-energy with respect to the non-polar environment at the membrane center or to potentially maximize their dipole or hydrogen bonding interactions.

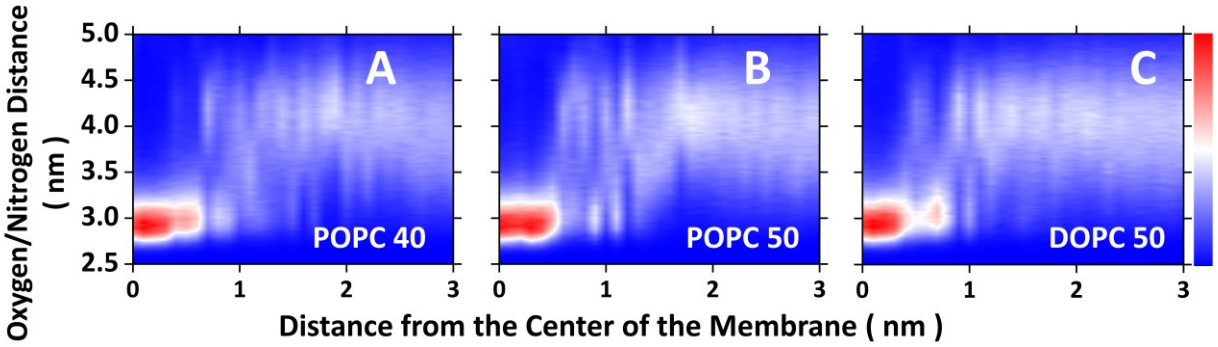


Figure 3.9. The distance between the oxygen and nitrogen atoms of the respective blocking groups – the oxygen atom of the acetylated N-terminus and the nitrogen atom of the amidated C-terminus – are plotted for each simulation window in terms of their associated probability distributions. Blue represents a small or zero likelihood of the phenylalanine dipeptide adopting a conformation with that particular oxygen/nitrogen distance, whereas red indicates a high likelihood. 3.9A depicts the oxygen/nitrogen distance for DOPC 50; 3.9B depicts the distance for POPC 50; 3.9C depicts the distance for POPC 40.

3.4.9. - Molecule Shapes and Sizes.

All three systems indicate little variation in both the radius of gyration (R_g) and the solvent accessible surface area (SASA) of the phenylalanine dipeptide. This is not surprising due to the small size of the peptide. The relevant figures (Figures 3.17 to 3.20) can be seen in the Supplementary Information. There is a small decrease in both the hydrophilic and total SASA as the dipeptide nears the center of all three membrane systems. Such a change is expected as the phenylalanine dipeptide attempts to lower its interfacial free energy within the hydrophobic membrane interior. However, this difference is not statistically significant. Conversely, there are small differences in the radius of gyration to within a couple picometers; however, they do not follow a consistent pattern, other than to indicate a slightly greater variation near the membrane transition as opposed to the center of the membrane or in the

aqueous region. Again, these differences are very small and may not be chemically significant; they are simply mentioned here for comparison with larger peptides that may possess greater changes.

3.4.10 - Specific Interactions.

The phenylalanine dipeptide interacts with a different environment at varying distances from the center of the model DOPC or POPC membrane. The number of water molecules within 0.3 nm of any part of the dipeptide is plotted in Figure 3.10. In the aqueous region 3.0 nm from the center of the model membranes, all three peptides were surrounded by around 45 water molecules, with the POPC and DOPC 50 systems at nearly the same value; whereas, the POPC 40 system has a slightly larger number. As the dipeptide moves closer to the membrane interface, the number of closely coordinated water molecules gradually decreases to around 20 at z = 1.5 nm. As expected for relatively small and non-ionic permeants, there are few, if any, water molecules surrounding the dipeptide at the center of membrane. However, it is interesting to note that at z = 0.0 nm from the membrane center, the DOPC 50, POPC 50, and POPC 40 systems possess the following number of nearby water molecules: 0.690 ± 0.007, 0.006 ± 0.001, 0.006 ± 0.001. The numbers for the POPC systems are essentially zero; however, the DOPC 50 system is statistically above zero. This may indicate a slightly greater permeability for the DOPC system due to its less ordered lipid tails.

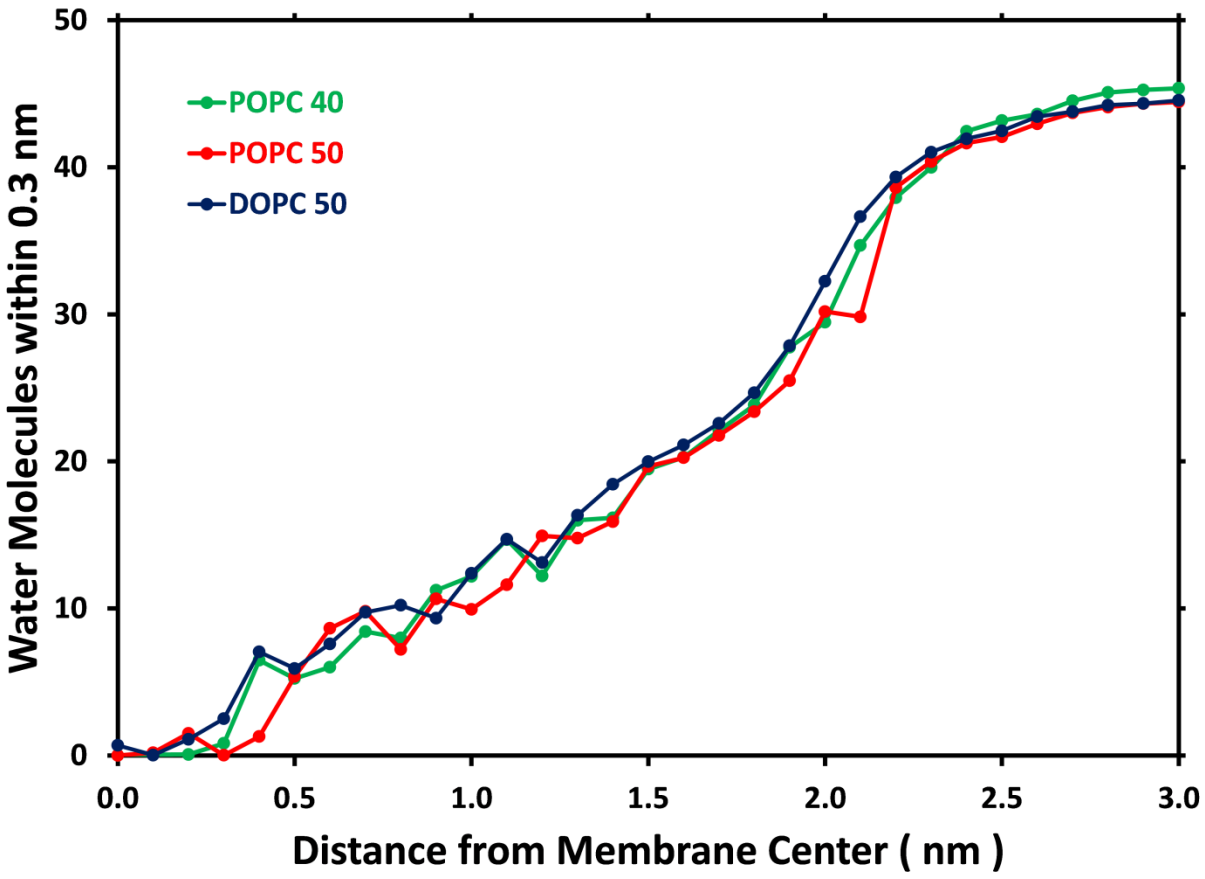


Figure 3.10. The number of water molecules within 0.3 nm of the phenylalanine dipeptide in all three membrane systems is displayed. At each membrane position, the standard error of the sample mean was calculated at a 95% confidence level and was less than 0.06 water molecules -- too small to appear on this figure.

Radial distribution functions (RDF) between either the backbone or sidechain moiety of the phenylalanine dipeptide interacting with the non-hydrogen atom components of water, the lipid tails, or the remaining lipid headgroup atoms are plotted in Figure 3.11 for the POPC 50 system. The corresponding figures for the DOPC 50 and POPC 40 systems are not included because they are indistinguishable to any but the most discerning eye. At the top of panel A and panel B, both moieties of the peptide are surrounded by a relatively uniform distribution of water molecules. However, this uniformity comes to gradual stop at 2.0 nm from the

membrane center for the sidechain; whereas, the backbone continues to display a relationship with the aqueous region until a distance of 1.0 nm. This trend is corroborated by the insertion angle data in Figure 3.5 where the peptide backbone remains oriented towards the aqueous region at greater membrane insertion depths. Similarly, both peptide moieties strongly associate with the lipid headgroups in the interfacial regions between 0.8 and 2.0 nm; however, the sidechain possess a more abrupt transition away from the lipid headgroups at 1.0 nm, as opposed to 0.4 nm for the backbone. Finally, the opposite trend is noticed with the lipid tails: the sidechain remains closer to them over a greater membrane distance (0.0 to 1.9 nm); whereas, the backbone strongly associates with the tail groups near the membrane center (0.0 to 1.0 nm.)

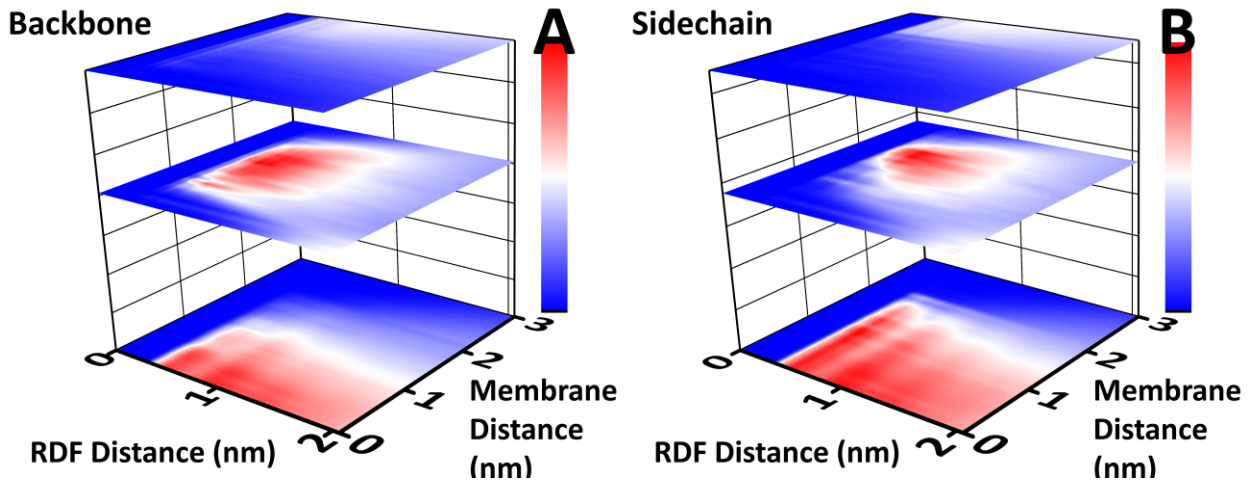


Figure 3.11. Radial distribution functions for either the peptide backbone (A) or the peptide sidechain (B) with water (top), lipid headgroups (middle), and lipid tails (bottom) are plotted as a function of the distance from either the dipeptide sidechain or backbone and the distance from the center of the membrane. Data shown is for POPC 50 systems; results for DOPC 50 and POPC 40 are highly similar and are not shown. Red indicates a large, relative value for the radial distribution function; white is an intermediate value; and blue is a small or zero value.

Coordination numbers representing the average number of atoms belonging to each of the previously mentioned groups within 0.5 nm of the backbone or sidechain are plotted in Figure 3.12. In Figure 3.12A, the coordination number between water and the backbone gradually decreases with distance from the center of the membrane; however, in Figure 3.12D, the sidechain undergoes a more distinct transition as distance increases and rapidly loses coordinating water atoms at 2.1 nm. In Figure 3.12B, the backbone is coordinated by a significant number of headgroup atoms over a large distance (0.6 to 2.5 nm) as compared to the sidechain (1.7 to 2.2 nm) in panel 3.12E. Finally, the sidechain coordinates with the lipid tails at around 2.0 m, in Figure 3.12F, before completing said transition at 1.0 nm; whereas, the backbone undergoes a more gradual process. Again, this matches both the RDF and *z*-insertion angle data that demonstrate that the phenylalanine sidechain enters the membrane first, and the backbone follows. Interestingly, there is not an observable difference in coordination number between the choice of lipid type (DOPC 50 in blue as compared to POPC 50 in green) or system size (POPC 50 in green and POPC 40 in red).

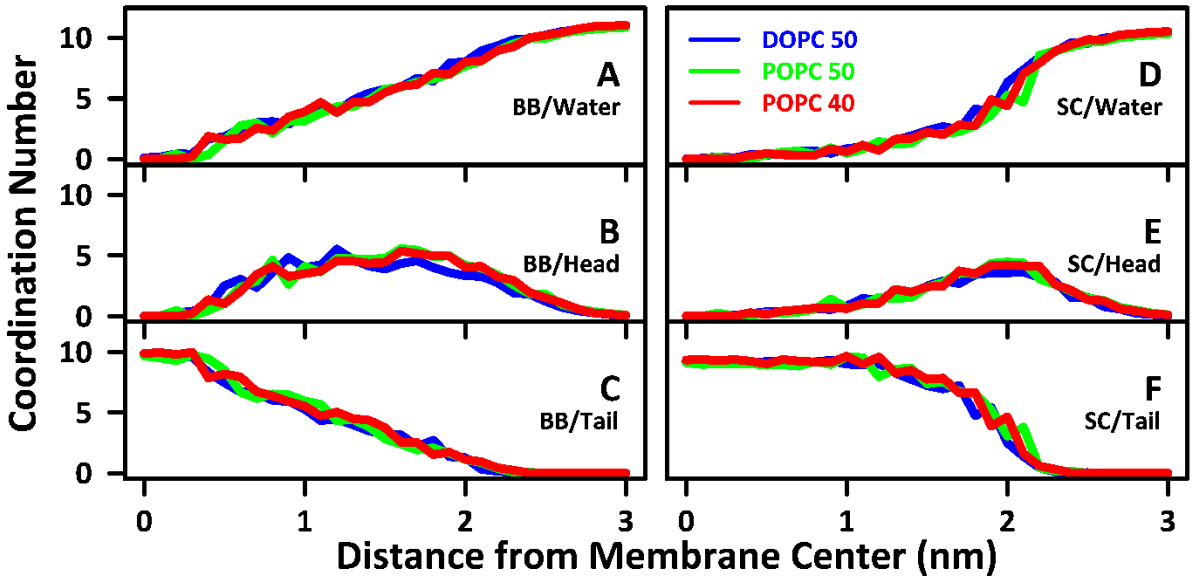


Figure 3.12. Coordination numbers are displayed and represent the average number of the specified sub-groups within 0.5 nm of either the dipeptide backbone or sidechain. The peptide sidechain was defined as all of the non-hydrogen atoms of the phenylalanine R-group, and the backbone consists of the remaining non-hydrogen atoms in the dipeptide, including those of the terminal blocking groups. Sub-groups were defined, as follows: water, only the oxygen atom; lipid tail, all of the carbon atoms in the acyl group of the lipid; lipid headgroup, all of the non-hydrogen and non-tail atoms remaining in the lipid. Panel 3.12A displays the coordinating number of water groups around the backbone; Panel 3.12B, the backbone and lipid headgroups; Panel 3.12C, the backbone and lipid tails; Panel 3.12D, the sidechain and water; Panel 3.12E, the sidechain and lipid headgroups; Panel 3.12F, the sidechain and lipid tails.

3.5. Conclusions

In this study, we conducted a wide variety of computational analyses to determine the effect of system size and lipid type on the passive permeation of the phenylalanine dipeptide. This was undertaken in the broader context of exploring and explaining the passive permeation of aromatic amino acids through model cell membranes. In terms of the effect of system size, the potential of mean force was lowered slightly by about 3 $kJmol^{-1}$, in a large bilayer, potentially indicating that membrane fluctuations are not being well sampled with only 40

POPC lipids or indicating that a simulation length of 50 ns per window is insufficient to sample these fluctuations. A similar trend was also noticed for the mean first passage time, which is highly coupled with the results for the potential of mean force. Along these lines, it was more difficult to achieve convergence in the rotational sidechain anisotropy studies, even with simulations 250 ns in length. A greater number of X_1 and X_2 sidechain conformations were witnessed in the interfacial region for POPC 50, indicating that the larger membrane system may accommodate more rarely sampled configurations. No difference was witnessed in terms of permeation coefficients, translational diffusion, insertion angles, Ramachandran plots, clustering, backbone oxygen/nitrogen distance, radius of gyration, solvent accessible surface area, radial distribution functions, and coordination numbers.

In terms of lipid type, recall that DOPC contains two, singly unsaturated acyl chains, whereas POPC contains only a single unsaturated acyl chain. The addition of a double bond creates greater disorder within the lipid structure. This effect is most easily seen within the potential of mean force for DOPC where the PMF is broader than that for either of the POPC systems. This then affects the mean first passage time which is highly sensitive to changes in the PMF. The clustering and backbone oxygen to nitrogen distance analyses also indicate that the DOPC system transitions to a C_{7eq} conformation sooner within the membrane. This also indicates a greater degree of disorder which allows the phenylalanine dipeptide to adopt this low energy conformation in more shallow regions of the lipid bilayer. No difference was witnessed in terms of permeation coefficients, translational diffusion, insertion angles, Ramachandran plots, radius of gyration, solvent accessible surface area, radial distribution functions, and coordination numbers.

Although only a few differences were witnessed in terms of system size and lipid type, a wide variety of observations were made in terms of the general permeation process. The potential of mean force exhibited the typical pattern for amphiphilic molecules: an energy minimum near the interface and an energy maximum near the interior. Translational diffusion coefficients were largely unchanged with insertion depth, with a small decrease near the interfacial region. Accordingly, the mean first passage time closely followed the trend in terms of the potential of mean force at the center of the membrane. Unfortunately, calculation of permeability coefficients was challenging due to the choice of starting location within the aqueous simulation region. Sidechain rotation was more inhibited near the interfacial region and freer in either the aqueous or central membrane regions. In the permeation process, the sidechain enters the membrane first, then followed by the backbone, adduced by the insertion angle, radial distribution function, and coordination number analyses. Finally, conformational analyses indicate more freedom of motion within the aqueous region and more restricted conformational freedom within the center of the membrane.

3.6. Supplementary Information

3.6.1. - Electron Density Profile.

The electron density profiles for the DOPC 50 and POPC 50 systems studied throughout this article are plotted below in Figure 3.13. Experimental data was obtained from the article by Liu and Nagle.[58]

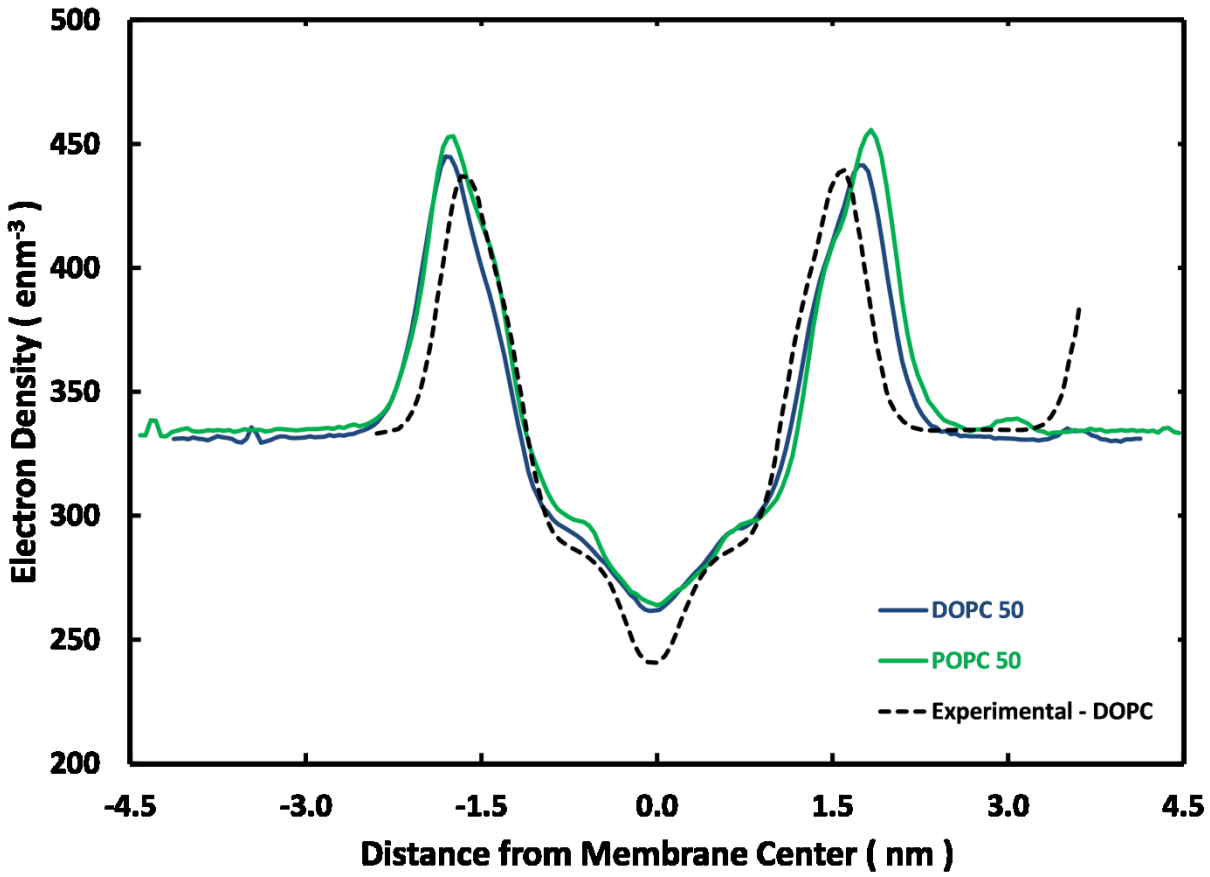


Figure 3.13. Electron density profiles for the DOPC and POPC 50 simulation systems.

3.6.2. - PMF Quartiles.

The Potential of Mean Force (PMF) data reported within the primary article was calculated from the last half of the simulation trajectory, and the standard error was calculated using the Bootstrap method. However, these errors were substantially smaller than the variation witnessed when the PMF was calculated over each quarter of the simulation, and varied between 1 and 20 $kJmol^{-1}$ for the membrane interface and between 2 and 10 $kJmol^{-1}$ for the central barrier. These PMFs are plotted below. The lightest line represents the 1^{st} quarter, and the darkest line represents the PMF for the 4^{th} and last quarter.

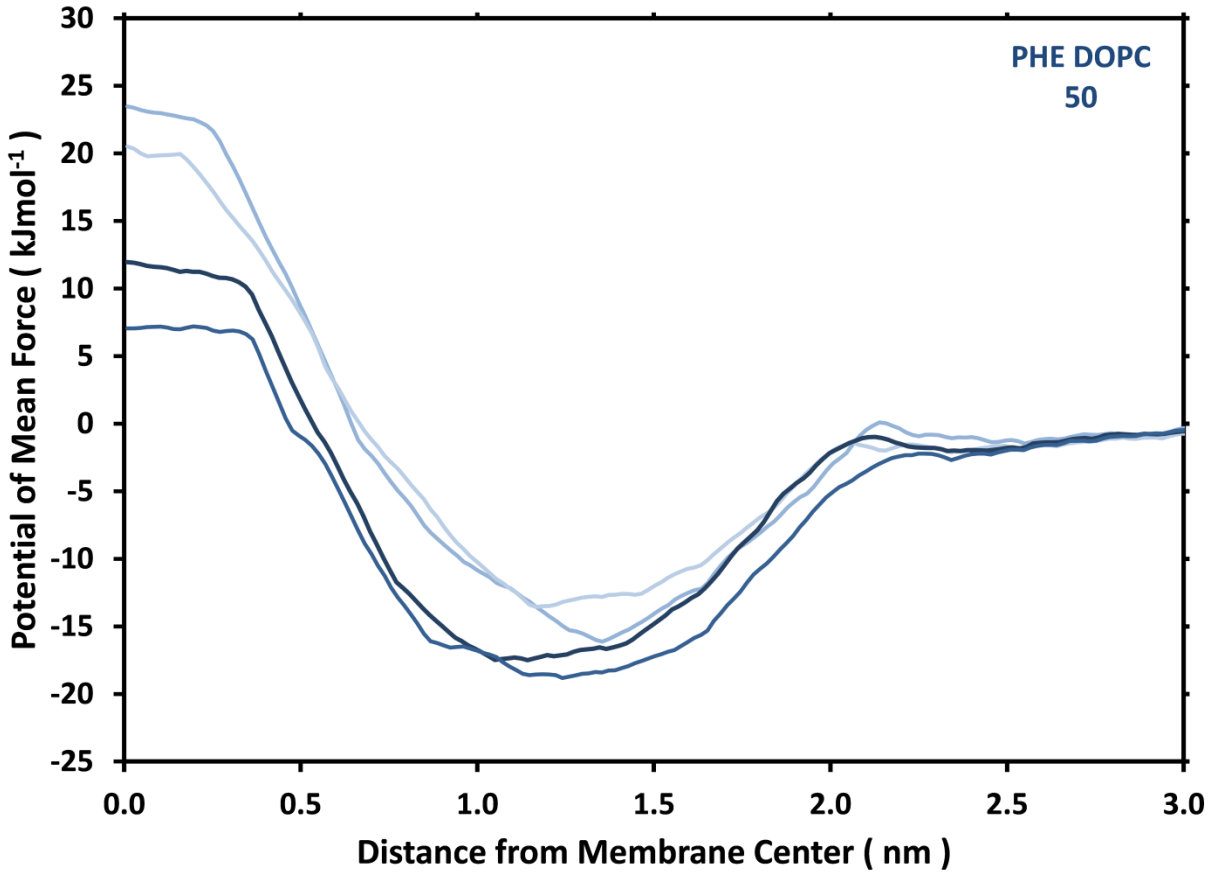


Figure 3.14. Potential of mean force quartiles for the phenylalanine dipeptide in the DOPC 50 system.

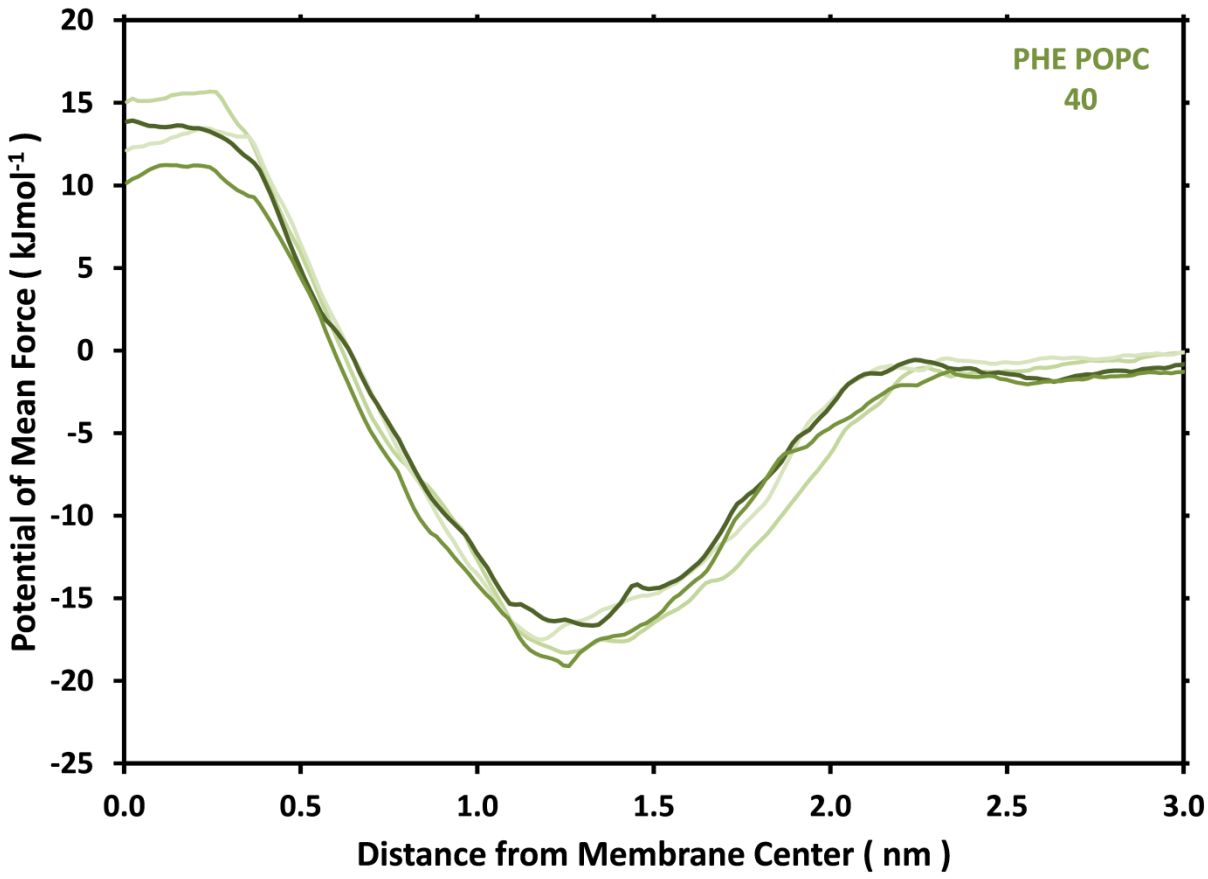


Figure 3.15. Potential of mean force quartiles for the phenylalanine dipeptide in the POPC 40 system.

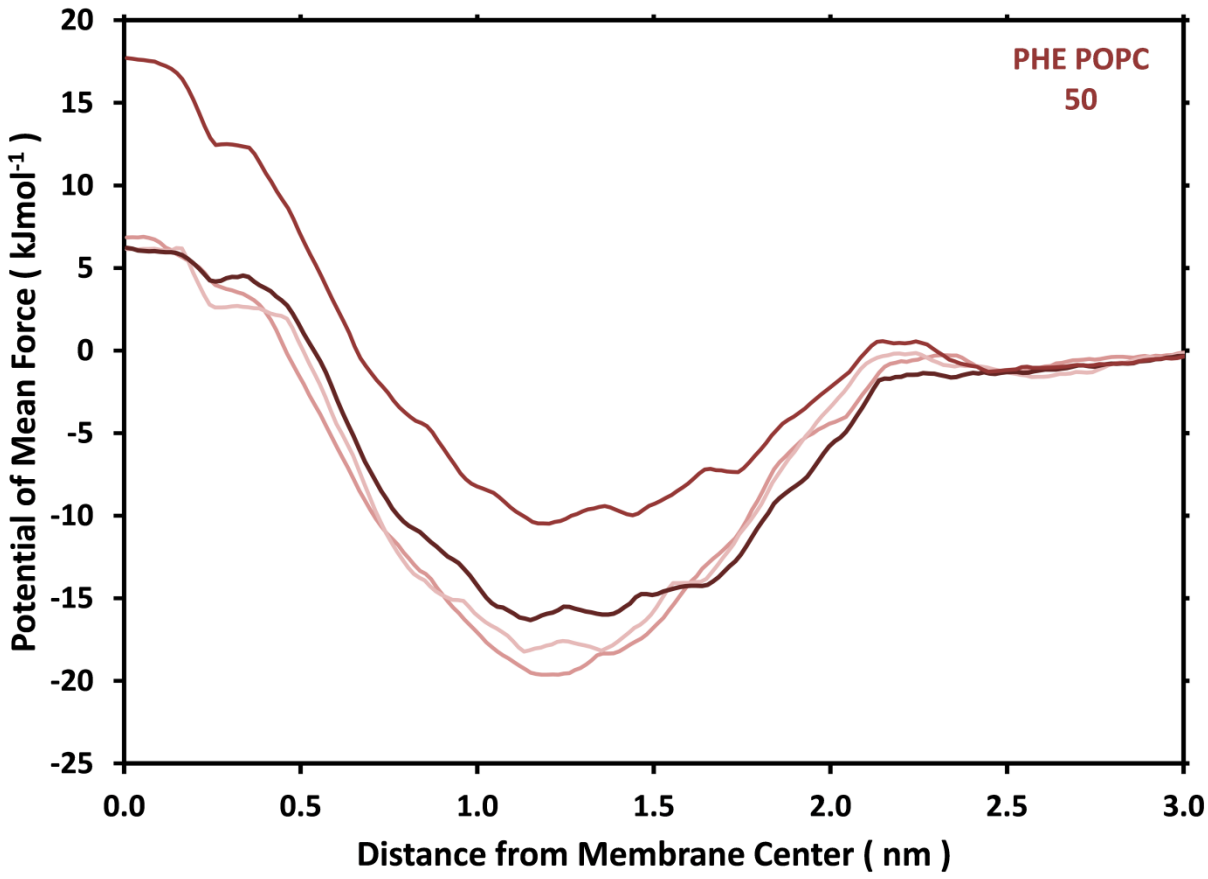


Figure 3.16. Potential of mean force quartiles for the phenylalanine dipeptide in the POPC 50 system.

3.6.3. - Radius of Gyration.

The radius of gyration with respect to the center of mass of the phenylalanine dipeptide is plotted below with respect to the simulation *z*-axis. The standard error of each sample mean was calculated and was not plotted because it was around 0.02% of the radius of gyration for all simulation depths. In contrast, the standard error of each sample was plotted in our previous article[56] and is around 3-5% of the radius of gyration, which is similar to these results. Either way, a discernible pattern is not present, as seen below in Figure 3.17.

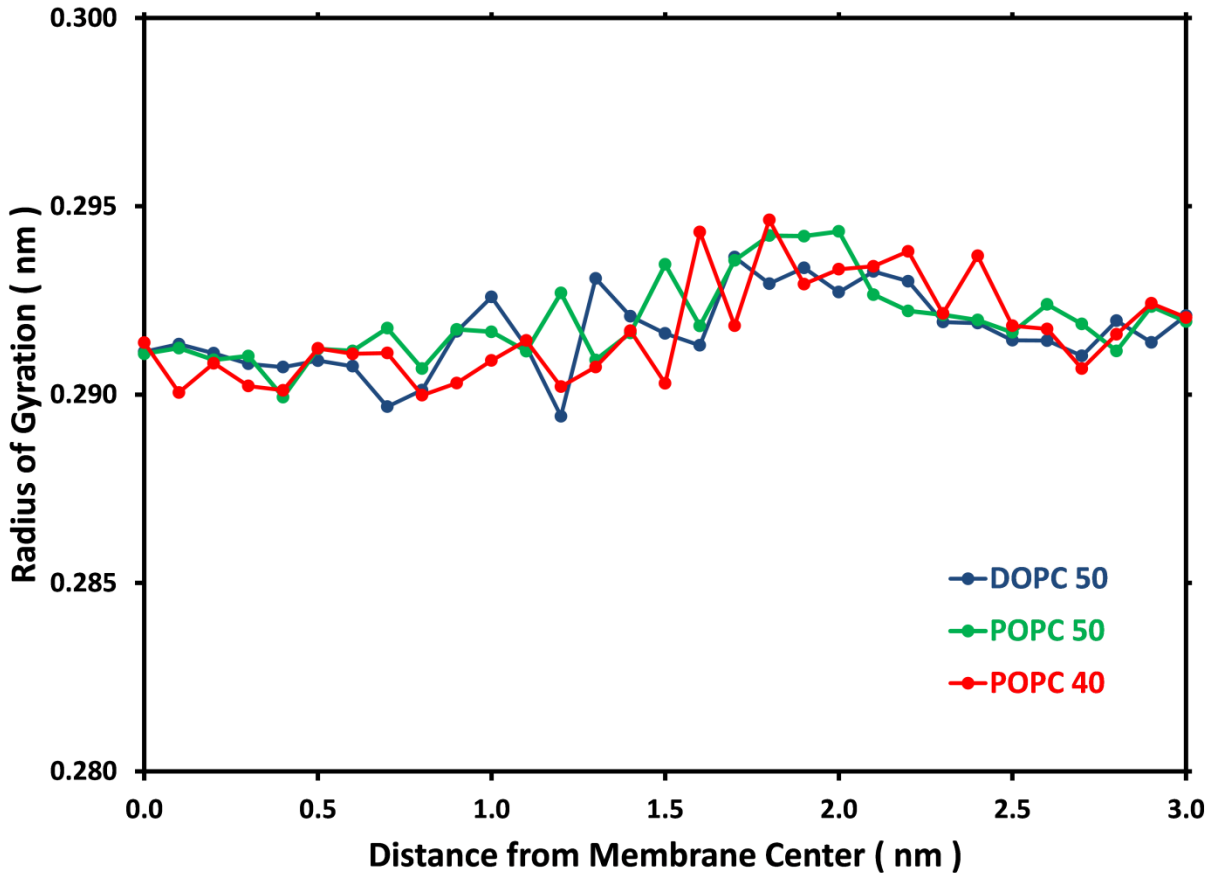


Figure 3.17. The radius of gyration for the phenylalanine dipeptide in DOPC and POPC lipid bilayers of varying sizes. 95% confidence intervals for the sample mean for each trajectory windows are plotted but are too small to be seen.

3.6.5. - Solvent Accessible Surface Area.

The solvent accessible surface area -- hydrophilic, hydrophobic, and total -- can be found below, respectively, in Figures 3.18, 3.19, and 3.20. The standard error for each sample mean at a 95% confidence is also plotted. There is little, if any, variation in all of these surface area measurements.

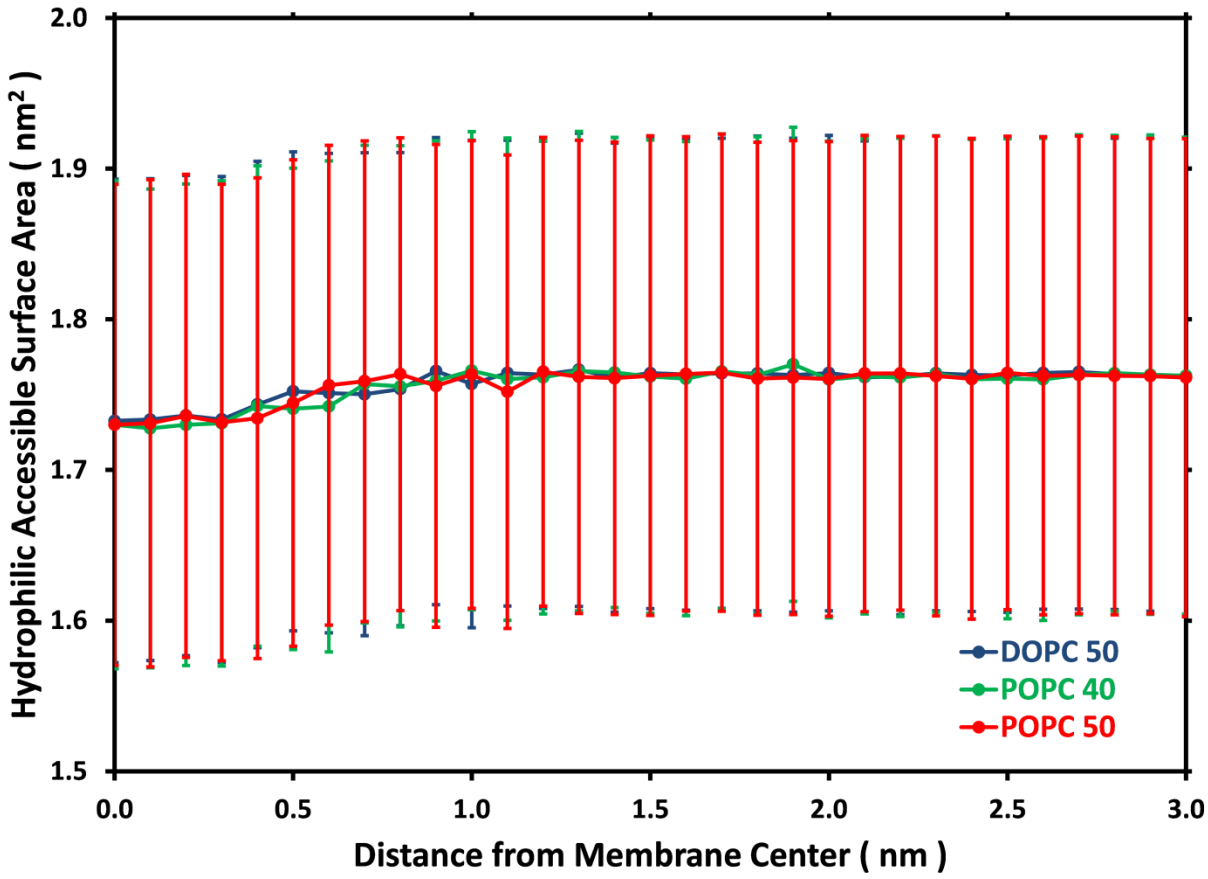


Figure 3.18. The hydrophilic accessible surface area for the phenylalanine dipeptide in DOPC and POPC lipid bilayers of varying sizes. The standard error (95% confidence) for the sample mean for each umbrella sampling window is plotted.

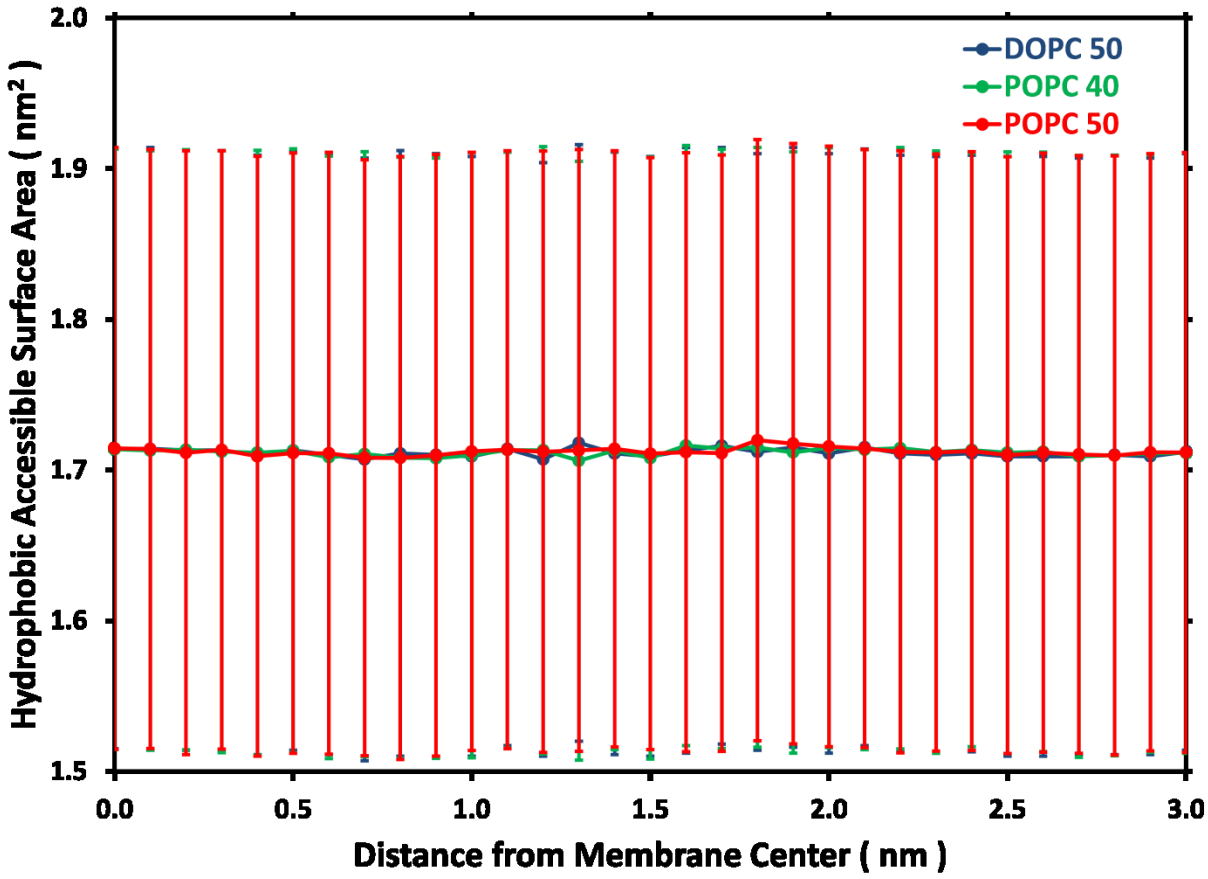


Figure 3.19. The hydrophobic accessible surface area for the phenylalanine dipeptide in DOPC and POPC lipid bilayers of varying sizes. The standard error (95% confidence) for the sample mean for each umbrella sampling window is plotted.

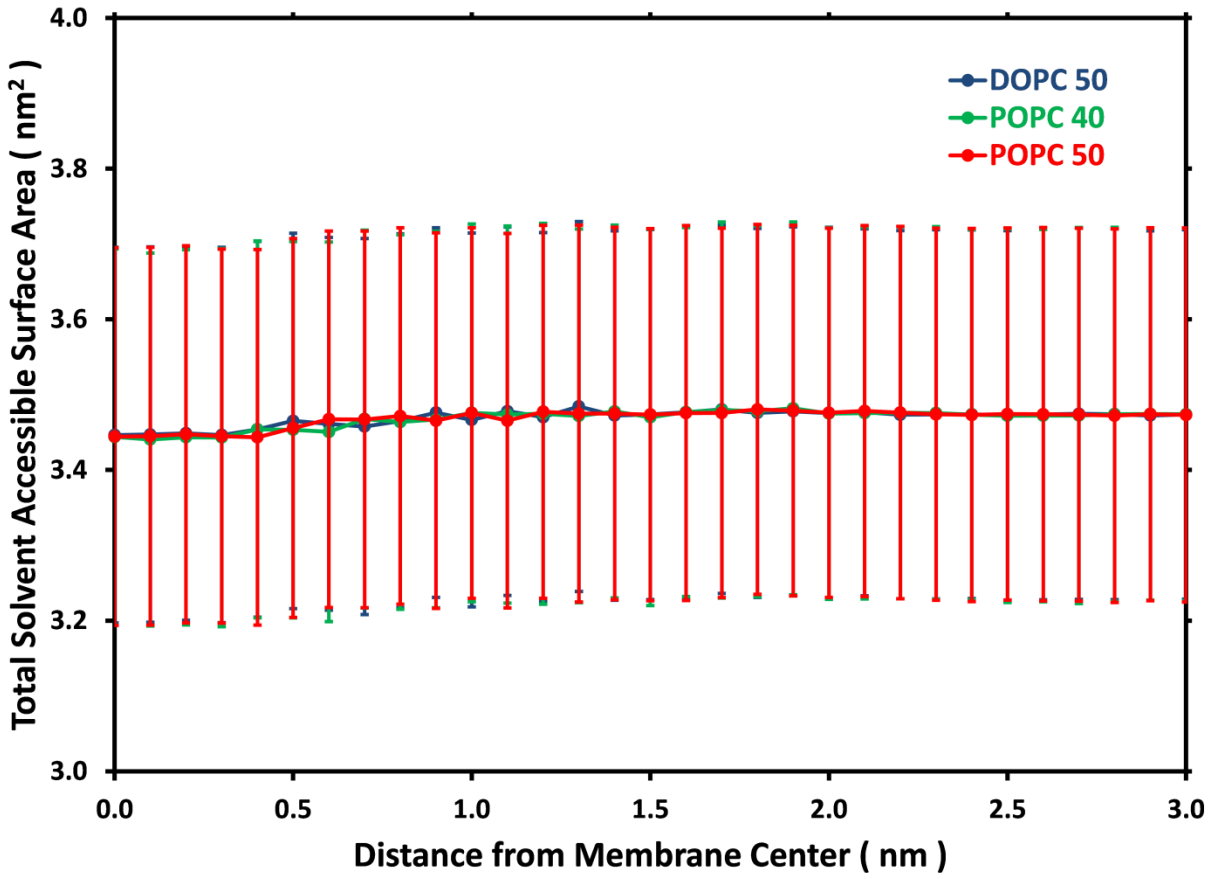


Figure 3.20. The total solvent accessible surface area for the phenylalanine dipeptide in DOPC and POPC lipid bilayers of varying sizes. The standard error (95% confidence) for the sample mean for each umbrella sampling window is plotted.

3.7. References

1. Whitney, W. R., A Dictionary of Applied Chemistry. *Science* **1914,** *40* (1033), 562--563.

2. Galili, G.; Amir, R.; Fernie, A. R., The Regulation of Essential Amino Acid Synthesis and Accumulation in Plants. In *Annual Review of Plant Biology, Vol 67*, Merchant, S. S., Ed. Annual Reviews: Palo Alto, 2016; Vol. 67, pp 153-+.

3. Furst, P.; Stehle, P., What are the essential elements needed for the determination of amino acid requirements in humans? *J. Nutr.* **2004,** *134* (6), 1558S-1565S.

4. Hinz, M.; Stein, A.; Uncini, T., Amino acid management of Parkinson's disease: a case study (vol 4, pg 165, 2011). *Int. J. Gen. Med.* **2016,** *9*, 205-205.

5. Hirayama, M.; Tsunoda, M.; Yamamoto, M.; Tsuda, T.; Ohno, K., Serum Tyrosine-to-Phenylalanine Ratio is Low in Parkinson's Disease. *J. Parkinsons Dis.* **2016,** *6* (2), 423-431.

6. Roiser, J. P.; McLean, A.; Ogilvie, A. D.; Blackwell, A. D.; Bamber, D. J.; Goodyer, I.; Jones, P. B.; Sahakian, B. J., The subjective and cognitive effects of acute phenylalanine and tyrosine depletion in patients recovered from depression. *Neuropsychopharmacology* **2005,** *30* (4), 775-785.

7. Cardenas, A. E.; Shrestha, R.; Webb, L. J.; Elber, R., Membrane Permeation of a Peptide: It Is Better to be Positive. *The Journal of Physical Chemistry B* **2015,** *119* (21), 6412-6420.

8. Kuczera, K.; Unruh, J.; Johnson, C. K.; Jas, G. S., Reorientations of Aromatic Amino Acids and Their Side Chain Models: Anisotropy Measurements and Molecular Dynamics Simulations. *The Journal of Physical Chemistry A* **2010,** *114* (1), 133-142.

9. Ano, R.; Kimura, Y.; Shima, M.; Matsuno, R.; Ueno, T.; Akamatsu, M., Relationships between structure and high-throughput screening permeability of peptide derivatives and related compounds with artificial membranes: application to prediction of Caco-2 cell permeability. *Bioorganic & Medicinal Chemistry* **2004,** *12* (1), 257-264.

10. Marrink, S.; Berendsen, H., Simulation of Water Transport Through a Lipid-Membrane. *Journal of Physical Chemistry* **1994,** *98* (15), 4155-4168.

11. Marrink, S. J.; Berendsen, H. J. C., Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations. *The Journal of Physical Chemistry* **1996,** *100* (41), 16729-16738.

12. Cascales, J. J. L.; Berendsen, H. J. C.; delaTorre, J. G., Molecular dynamics simulation of water between two charged layers of dipalmitoylphosphatidylserine. *Journal of Physical Chemistry* **1996,** *100* (21), 8621-8627.

13. Lee, B. L.; Kuczera, K., Simulating the free energy of passive membrane permeation for small molecules. *Molecular Simulation* **2018,** *44* (13-14), 1147-1157.

14. Awoonor-Williams, E.; Rowley, C. N., Molecular simulation of nonfacilitated membrane permeation. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2015**.

15. Lee, C. T.; Comer, J.; Herndon, C.; Leung, N.; Pavlova, A.; Swift, R. V.; Tung, C.; Rowley, C. N.; Amaro, R. E.; Chipot, C.; Wang, Y.; Gumbart, J. C., Simulation-Based Approaches for Determining Membrane Permeability of Small Compounds. *Journal of Chemical Information and Modeling* **2016,** *56* (4), 721-733.

16. Hansen, F. Y.; Peters, G. H.; Taub, H.; Miskowiec, A., Diffusion of water and selected atoms in DMPC lipid bilayer membranes. *The Journal of Chemical Physics* **2012,** *137* (20), 204910.

17. Wei, T.; Huang, T.; Qiao, B.; Zhang, M.; Ma, H.; Zhang, L., Structures, Dynamics, and Water Permeation Free Energy across Bilayers of Lipid A and Its Analog Studied with Molecular Dynamics Simulation. *The Journal of Physical Chemistry B* **2014,** *118* (46), 13202-13209.

18. Comer, J.; Schulten, K.; Chipot, C., Diffusive Models of Membrane Permeation with Explicit Orientational Freedom. *Journal of Chemical Theory and Computation* **2014,** *10* (7), 2710-2718.

19. Orsi, M.; Sanderson, W. E.; Essex, J. W., Permeability of Small Molecules through a Lipid Bilayer: A Multiscale Simulation Study. *The Journal of Physical Chemistry B* **2009,** *113* (35), 12019-12029.

20. Riahi, S.; Rowley, C. N., Why Can Hydrogen Sulfide Permeate Cell Membranes? *Journal of the American Chemical Society* **2014,** *136* (43), 15111-15113.

21. Bemporad, D.; Luttmann, C.; Essex, J. W., Behaviour of small solutes and large drugs in a lipid bilayer from computer simulations. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2005,** *1718* (1-2), 1-21.

22. Loverde, S. M., Molecular Simulation of the Transport of Drugs across Model Membranes. *The Journal of Physical Chemistry Letters* **2014,** *5* (10), 1659-1665.

23. Alper, H. E.; Stouch, T. R., Orientation and diffusion of a drug analog in biomembranes: Molecular dynamics simulations. *The Journal of Physical Chemistry* **1995,** *99* (15), 5724-5731.

24. Orsi, M.; Essex, J. W., Permeability of drugs and hormones through a lipid bilayer: insights from dual-resolution molecular dynamics. *Soft Matter* **2010,** *6* (16), 3797.

25. Martin, L. J.; Chao, R.; Corry, B., Molecular dynamics simulation of the partitioning of benzocaine and phenytoin into a lipid bilayer. *Biophysical Chemistry* **2014,** *185*, 98-107.

26. Porasso, R. D.; Drew Bennett, W. F.; Oliveira-Costa, S. D.; López Cascales, J. J., Study of the Benzocaine Transfer from Aqueous Solution to the Interior of a Biological Membrane. *The Journal of Physical Chemistry B* **2009,** *113* (29), 9988-9994.

27. Nademi, Y.; Iranagh, S. A.; Yousefpour, A.; Mousavi, S. Z.; Modarress, H., Molecular dynamics simulations and free energy profile of Paracetamol in DPPC and DMPC lipid bilayers. *Journal of Chemical Sciences* **2014,** *126* (3), 637-647.

28. Khuntawee, W.; Wolschann, P.; Rungrotmongkol, T.; Wong-ekkabut, J.; Hannongbua, S., Molecular Dynamics Simulations of the Interaction of Beta Cyclodextrin with a Lipid Bilayer. *Journal of Chemical Information and Modeling* **2015,** *55* (9), 1894-1902.

29. dos Santos, D. J. V. A.; Eriksson, L. A., Permeability of psoralen derivatives in lipid membranes. *Biophys. J.* **2006,** *91* (7), 2464-2474.

30. Cascales, J. J. L.; Huertas, M. L.; de la Torre, J. G., Molecular dynamics simulation of a dye molecule in the interior of a bilayer: 1,6-diphenyl-1,3,5-hexatriene in dipalmitoylphosphatidylcholine. *Biophysical Chemistry* **1997,** *69* (1), 1-8.

31. Kyrychenko, A.; Sevriukov, I. Y.; Syzova, Z. A.; Ladokhin, A. S.; Doroshenko, A. O., Partitioning of 2,6-Bis(1H-Benzimidazol-2-yl)pyridine fluorophore into a phospholipid bilayer: Complementary use of fluorescence quenching studies and molecular dynamics simulations. *Biophysical Chemistry* **2011,** *154* (1), 8-17.

32. Nolte, T. M.; Kettler, K.; Meesters, J. A. J.; Hendriks, A. J.; van de Meent, D., A semi-empirical model for transport of inorganic nanoparticles across a lipid bilayer: Implications for uptake by living cells: Passive uptake of nanoparticles. *Environmental Toxicology and Chemistry* **2015,** *34* (3), 488-496.

33. Oroskar, P. A.; Jameson, C. J.; Murad, S., Surface-Functionalized Nanoparticle Permeation Triggers Lipid Displacement and Water and Ion Leakage. *Langmuir* **2015,** *31* (3), 1074-1085.

34. Fiedler, S. L.; Violi, A., Simulation of Nanoparticle Permeation through a Lipid Membrane. *Biophys. J.* **2010,** *99* (1), 144-152.

35. Herce, H. D.; Garcia, A. E., Cell Penetrating Peptides: How Do They Do It? *Journal of Biological Physics* **2007,** *33* (5-6), 345-356.

36. Herce, H. D.; Garcia, A. E., Molecular dynamics simulations suggest a mechanism for translocation of the HIV-1 TAT peptide across lipid membranes. *Proceedings of the National Academy of Sciences* **2007,** *104* (52), 20805-20810.

37. Khajeh, A.; Modarress, H., Effect of cholesterol on behavior of 5-fluorouracil (5-FU) in a DMPC lipid bilayer, a molecular dynamics study. *Biophysical Chemistry* **2014,** *187-188*, 43-50.

38. Söderhäll, J. A.; Laaksonen, A., Molecular Dynamics Simulations of Ubiquinone inside a Lipid Bilayer. *The Journal of Physical Chemistry B* **2001,** *105* (38), 9308-9315.

39. Cardenas, A. E.; Jas, G. S.; DeLeon, K. Y.; Hegefeld, W. A.; Kuczera, K.; Elber, R., Unassisted Transport of N-Acetyl-L-tryptophanamide through Membrane: Experiment and Simulation of Kinetics. *The Journal of Physical Chemistry B* **2012,** *116* (9), 2739-2750.

40. Wood, I.; Martini, M. F.; Pickholz, M., Similarities and differences of serotonin and its precursors in their interactions with model membranes studied by molecular dynamics simulation. *Journal of Molecular Structure* **2013,** *1045*, 124-130.

41. MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P., Transfer of Arginine into Lipid Bilayers Is Nonadditive. *Biophys. J.* **2011,** *101* (1), 110-117.

42. Bonhenry, D.; Tarek, M.; Dehez, F., Effects of Phospholipid Composition on the Transfer of a Small Cationic Peptide Across a Model Biological Membrane. *Journal of Chemical Theory and Computation* **2013,** *9* (12), 5675-5684.

43. Hu, Y.; Ou, S. C.; Patel, S., Free Energetics of Arginine Permeation into Model DMPC Lipid Bilayers: Coupling of Effective Counterion Concentration and Lateral Bilayer Dimensions. *Journal of Physical Chemistry B* **2013,** *117* (39), 11641-11653.

44. MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P., Distribution of Amino Acids in a Lipid Bilayer from Computer Simulations. *Biophys. J.* **2008,** *94* (9), 3393-3404.

45. Baoukina, S.; Rozmanov, D.; Tieleman, D. P., Composition Fluctuations in Lipid Bilayers. *Biophys. J.* **2017,** *113* (12), 2750-2761.

46. Issack, B. B.; Peslherbe, G. H., Effects of Cholesterol on the Thermodynamics and Kinetics of Passive Transport of Water through Lipid Membranes. *The Journal of Physical Chemistry B* **2015,** *119* (29), 9391-9400.

47. Rabinovich, A. L.; Lyubartsev, A. P., Computer Simulation of Lipid Membranes: Methodology and Achievements. *Polymer Science Series C* **2013,** *55* (1), 162-180.

48. Sugii, T.; Takagi, S.; Matsumoto, Y., A molecular-dynamics study of lipid bilayers: Effects of the hydrocarbon chain length on permeability. *The Journal of Chemical Physics* **2005,** *123* (18), 184714.

49. Nitschke, N.; Atkovska, K.; Hub, J. S., Accelerating potential of mean force calculations for lipid membrane permeation: System size, reaction coordinate, solute-solute distance, and cutoffs. *J. Chem. Phys.* **2016,** *145* (12), 9.

50. Lee, B. L.; Kuczera, K., Simulating the free energy of passive membrane permeation for small molecules. *Molecular Simulation* **2017**, 1-11.

51. Tieleman, D. P., Chapter 1 Methods and Parameters for Membrane Simulations. In *Molecular Simulations and Biomembranes: From Biophysics to Function*, The Royal Society of Chemistry: 2010; pp 1-25.

52. Orsi, M.; Essex, J. W., Chapter 4 Passive Permeation Across Lipid Bilayers: a Literature Review. In *Molecular Simulations and Biomembranes: From Biophysics to Function*, The Royal Society of Chemistry: 2010; pp 76-90.

53. Dickey, A. N.; Faller, R., Molecular Modeling of Biomembranes: A How-to ApproachIn *Biomedical Applications of Biophysics*, Jue, T., Ed. Springer: Totowa, NJ, 2010.

54. Xiang, T. X.; Anderson, B. D., Liposomal drug transport: A molecular perspective from molecular dynamics simulations in lipid bilayers. *Advanced Drug Delivery Reviews* **2006,** *58* (12-13), 1357-1378.

55. Tieleman, D. P., Computer Simulations of Transport through Membranes: Passive Diffusion, Pores, Channels and Transporters. *Clinical and Experimental Pharmacology and Physiology* **2006,** *33* (10), 893-903.

56. Lee, B. L.; Kuczera, K.; Middaugh, C. R.; Jas, G. S., Permeation of the three aromatic dipeptides through lipid bilayers: Experimental and computational study. *J. Chem. Phys.* **2016,** *144* (24), 14.

57. Petrache, H. I.; Tristram-Nagle, S.; Gawrisch, K.; Harries, D.; Parsegian, V. A.; Nagle, J. F., Structure and fluctuations of charged phosphatidylserine bilayers in the absence of salt. *Biophys. J.* **2004,** *86* (3), 1574-1586.

58. Liu, Y.; Nagle, J. F., Diffuse scattering provides material parameters and electron density profiles of biomembranes. *Physical Review E* **2004,** *69* (4).

59. Leftin, A.; Molugu, T. R.; Job, C.; Beyer, K.; Brown, M. F., Area per lipid and cholesterol interactions in membranes from separated local-field (13)C NMR spectroscopy. *Biophys. J.* **2014,** *107* (10), 2274-2286.

60. Jo, S.; Kim, T.; Im, W., Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS ONE* **2007,** *2* (9), e880.

61. Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W., CHARMM-GUI *Membrane Builder* toward realistic biological membrane simulations. *Journal of Computational Chemistry* **2014,** *35* (27), 1997-2004.

62. Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W., CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **2009,** *97* (1), 50-58.

63. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013,** *29* (7), 845-854.

64. Pastor, R. W.; MacKerell, A. D., Development of the CHARMM Force Field for Lipids. *The Journal of Physical Chemistry Letters* **2011,** *2* (13), 1526-1532.

65. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **2009**, NA-NA.

66. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926.

67. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995,** *103* (19), 8577.

68. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007,** *126* (1), 014101.

69. Allen, M. P.; Tildesley, D. J., *Computer simulation of liquids*. Clarendon Press: 1989; p 385.

70. Torrie, G.; Valleau, J., Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *Journal of Computational Physics* **1977,** *23* (2), 187-199.

71. Kästner, J., Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011,** *1* (6), 932-942.

72. Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. the Method. *Journal of Computational Chemistry* **1992,** *13* (8), 1011-1021.

73. Souaille, M.; Roux, B., Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer Physics Communications* **2001,** *135* (1), 40-57.

74. Roux, B., The calculation of the potential of mean force using computer-simulations. *Computer Physics Communications* **1995,** *91* (1-3), 275-282.

75. Hummer, G., Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New Journal of Physics* **2005,** *7*, 34-34.

76. Zusman, L. D., Outer-Sphere Electron-Transfer in Polar-Solvents. *Chemical Physics* **1980,** *49* (2), 295-304.

77. Bicout, D. J.; Szabo, A., Electron transfer reaction dynamics in non-Debye solvents. *J. Chem. Phys.* **1998,** *109* (6), 2325-2338.

78. Diamond, J. M.; Katz, Y., Interpretation of Nonelectrolyte Partition-Coefficients between Dimyristoyl Lecithin and Water. *Journal of Membrane Biology* **1974,** *17* (2), 121-154.

79. Schulten, K., Dynamics of reactions involving diffusive barrier crossing. *The Journal of Chemical Physics* **1981,** *74* (8), 4426.

80. Ulander, J.; Haymet, A. D. J., Permeation across hydrated DPPC lipid bilayers: Simulation of the titrable amphiphilic drug valproic acid. *Biophys. J.* **2003,** *85* (6), 3475-3484.

81. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **2009,** *30* (10), 1545-1614.

82. Hub, J. S.; de Groot, B. L.; van der Spoel, D., g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *Journal of Chemical Theory and Computation* **2010,** *6* (12), 3713-3720.

83. Carpenter, Timothy S.; Kirshner, Daniel A.; Lau, Edmond Y.; Wong, Sergio E.; Nilmeier, Jerome P.; Lightstone, Felice C., A Method to Predict Blood-Brain Barrier Permeability of Drug-Like Compounds Using Molecular Dynamics Simulations. *Biophys. J.* **2014,** *107* (3), 630-641.

84. Filipe, H. A. L.; Moreno, M. J.; Róg, T.; Vattulainen, I.; Loura, L. M. S., How To Tackle the Issues in Free Energy Simulations of Long Amphiphiles Interacting with Lipid Membranes: Convergence and Local Membrane Deformations. *The Journal of Physical Chemistry B* **2014,** *118* (13), 3572-3581.

85. Neale, C.; Bennett, W. F. D.; Tieleman, D. P.; Pomes, R., Statistical Convergence of Equilibrium Properties in Simulations of Molecular Solutes Embedded in Lipid Bilayers. *Journal of Chemical Theory and Computation* **2011,** *7* (12), 4175-4188.

86. Klauda, J. B.; Roberts, M. F.; Redfield, A. G.; Brooks, B. R.; Pastor, R. W., Rotation of lipids in membranes: Molecular dynamics simulation, P-31 spin-lattice relaxation, and rigid-body dynamics. *Biophys. J.* **2008,** *94* (8), 3074-3083.

87. Wei, C.; Pohorille, A., Flip-Flop of Oleic Acid in a Phospholipid Membrane: Rate and Mechanism. *The Journal of Physical Chemistry B* **2014,** *118* (45), 12919-12926.

88. Cardenas, A. E.; Elber, R., Computational study of peptide permeation through membrane: searching for hidden slow variables. *Molecular Physics* **2013,** *111* (22-23), 3565-3578.

89. Comer, J.; Schulten, K.; Chipot, C., Calculation of Lipid-Bilayer Permeabilities Using an Average Force. *Journal of Chemical Theory and Computation* **2014,** *10* (2), 554-564.

90. Neale, C.; Pomes, R., Sampling errors in free energy simulations of small molecules in lipid bilayers. *Biochim. Biophys. Acta-Biomembr.* **2016,** *1858* (10), 2539-2548.

91. Mills, T. T.; Toombes, G. E. S.; Tristram-Nagle, S.; Smilgies, D. M.; Feigenson, G. W.; Nagle, J. F., Order parameters and areas in fluid-phase oriented lipid membranes using wide angle x-ray scattering. *Biophys. J.* **2008,** *95* (2), 669-681.

92. Vaz, W. L.; Almeida, P. F., Microscopic versus macroscopic diffusion in one-component fluid phase lipid bilayer membranes. *Biophys. J.* **1991,** *60* (6), 1553-1554.

93. Comer, J.; Phillips, J. C.; Schulten, K.; Chipot, C., Multiple-Replica Strategies for Free-Energy Calculations in NAMD: Multiple-Walker Adaptive Biasing Force and Walker Selection Rules. *Journal of Chemical Theory and Computation* **2014,** *10* (12), 5276-5285.

94. Parisio, G.; Stocchero, M.; Ferrarini, A., Passive Membrane Permeability: Beyond the Standard Solubility-Diffusion Model. *Journal of Chemical Theory and Computation* **2013,** *9* (12), 5236-5246.

95. Jämbeck, J. P. M.; Lyubartsev, A. P., Exploring the Free Energy Landscape of Solutes Embedded in Lipid Bilayers. *The Journal of Physical Chemistry Letters* **2013,** *4* (11), 1781-1787.

96. Cardenas, A. E.; Elber, R., Modeling kinetics and equilibrium of membranes with fields: Milestoning analysis and implication to permeation. *The Journal of Chemical Physics* **2014,** *141* (5), 054101.

97. Shinoda, W.; Okazaki, S., A Voronoi analysis of lipid area fluctuation in a bilayer. *The Journal of Chemical Physics* **1998,** *109* (4), 1517.

98. Neale, C.; Madill, C.; Rauscher, S.; Pomès, R., Accelerating Convergence in Molecular Dynamics Simulations of Solutes in Lipid Membranes by Conducting a Random Walk along the Bilayer Normal. *Journal of Chemical Theory and Computation* **2013,** *9* (8), 3686-3703.

99. Pal, D.; Chakrabarti, P., On residues in the disallowed region of the Ramachandran map. *Biopolymers* **2002,** *63* (3), 195-206.

100. Ramachandran, G.; Ramakrishnan, C.; Sasisekharan, V., Stereochemistry of Polypeptide Chain Configurations. *Journal of Molecular Biology* **1963,** *7* (1), 95-&.

101. Tobias, D. J.; Brooks, C. L., Conformational Equilibrium in the Alanine Dipeptide in the Gas-Phase and Aqueous-Solution - a Comparison of Theoretical Results. *Journal of Physical Chemistry* **1992,** *96* (9), 3864-3870.

Chapter 4

On the Calculation of Diffusion Constants in Passive Membrane Permeation Studies

4.1 Overview:

Most simulations of membranes and permeation processes focus on calculating the potential of mean force. However, estimating diffusion within lipid bilayers is important in biology and pharmacology. Many different methods and alternative approaches have been discussed in scientific literature. In this Chapter, we explore several different methods, as applied to our previous simulation of the permeation of the phenylalanine dipeptide from Chapter 3. Specifically, diffusion constants were calculated with a wide variety of methods and sub-cases. Each method and case was then evaluated within the context recent literature reports in order to determine optimal approaches and in order to clarify a wide variety of anomalous behavior. Position dependent diffusion constants in the *z*-direction were calculated with the following methods: 1.) The Fluctuation Dissipation Theorem, 2.) Green-Kubo Relations, 3.) Einstein Relations, 4.) The Hummer Displacement Method, and 5.) a Numerical Approximation to the Smoluchowski Equation. We found that the Fluctuation Dissipation theorem produces reasonable data; however, the common practice of using umbrella pulling forces may be producing anomalous results. Green-Kubo relations converge rapidly but the generation of reliable estimates is hindered due to backscattering and negative autocorrelations. Einstein relations suffer many of the same problems as Green-Kubo relations and often depend on linear fits in subjective stability regions. The Hummer Displacement method, although promising, did not work for us. Finally, we prefer an approach that uses a numerical approximation to the Smoluchowski equation that likely avoids most of the previous problems due to averaging and binning time series data. Ultimately, though, one has to

wonder if any of these approaches are appropriate because each possess serious shortcomings from the complex, anisotropic interactions between the peptide and its environment, including the effects of permeant rotational motion and lipid bilayer fluctuations.

4.2 Introduction:

In the research literature, at large, two of the most often discussed quantities are diffusion constants and position dependent diffusion coefficients. These quantities are essential towards evaluating various kinetic properties of a passive membrane permeation process. The most common of these quantities are permeability coefficients and mean first passage times, which are calculated by using the inhomogeneous solubility diffusion model.[1-4] Recall from previous Chapters that these equations are, as follows:

$$P = 1/\int_a^b e^{\beta w(z)} D_z(z)^{-1} dz \tag{4.1}$$

where *P* is the permeability coefficient; *z* is a position along the *z*-axis of the simulation which is perpendicular to the plane of the lipid bilayer; *a* is a location in the aqueous region on one side of the membrane along the *z* axis; *b* is a location in the aqueous region on the opposite side of the membrane; $\beta = (k_B T)^{-1}$; *w(z)* is the potential of mean force at a given location along the *z*-axis; and $D_z(z)$ is the position dependent diffusion coefficient in the *z*-direction as a function of *z*. The mean first passage time, $\langle\tau\rangle$, is calculated by:[5-6]

$$\langle\tau\rangle = \int_a^b \left[e^{\beta w(z)} D_z(z)^{-1} \int_a^z e^{-\beta w(z')} dz' \right] dz \tag{4.2}$$

where all variables are defined in the same manner as for the calculation of P. Notice that $D_z(z)$ plays a prominent role in both calculations and is also an interesting quantity to examine by itself.

However, given the importance of $D_z(z)$, most articles within the literature offer few details concerning its actual calculation. Most simply state a general method and provide a figure a two; many fail to provide a figure or table of $D_z(z)$ and simply report P, $\langle\tau\rangle$, or bulk water diffusion constants. Of those that do report $D_z(z)$ values, the precise timescale and calculation details are often glossed over with statements similar to "fit within the stability region", "integrated until convergence has been achieved," or "averaged over 5 – 10 trajectories." Note: formal citations were intentionally omitted to protect the 'innocent.' Such statements tend to pose more questions than they answer. What is a "stability region" and what prevents stability? How is 'convergence' defined? Why were 5 – 10 trajectories averaged? Was one value averaged over 6 trajectories and another averaged over 10? If so, why were 6 trajectories used in one case, but 10 in another? In addition to this ambiguity, were diffusion coefficients calculated with restrained parameters from umbrella simulations or from free, unconstrained simulations? If so, was the same force field, temperature, simulation package, and so forth kept constant for multiple trials and comparisons. With these questions in mind, we summarize various methods for calculating diffusion constants within the context of recent literature.

One of the most common methods for calculating diffusion constants are Einstein relations which use the mean square displacement (MSD). They are often used to calculate

translational diffusion constants along the *z*-axis, $D_z(z)$, and lateral diffusion constants in the xy plane parallel to the lipid bilayer, $D_{xy}(z)$:

$$D_z(z) = \frac{1}{2} \lim_{t\to\infty} \left\{ \frac{1}{t} \langle [r_z(t_0 + t) - r_z(t_0)]^2 \rangle \right\} \tag{4.3}$$

$$D_{xy}(z) = \frac{1}{4} \lim_{t\to\infty} \left\{ \frac{1}{t} \langle \left| \boldsymbol{r}_{xy}(t_0 + t) - \boldsymbol{r}_{xy}(t_0) \right|^2 \rangle \right\} \tag{4.4}$$

where t is the current time of the simulation, t_0 is the starting time of the calculation, r_z is the position of the permeant along the *z*-axis, $\boldsymbol{r}_{xy}$ is the position of the permeant in the *xy* plane of the simulation as a vector, where the vertical bars in Equation 4.4 represent the magnitude of the displacement vector, and where the angled brackets, <…>, indicate an ensemble average. In either case, the diffusion constant is the slope of a line of best fit from a plot of the MSD with respect to time. Such analyses invariably start at a time other than $t = 0$ in order to avoid the initial ballistic regime and are averaged over different time origins, and the squared displacement is averaged over multiple permeant molecules when more than one permeant is present in the simulation. However, most research journal articles, except those that look at water or translational lipid diffusion, typically contain analyses concerning a single permeant molecule. Finally, note that the dimensional factor in the prefix changes from 2, to 4, to 6, in 1, 2, and 3 dimensional Einstein relations, respectively. Oddly enough, none of the studies in the literature use a three dimensional Einstein relation.

The Einstein MSD method has been used to calculated diffusion constants for many different systems. For example, Yousefpour *et. al.* studied the permeation of the anti-inflammatory drugs naproxen and Relafen, which have relatively similar properties to a phenylalanine dipeptide, through a lipid bilayer consisting of 128 1,2-dimyristoyl-*sn*-glycero-3-

phosphocholine (DMPC) molecules.[7] They found that the MSD from 0 to 4 ns fluctuated greatly and conducted their linear fit from 4 to 40 ns to arrive at diffusion constants of 1.463×10^{-10} m^2s^{-1} for Relafen and 0.863×10^{-10} m^2s^{-1} for naproxen in water. The published plot of their MSD data displays substantial positive and negative deviations from a simple linear relationship. As examined in more detail in Chapter 3, Nademi *et. al.* calculated lateral diffusion coefficients of paracetamol in the center of a lipid bilayer, and their MSD curves followed a similar pattern.[8] In contrast, Hansen *et. al.* conducted a rather comprehensive study of water diffusion within a lipid bilayer, and their MSD data at various depths – even with a large collection of water molecules and superior statistical sampling of 100 trajectories – displayed deviations from linearity until roughly 20 to 80 ps after the start of their simulation.[9] For comparison, their water diffusion constant inside of a DMPC bilayer was 1.80×10^{-10} m^2s^{-1}.[9] On the other end of the molecular spectrum, Cascales *et. al.* studied a large molecular dye, 1,6-diphenyl-1,3,5-hexatriene (DPH), which is also a popular molecule used in undergraduate chemistry laboratories involving the particle in a box quantum mechanical model. Surprisingly, their data, albeit with a low temporal resolution of 5 ps, did follow a linear trend, and their bulk water diffusion constant for DPH was 1.4×10^{-9} m^2s^{-1}.[10] Others used Einstein's MSD method to determine diffusion constants in model membrane systems for oleic acid,[11] benzene,[12-13] ubiquinone,[14] nifedipine,[15] psoralen,[16] β-cyclodextrin,[17] 5-flurouracil,[18] other fluorescent molecules,[19] a small helical peptide (wl5),[20] and lipid A.[21]

On average, these studies indicate a few general patterns. Sometimes the MSD data is generally linear. Most of the time, however, substantial negative deviations from a linear relationship appears. Sometimes, a positive deviation is observed. All but one study fit the

data to a region where the researchers felt the data best represented a linear relationship. In the single outlying study, the MSD was fit over a very long, 100 ns trajectory. Most of the time, a fine, sub-picosecond temporal resolution is used. Finally, of the articles that compared both lateral, $D_{xy}(z)$, and transverse, $D_z(z)$, diffusion (the MSD method is the overwhelming favorite for calculating lateral, $D_{xy}(z)$, diffusion constants), all but one observed an isotropic relationship between $D_z(z)$ and $D_{xy}(z)$.

In our opinion, the difficulties with this method arise from the positive (super-diffusion) and negative (sub-diffusion) deviations from linearity. Sub-diffusion has long been experimentally witnessed within living organisms and arises from crowding inside complicated biological environments. For example, the diffusion of a molecule within the cytoplasm typically follows a linear relationship for small timescales, then the mean squared displacement begins to level off as the molecule of interest encounters much larger chemical structures that do not move quickly, behaving in a manner markedly different from typical solvents.[22-23] Within the context of a lipid bilayer, a small molecule can undergo Brownian diffusion over short distances; however, as it moves greater distances (after longer times) it begins to push against lipid molecules, thereby either slowing down its motion or completely reversing its trajectory.[24] This scenario generally occurs on a nanosecond timescale; however, according to Balbo *et. al.*, normal diffusive motion returns on microsecond timescales, and these effects are averaged out.[25] In contrast, permeant diffusion is often characterized by a combination of both Brownian motion and a series of jumps between transient void spaces within the lipid bilayer.[26] Conversely, super-diffusion can occur when a permeant encounters a transient void space and undergoes near ballistic motion into a vacant pocket.

The second most common method for determining diffusion constants is a variation of the Fluctuation-Dissipation theorem that has been modified to make use of the constraint forces within umbrella sampling simulations:[1, 27]

$$D_z(z) = \frac{(k_B T)^2}{\int_0^\infty \langle \Delta F_z(z,0) * \Delta F_z(z,t) \rangle dt} \tag{4.5}$$

where k_B is the Boltzmann constant; T is the temperature; and $\Delta F_z(z, t)$ is the deviation of the z-component of the force experienced on the center of mass of the permeant from its average value, as a function of the permeant depth, z, and time, t. Note that the angled brackets represent an autocorrelation function that is an average over initial starting times. It's unclear as to why this method is applicable to pulling forces from constrained umbrella sampling simulations, and we have also used this method with forces from free, unconstrained simulations. The only explanation provided in the literature is a vague argument concerning these effects either being small or canceling each other out. Regardless, the main advantage of this method is that force data is already available, as it is also required to calculate the potential of mean force, greatly simplifying data analysis and storage. Therefore, it is most commonly used in simulations that also calculate position dependent diffusion constants for use with the inhomogeneous solubility diffusion model. For example, Equation 4.5 has been used to determine diffusion constants for tryptophan,[3, 28] oleic acid,[29] water,[1-2, 30-31] other small molecules,[2, 32] model pharmaceutical compounds,[33] and nanoparticles.[34] Only Fiedler *et. al.*, who studied nanoparticle permeation, discussed their calculation procedure for this method: they first averaged 7 – 10 force autocorrelation functions, then fit the averaged autocorrelation

function to a two-parameter exponential decay function, and finally integrated the fitted function to evaluate the integral in Equation 4.5.

The third most popular method involves taking the Laplace transform of Equation 4.5 and manipulating it in order to convert the force fluctuation autocorrelation function into a position fluctuation autocorrelation function. It was first devised by Woolf and Roux[35] and was then modified by Hummer to arrive at the following equation:[36]

$$D_z(z) = \frac{\langle \delta Q(t)^2 \rangle^2}{\int_0^\infty \langle \delta Q(t) \delta Q(0) \rangle dt} \tag{4.6}$$

where

$$\delta Q(t) = Q(t) - \langle Q(t) \rangle \tag{4.7}$$

and where *Q(t)* is the position of the permeant along the reaction coordinate, which is in the *z* direction for membrane permeation studies, as a function of time; and, where *δQ(t)* is the deviation of the current value of *Q(t)* from the average value of *Q(t)*. Hummer certifies that this equation is exact for an overdamped harmonic oscillator; however, it is unclear as to precisely why a membrane permeation process can be modeled in this manner. This method has been used to study the passive permeation of H_2S,[37] small ions,[38] water,[39] and a wide variety of model drug compounds.[40] None of these articles provides a description concerning how they calculated their position autocorrelation functions; however, Vorobyov *et. al.*, who studied small ions and who provided a large amount of supplementary material, did mention that this method is prone to unusual, oscillatory behavior.

Green-Kubo expressions round up the last of the autocorrelation methods and represent the fourth method for calculating diffusion constants in passive membrane permeation processes. Oddly enough, given its ubiquity in other fields, this method is rarely used. Green-Kubo relations are a modification of the Einstein MSD method and utilize velocity autocorrelation functions:[27]

$$D_z(z) = \int_0^\infty \langle\, v_z(t_0)\, v_z(t_0 + t)\, \rangle_{t_o}\, dt \tag{4.8}$$

where t_0 is the starting time of the simulation, t is the current time, v_z is the velocity in the z direction, and the angled brackets indicate an autocorrelation function. For lateral diffusion coefficients, $D_{xy}(z)$, this becomes:

$$D_{xy}(z) = \frac{1}{2}\int_0^\infty \langle\, \boldsymbol{v_{xy}}(t_0) \cdot\, \boldsymbol{v_{xy}}(t_0 + t)\, \rangle_{t_0}\, dt \tag{4.9}$$

where $\boldsymbol{v}_{xy}$ is the velocity vector of the permeant within the xy plane of the simulation box. Please note that this is a vector and that the dot in the middle of this equation represents the vector dot product. Finally, for diffusion in three dimensions, this equation is:

$$D_{3D}(z) = \frac{1}{3}\int_0^\infty \langle\, \boldsymbol{v}(t_0) \cdot\, \boldsymbol{v}(t_0 + t)\, \rangle_{t_o}\, dt \tag{4.10}$$

where $\boldsymbol{v}$ is the velocity vector for the permeant in Cartesian coordinates.

Only two articles have used this method to determine diffusion constants. Joshi *et. al.* studied the diffusion of water across membrane pores created by oscillating electric fields. They chose this method due to the rapid decay of the velocity autocorrelation function and noted a strong backscattering effect that produced a pronounced negative autocorrelation.

Their autocorrelation functions did not exhibit exponential decay, as predicted for Brownian motion, due to the effect of strong intermolecular interactions and memory effects. Sugii *et. al.* studied the permeation of small gas molecules – O_2, CO, NO, and H_2O – and chose this method due to its lower computational cost and faster convergence. In addition, all of the velocity autocorrelation data could be used, as opposed to MSD data that has to be fit within an 'optimal' range that is subjectively chosen by the researcher.

Finally, the remaining methods employed in recent literature involve numerical solutions to various mathematical models. For example, Comer *et. al.* utilized a Bayesian inference scheme to reverse engineer diffusion constants from an adaptive force biasing simulation.[41] Holland *et. al.* also devised an improved, oscillating forward/reverse method to determine diffusion constants from steered molecular dynamics simulations.[42] Unfortunately, both of these methods do not apply to either free or constrained molecular dynamics simulations. Therefore, we'll focus on the next and final method considered in this section: numerically integrating the Smoluchowski equation. The Smoluchowski equation describes the time evolution of the probability density, *p(Q, t)*, along a diffusion coordinate, *Q*, and is:[36]

$$\frac{\partial p(Q,t)}{\partial t} = \frac{\partial}{\partial Q}\left\{D(Q)e^{-\beta G(Q)}\frac{\partial}{\partial Q}\left[e^{\beta G(Q)}p(Q,t)\right]\right\} \qquad (4.11)$$

where *G(Q)* is the potential of mean force, and *D(Q)* are the position dependent diffusion constants. Our numerical approach uses the method described by Hummer[36] that is, in turn, based on ideas developed by Zusman[43] and Bicout and Szabo.[44] Here, the locations of the permeant center-of-mass, *z(t)*, are histogrammed within each umbrella sampling window.

From these histograms, we can calculate the local, biased probability distributions across each bin, as well as the transition rates between neighboring bins. The diffusion rates, *D*, can then be obtained from:

$$D_{n+1/2} = w_{n+1,n}\left[\frac{p^*(n)}{p^*(n+1)}\right]^{1/2} d^2 \qquad (4.12)$$

where *n* is the bin number, $w_{n+1,n}$ is the transition rate between neighboring bins, *p** is the biased probability distribution, and d is the width of each bin. The primary advantage of this method is that the actual umbrella sampling simulation data is used and absolutely zero autocorrelation functions are calculated.

The primary objective of this chapter is to utilize as many of the above methods, as is practical, to calculate diffusion constants – mainly $D_z(z)$, but also $D_{xy}(z)$ and $D_{3d}(z)$, depending on the method. Calculated diffusion coefficients will then be compared, as much as is practical, and commentary will be provided concerning the ease of calculation and convergence. In addition, we attempt to provide as many answers as possible to the lingering questions from the beginning of this Introduction. We chose the same chemical systems that were used in Chapter 3 as diffusion test subjects; namely, the phenylalanine dipeptide (N-acetylphenylalanineamide) passively permeating through lipid bilayers consisting of 40 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) lipid molecules, 50 POPC lipid molecules, or 50 1,3-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) molecules.

Specifically, diffusion constants were calculated with the following methods and sets of data: 1.) the Fluctuation Dissipation theorem, Equation 4.5; 2.) Green-Kubo expressions,

Equations 4.8, 4.9, and 4.10; 3.) Einstein Relations, Equations 4.3; 4.) the Hummer Displacement method, Equation 4.6; and 5.) Numerical Approximation of the Smoluchowski Equation, Equation 4.12. For Method 1, diffusion constants were calculated using pulling forces from our umbrella sampling simulations, from forces in the *z*-direction from separate z-position restrained simulations in CHARMM, and from forces in the *z*-direction from free, completely unrestrained CHARMM simulations. These different methods provide insight into the validity and effect of constraining the position of the permeant. For Method 2, diffusion constants were calculated from free GROMACS simulations in the *z*-direction, the *xy*-plane, and in 3D space; these sets of data provide insight into the isotropic nature of passive membrane permeation processes. For Method 3, diffusion constants were only calculated from data from free simulations in GROMACS in the *z*-direction and highlight the challenges of this method. For Method 4, only position data from our umbrella sampling simulations were used. The same is true for Method 5, as both of these methods are based on umbrella sampling simulations. Finally, all methods are compared and contrasted.

4.3 Methods:

GROMACS umbrella sampling simulations were used for three methods: the Fluctuation-Dissipation theorem, the Hummer Displacement method, and the Numerical Solution to the Smoluchowski Equation method. The methodology for obtaining the relevant umbrella pulling forces and displacement data is the same as those used in Chapters 2 and 3.

Our simulation windows consisted of one phenylalanine dipeptide, also known as N-acetylphenylalanineamide, Ac-Phe-NH2, or NAFA; water; and a phospholipid bilayer consisting of either 1,3-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) or 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC). Specifically, three systems were studied and were classified based upon the number and type of lipid molecules comprising the bilayer. The DOPC 50 system contained the following molecules: 1 phenylalanine dipeptide, 50 DOPC lipids, 2991 TIP3P waters, 8 sodium atoms, and 8 chlorine atoms, for a grand total of 15,764 atoms. The POPC 50 system contained 1 phenylalanine dipeptide, 50 POPC lipids, 2996 TIP3P waters, 6 sodium atoms, and 6 chlorine atoms, for a grand total of 15,578 atoms. Finally, the POPC 40 system contained 1 phenylalanine dipeptide, 40 POPC lipids, 2483 TIP3P waters, 7 sodium atoms, and 7 chlorine atoms, for a grand total of 15,578 atoms. Sodium and chlorine ions were added to maintain a physiologically relevant ionic strength. All systems were confined to tetragonal boxes of dimensions: 4.29 × 4.29 × 8.30 nm for DOPC 50, 4.15 × 4.15 × 8.74 nm for POPC 50, 3.70 × 3.70 × 8.92 nm for POPC 40, corresponding to respective lipid headgroup areas of 0.736 nm^2, 0.689 nm^2, and 0.685nm^2. For the DOPC 50 system, 0.736 nm^2 is in close agreement with an experimental average of 0.723 nm^2,[45-46] and the electron density profile is similar to one obtained from X-ray scattering experiments.[46-47] Experimentally, POPC headgroup areas are 70.5 nm^2, which are close to our simulated values.[48] Unfortunately, experimental data for a POPC electron density profile does not exist; however, our data favorably compares to an experimentally determined DOPC electron density profile,[46] and a plot of this information (Figure 3.13) can be found in Chapter 3. The *x* and *y* axes of the simulation box are parallel to the plane of the lipid bilayer, and the *z* axis is defined as being perpendicular to the plane of the

lipid bilayer. The initial coordinates for all systems, including the phospholipid bilayer, were determined by using CHARMM-GUI.[49-51]

Molecular dynamics simulations for the DOPC 50 system were carried out with GROMACS 4.5.4; simulations for the POPC 50 and POPC 40 systems were carried out with GROMACS 4.5.6.[52] The following options were chosen within GROMACS for all simulations. Forces for DOPC, POPC, and the phenylalanine dipeptide were all represented by the version 36 CHARMM force field.[53-54] Water was represented by using the TIP3P model.[55] Electrostatic interactions were truncated at 0.13 nm, and a long-range correction was added with the Particle-Mesh Ewald method.[56] Van der Waals interactions were truncated at 1.2 nm; however, a switching function was used to smooth the transition between 1.0 and 1.2 nm. A canonical ensemble was used; and, therefore, the number of system particles, the system volume, and temperature were constant. Temperature was held constant at 300K by using velocity rescaling.[57] Initial velocities were sampled from a Maxwell distribution at 300K. The default leap-frog algorithm was used to integrate the relevant equations of motion with a time step of 2 fs. Periodic boundary conditions in the *x*, *y*, and *z* simulation box directions were used. All bond constraints were enforced by using the LINCS algorithm.[58] Position, energy, and other simulation data were saved every 500 time steps and once every picosecond.

Simulations began by equilibrating each system, as setup by CHARMM-GUI, over 500 ps intervals with the parameters described in the previous paragraph. Once equilibrated, an unrestrained simulation was run for at least 50 ns, and a system image was obtained at $z = 1.6$ nm. This image was then used as the starting point for further umbrella sampling windows in both positive and negative *z* directions. Umbrella sampling windows for all systems were run

from z = 0.0 nm to z = 3.0 nm in 0.1 nm increments. A harmonic restraining potential of 3000 kJ mol^{-1} nm^{-1}, located at the center of each umbrella sampling window, was applied to the center of mass of the phenylalanine dipeptide.[59-60] Each window ran from 50 to 100 ns; as needed, simulation windows near the center of the membrane were extended up to 100 ns in order to improve statistical overlap for potential of mean force calculations.

Equation 4.5 was then used to calculate position dependent diffusion constants from the constrained, umbrella pulling forces that were run with GROMACS. Force fluctuations over the first 50 ns of each window, with a temporal resolution of 20 fs, were calculated and were then split into 5, 10 ns time series for each window in order to calculate statistical error for each position. Autocorrelation functions for each 10 ns piece, per window, were calculated using a Fortran program. Autocorrelation functions were then numerically integrated using the trapezoid rule until reaching a maximum, stable value, as chosen by our custom made integration picking routine, also written in Fortran, and confirmed manually by a random sampling of autocorrelation functions. The same procedure was then repeated using Equation 4.6; where, instead of using pulling forces, the position of the permeant over the first 50 ns of the trajectory was used to calculate actual forces acting on the center of mass in the absence of restraints, instead, to create the desired autocorrelation functions. Again, position data for each 50 ns trajectory was split into 5, 10 ns pieces and autocorrelation, integration, and convergence were determined in the same manner as before

Finally, our GROMACS, umbrella sampling simulations were used with Equation 4.12 to calculate diffusion constants one, final time. Specifically, this approach begins by binning $z(t)$, the distance between the center of mass of the phenylalanine dipeptide and the center of mass

of the lipid bilayer, over the course of the entire time series. Similar to before, *z(t)* was split into 4 separate pieces of equal length in order to assess statistical error. The biased probability distribution, *p*(z)*, is then approximated by the resulting, normalized histogram as *p*(n)*. The average number of transitions between neighboring bins is then calculated from the position time series *z(t)*. Next, $w_{n+1,n}$ is calculated from the average number of transitions between neighboring bins divided by the bin residence time. We chose a bin width of 0.025 nm so that a bin transition is located at the center of any given umbrella window and in order to minimize error. Using these parameters, the position dependent translational diffusion coefficient, $D_z(z)$, for each of our umbrella windows is then found when $n = 1$. Please note: although we are using our harmonically restrained umbrella sampling windows, the diffusion coefficient is calculated at the center of each window where the restraining potential is equal to zero, thereby minimizing the impact of the harmonic umbrella sampling constraint.

Next, a custom made version of CHARMM v. 38 was used to calculate diffusion constants at fixed *z*-values by using both Equations 4.5 and 4.9. CHARMM simulations constrained in the *z*-direction were run with the following parameters, which are the same as those used in Chapter 2. For each umbrella sampling window, ten independent molecular dynamics trajectories were run with the same CHARMM v. 36 protein and lipid parameters as were used in the GROMACS simulations.[61] The starting images were extracted from the entire course of the umbrella sampling trajectories in order to sample a greater variety of initial, starting systems. In other words, if ten images were extracted from a 50 ns umbrella trajectory, then a single image was selected from each 5 ns subdivision of the overall, 50 ns trajectory. The constrained CHARMM simulations were then run for 20 ps with a time step of 2 fs. Velocity and

coordinate data were recorded every 2 fs. Force data was then calculated using the coordinate data in a separate calculation once molecular dynamics simulations were completed. Temperature was kept constant at 300K by using a Nosé-Hoover thermostat.[62-63] Where possible, remaining molecular dynamics parameters were set to match those within the GROMACS umbrella simulations. Force autocorrelation functions were then calculated in the same manner as before, before arriving at the desired diffusion constants. Velocity autocorrelation functions, on the hand, were integrated until they crossed the *x*-axis twice.

Unconstrained, free simulations were used to calculate diffusion constants with both GROMACS and CHARMM. Free, unconstrained CHARMM simulations were used to calculate diffusion constants using Equation 4.5. The same parameters and settings were used as those listed in the previous paragraph, except that an unmodified, normal version of CHARMM v.38 was used. Force data, autocorrelation functions, and integration were then conducted in the same manner, too. Finally, free GROMACS simulations were run in the same manner as the GROMACS umbrella sampling runs, except with a few changes. Namely, the umbrella sampling potential was removed and 10 separate simulations of 20 ps in length were run with velocity and position data being saved after every 2 fs time step. This data was then used to calculate diffusion constants using Equations 4.3, 4.8, 4.9, and 4.10. Autocorrelation functions were calculated and integrated as before. However, for one set of data, the GROMACS g_velacc program was used to calculate autocorrelation functions. Position data was then used with Equation 4.3 to calculate diffusion constants by plotting the mean square displacement as a function of time and by fitting data from approximately 1 to 5 ps to the equation of a line. The

slope of said line is then proportional to the diffusion constant. Occasionally, the fit was only conducted from 1 to 3 or 4 ns due to oscillatory behavior that was manually observed.

4.4 Results and Discussion:

4.4.1. – The Fluctuation Dissipation Theorem.

The most common method for calculating diffusion constants is with the Fluctuation Dissipation theorem applied to umbrella sampling pulling forces. Figure 4.1, shown below, depicts these results.

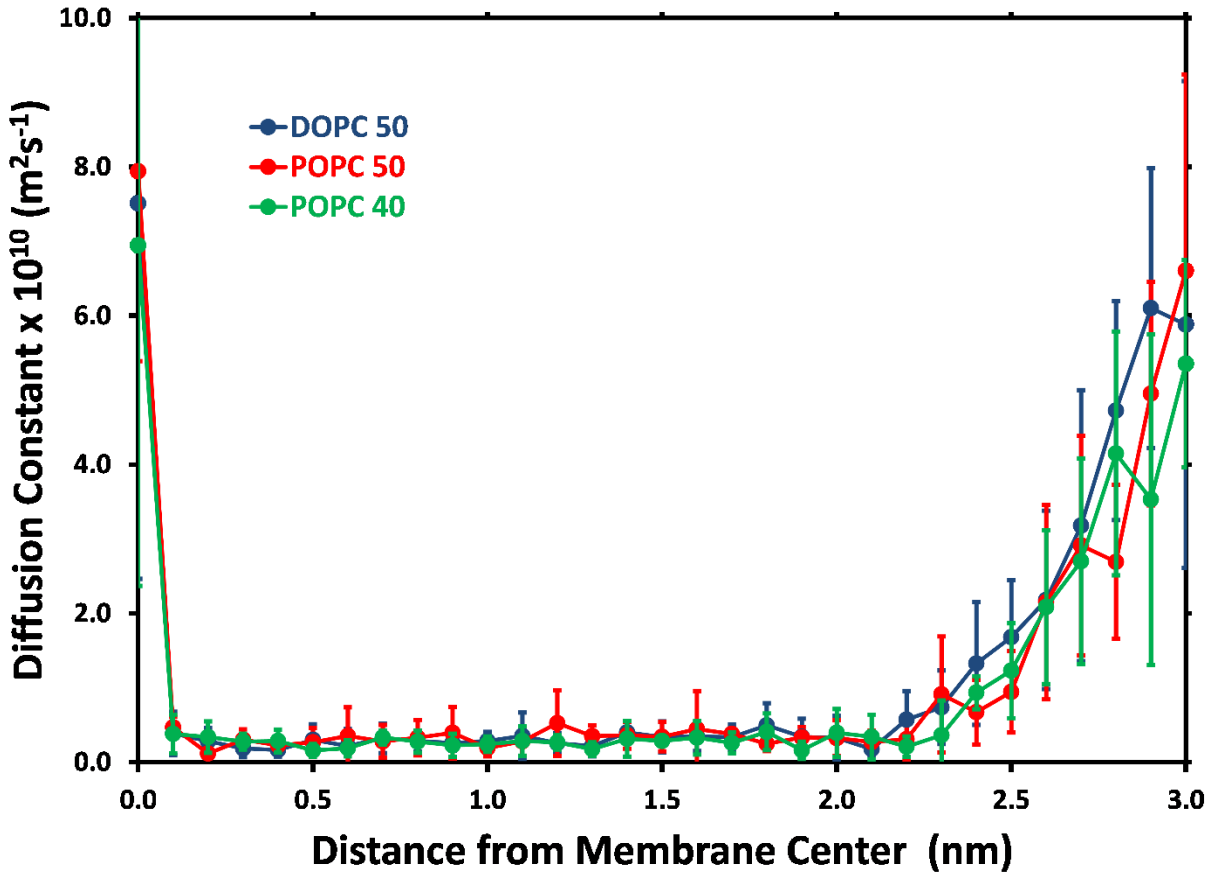


Figure 4.1. *$D_z(z)$* calculated from the Fluctuation-Dissipation theorem using umbrella sampling pulling forces from GROMACS simulations. Diffusion constants are for the passive permeation of the phenylalanine dipeptide through DOPC 50 (blue), POPC 50 (red), and POPC 40 (green) lipid bilayer systems. Error bars represent a 95% confidence interval around the mean value from 5 contiguous pieces of the overall, 50 ns trajectory.

Diffusion constants in the aqueous region are around 6×10^{-10} m^2s^{-1}, which is a common result for medium sized permeants like the phenylalanine dipeptide. As the permeant nears the interfacial region of the membrane, the diffusion constants decrease dramatically to around 2×10^{-11} m^2s^{-1}. These diffusion constants, on the order of 10^{-11}, are rarely reported in the literature for any but the largest of permeants, such as the work that dos Santos and Eriksson conducted on a tricyclic pharmaceutical compound, psoralen.[16] In general, one would expect transverse

diffusion to dramatically decrease near the lipid-water interface; however, such a result has not been witnessed in any other studies of this type. If a decrease is witnessed relative to the center of the membrane and the aqueous region, it is generally within a factor of 2 or 3, not an entire order of magnitude or greater. Diffusion constants then increase to around 7.5×10^{-10} m^2s^{-1}, which is around the same value as in the aqueous region. This result is very typical from the aqueous region to membrane center: a slight decrease in diffusion constants before then increasing and has been attributed to the formation of transient void spaces where the tails of opposing lipid molecules interdigitate.

The anomalous diffusion constants in the center may be caused by a variety of factors. This method uses the umbrella sampling pulling forces to create the force autocorrelation functions in Equation 4.5 that are then integrated. In the aqueous region, the permeant is not confined and is free to wander about in either direction; therefore, forces should fluctuate in a random manner and result in a poor autocorrelation and larger diffusion constants. However, at the interface and within the lipid bilayer, the permeant is trapped in a small region of space. Furthermore, the forces that it experiences are directional in nature. For example, if the permeant is inside the lipid bilayer at z = 0.8 nm, it would experience a force pulling it towards the interfacial region at x = 1.3 nm where the potential of mean force is at a minimum (see Figure 3.2 in Chapter 3) and where the hydrophilic and hydrophobic moieties of the dipeptide can associate with either water or the lipid molecules. As a result, the umbrella pulling forces are also unidirectional in the opposite direction, and any random fluctuations would be minimized, thereby resulting in a stronger force fluctuation autocorrelation: fluctuations are not based on the random motions of nearby atoms and molecules, but are coupled to the much

slower motions of the lipid bilayer. This would result in smaller than anticipated diffusion constants. However, it remains unclear as to how, precisely, the umbrella potential is attenuating this affect, otherwise, other researchers using this method would observe the same precipitous drop as seen in Figure 4.1.

In order to determine if the umbrella potential is responsible for the aforementioned anomaly, a custom made version of CHARMM v. 38 was used to fix the *z*-position of the permeant by gently nudging the velocity of the permeant in the appropriate direction; forces were not directly influenced. Using this force data and Equation 4.5, diffusion constants were calculated and plotted in Figure 4.2, as seen below.

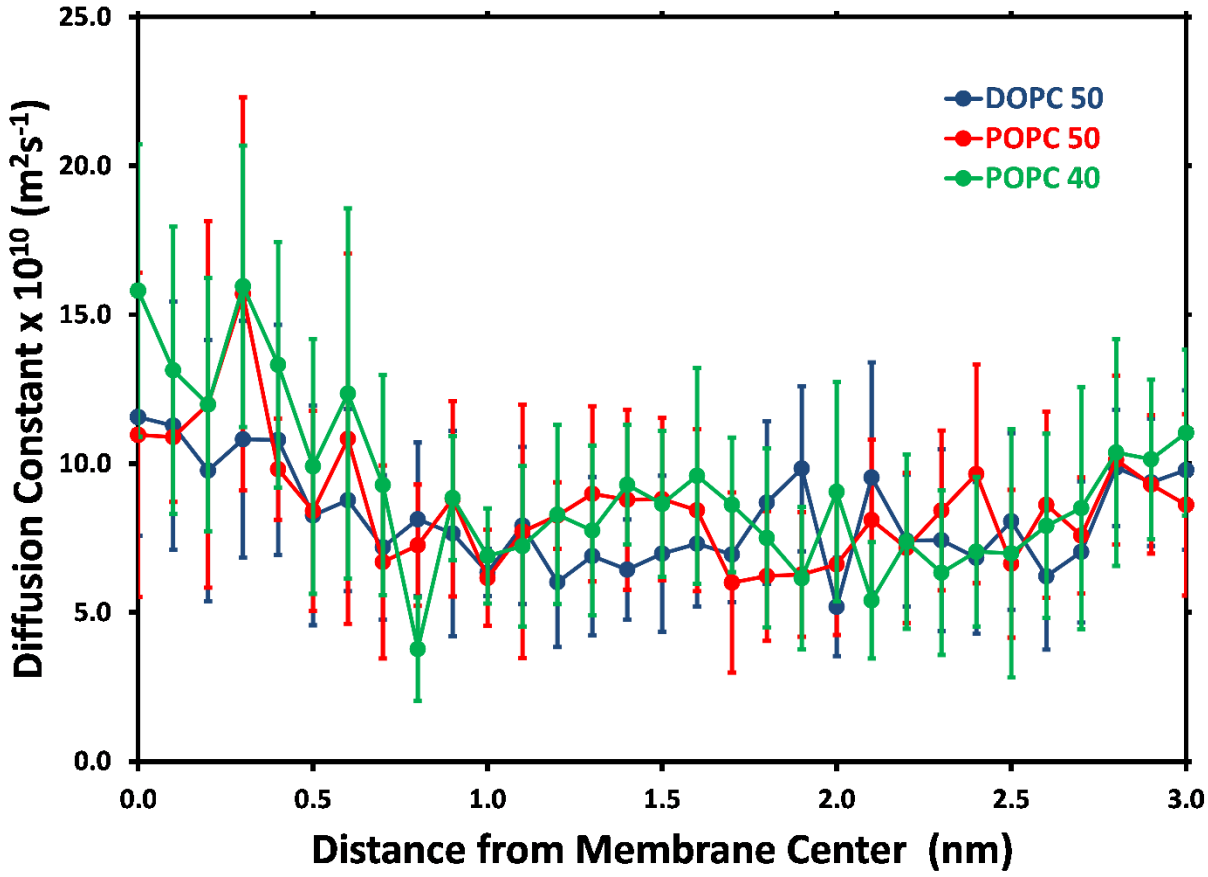


Figure 4.2. *$D_z(z)$* calculated from the Fluctuation-Dissipation theorem using force data from a *z*-constrained CHARMM simulation. Diffusion constants are for the passive permeation of the phenylalanine dipeptide through DOPC 50 (blue), POPC 50 (red), and POPC 40 (green) lipid bilayer systems. Error bars represent a 95% confidence interval around the mean value from 10 separate trajectories of 20 ps.

In the aqueous region, diffusion constants are around 10×10^{-10} m^2s^{-1}, which is very similar to those from the umbrella pulling forces. In contrast, these values remain constant throughout the entire distance from the center of the membrane, with a slight and statistically insignificant increase in the center of the bilayer, as reported by most studies of small permeants. The precipitous drop in the central region has vanished! This provides some evidence that the umbrella potential may be accentuating the central diffusion constants. Alternatively, the difference may be due to changes in time scale. The force data for these results were obtained

as an average over several 20 ps simulations; whereas, the data in Figure 4.1 is from umbrella pulling forces over a 10 ns piece of a 50 ns umbrella window. For example, a typical phosphatidylcholine lipid molecule exchanges positions with its nearest neighbor lipid, roughly, once every 50 ns.[64] A calculation spanning 10 ns may sample parts of this motion; however, it is unlikely that a 20 ps simulation would sample this motion. Therefore, assuming that the umbrella pulling force fluctuations are related to the unconstrained force fluctuations, the difference between these two methods can be explained. In Figure 4.1, force fluctuations can be associated with the macroscopic lipid motions; whereas, in Figure 4.2, they are likely not, thereby weakening the autocorrelation and increasing the value of the diffusion constants.

However, in the previous case, the position of the permeant is still fixed in an artificial manner. For the next case, the same CHARMM simulations were run, except that the position of the permeant was not fixed, and multiple short trajectories of the free phenylalanine dipeptide were generated, starting at the center of each umbrella window, with the peptide free to wander about wherever it so pleases. These data are plotted in Figure 4.3, as seen below.

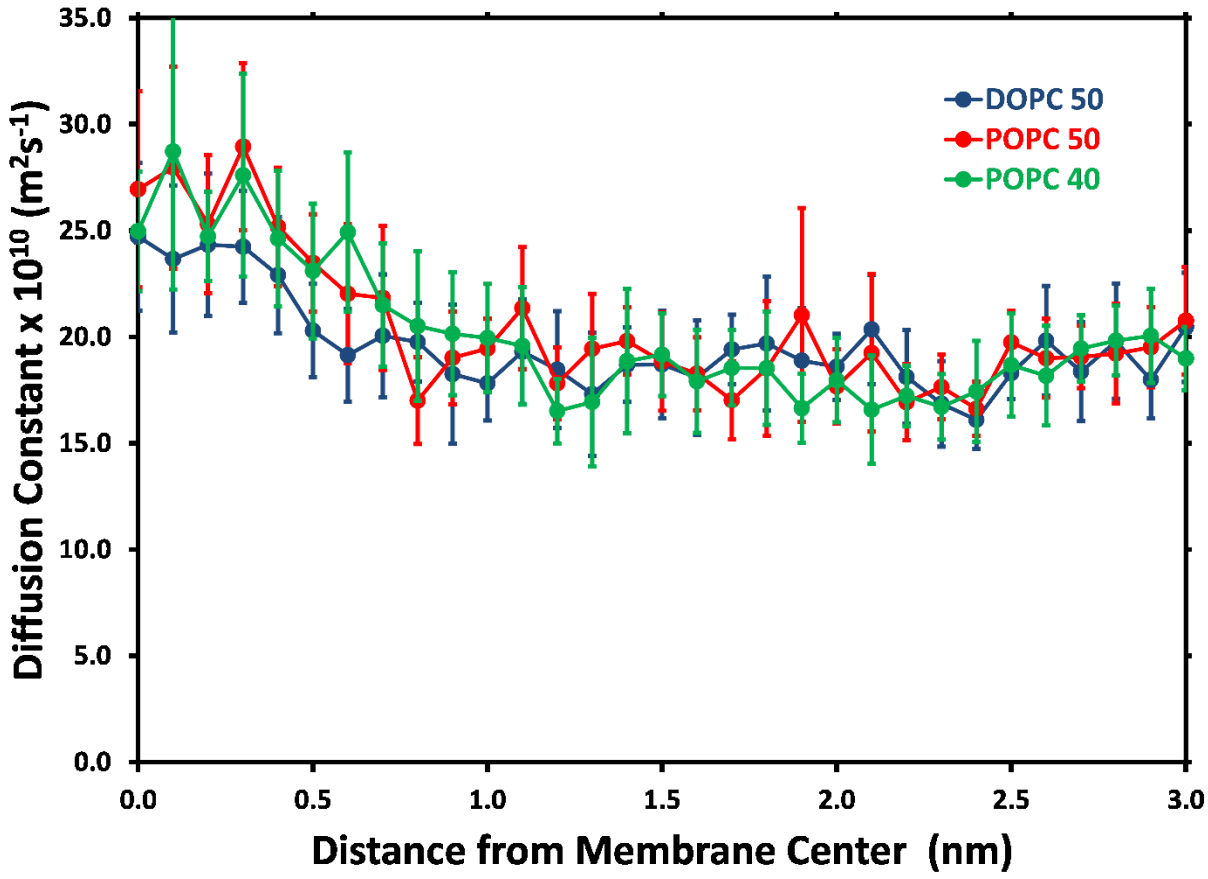


Figure 4.3. *$D_z(z)$* calculated from the Fluctuation-Dissipation theorem using forces from a free, unconstrained CHARMM simulation. Diffusion constants are for the passive permeation of the phenylalanine dipeptide through DOPC 50 (blue), POPC 50 (red), and POPC 40 (green) lipid bilayer systems. Error bars represent a 95% confidence interval around the mean value from 10 separate trajectories of 20 ps.

Diffusion constants start out at around 20×10^{-10} m^2s^{-1} and remain constant before increasing to around 25×10^{-10} m^2s^{-1} within the interior of the membrane, which, for POPC 40 and DOPC 50 is a statistically significant difference. It is worth noting that, in this case only, the starting *z*-location of the peptide is well determined. In the course of the 20 ps trajectories, the peptide explores a small range of *z*-values around this starting position (roughly 0.1 nm.) Compared to the *z*-constrained CHARMM simulations, these values are roughly twice as large, indicating that

the presence of the *z*-constraint reduces the diffusion constants because nearly all other potential variables, such as the timescale, were kept constant. As with the umbrella pulling forces, it is not entirely clear why a constraint would increase the autocorrelation of the force fluctuations and therefore lower the diffusion constants. Alternatively, the permeant was allowed to roam freely and follow the gradient of the forces to lower energy states. This could expose the permeant to a larger number of random force fluctuations and increase the calculated diffusion constant by weakening the autocorrelation. In contrast, the constraining velocity and force adjustments of the previous two cases, may keep the permeant in energetically unfavorable regions that would then strengthen the correlation of the force fluctuations.

Let's reconsider the hypothetical situation from the first case of this section: the permeant is located at z = 0.8 nm, with a free energy minimum located at z = 1.3 nm. The permeant is then constrained at z = 0.8 nm by a harmonic constraining force. Over time, the permeant will most likely travel towards the interfacial region, thus straining the boundary of its constraint. Every time it does that, it experiences a push or force in the opposite direction in order to maintain its constraint. In between constraining events, forces would fluctuate in a typical manner. However, if it strays too far, it then experiences an increasingly aggressive constraining response. This constraining response would then cause a large increase in the deviation of the forces from average. Eventually, the permeant will return to its constrained *z*-position before drifting, once again. It will then experience an increasing response from the constraining method, which will cause a large fluctuation in forces, and the cycle repeats. Force fluctuations are then not correlated with only the system but are also correlated with respect

to the timing and mechanics of the constraining method. The greater correlation of the force fluctuations would then decrease the calculated diffusion constants. At this point, this is somewhat speculative, but it appears that it is important to consider these potential effects when evaluating diffusion constant data in constrained simulations.

4.4.2. – Green-Kubo Relations.

At first glance, Green-Kubo relations appeared to be a promising way to calculate diffusion constants due to the rapid convergence of velocity autocorrelation functions and short simulations times. However, Figure 4.4 depicts data using Equation 4.8, as seen below.

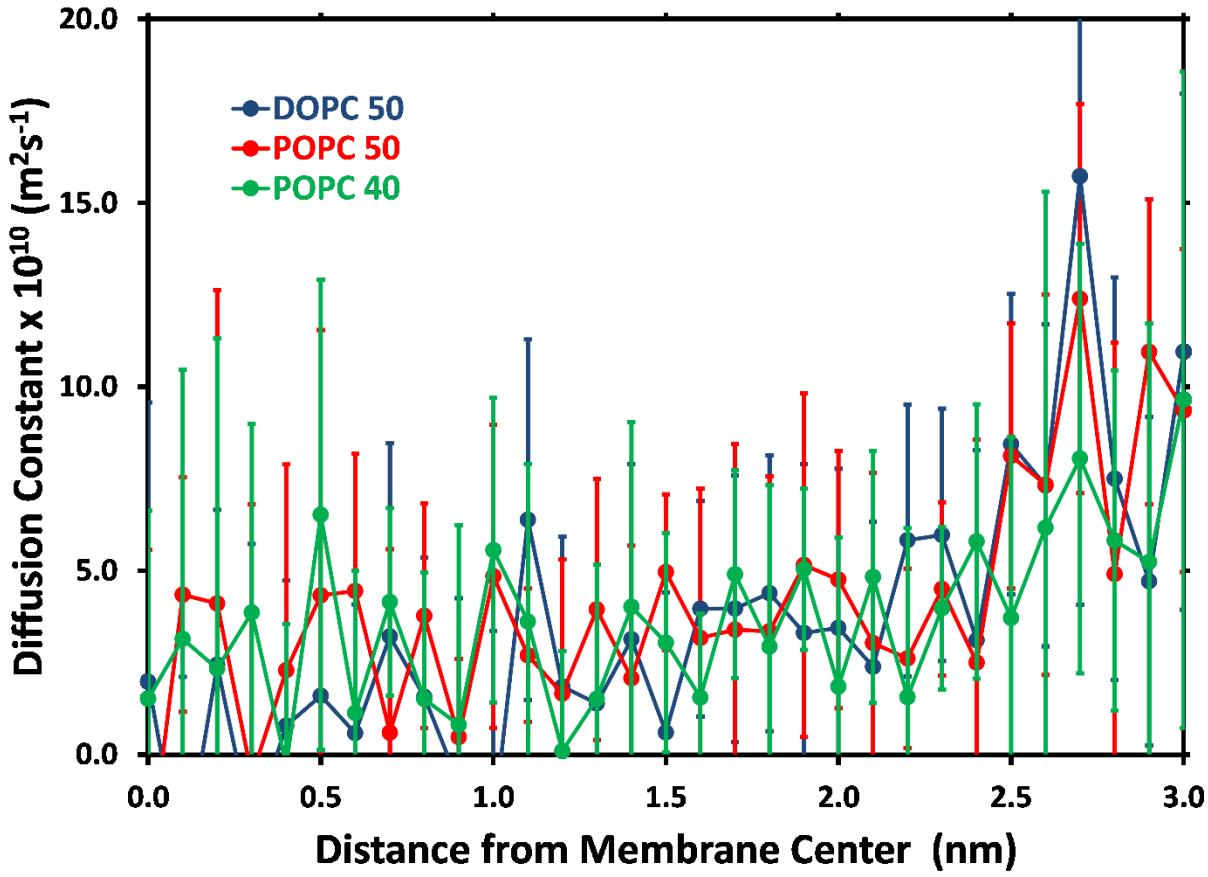


Figure 4.4. *$D_z(z)$* calculated from a Green-Kubo expression using velocity data from unconstrained GROMACS simulations. Diffusion constants are for the passive permeation of the phenylalanine dipeptide through DOPC 50 (blue), POPC 50 (red), and POPC 40 (green) lipid bilayer systems. Error bars represent a 95% confidence interval around the mean value from 10 separate trajectories of 20 ps.

Average diffusion constants vary from around 1 to 15 × 10^{-10} m^2s^{-1} which is similar to our previous results and other reported values for similar molecules. However, the confidence intervals and the uncertainty in the data vary greatly. Individual autocorrelation functions would often decay to zero within a single ps; however, the resulting integrals varied greatly over a couple orders of magnitude. Alarmingly, some autocorrelation functions were even negative, resulting in nonsensical and negative diffusion constants. Joshi *et. al.* did mention in

their study of water that they observed backscattering in their data and that these short timescales represented molecules rattling around in a cage. Anti-correlated velocities were primarily witnessed within the lipid bilayer and may represent the permeant moving forward, only to encounter a lipid molecule and being forced backward, hence resulting in a negative velocity correlation.

Data for DOPC 50 at z = 1.0 nm was particularly problematic with an average in negative territory. This data warrants a closer examination. Below, all ten of the velocity autocorrelation functions are plotted in Figure 4.5. 9 of the 10 autocorrelation functions are relatively well behaved; however, the tenth one, in light blue, decays extremely rapidly and oscillates around zero, as if the permeant is trapped in place and cannot move in either direction. Figure 4.6 indicates how these oscillations are compounded when the velocity autocorrelation function is integrated from 0 ps until the indicated time on the figure. How is such an autocorrelation function to be integrated? Or should it be discarded, resulting in an average over 9 trajectories? For all results within this Chapter, no trajectories were left behind and discarded. For velocity autocorrelations, the integration was carried out to the second zero of the autocorrelation functions so that both the first positive and negative regions are included. Obviously, this choice impacts the results and represents the best of many poor choices, such as integrating for a longer time or using maximum integrated values.

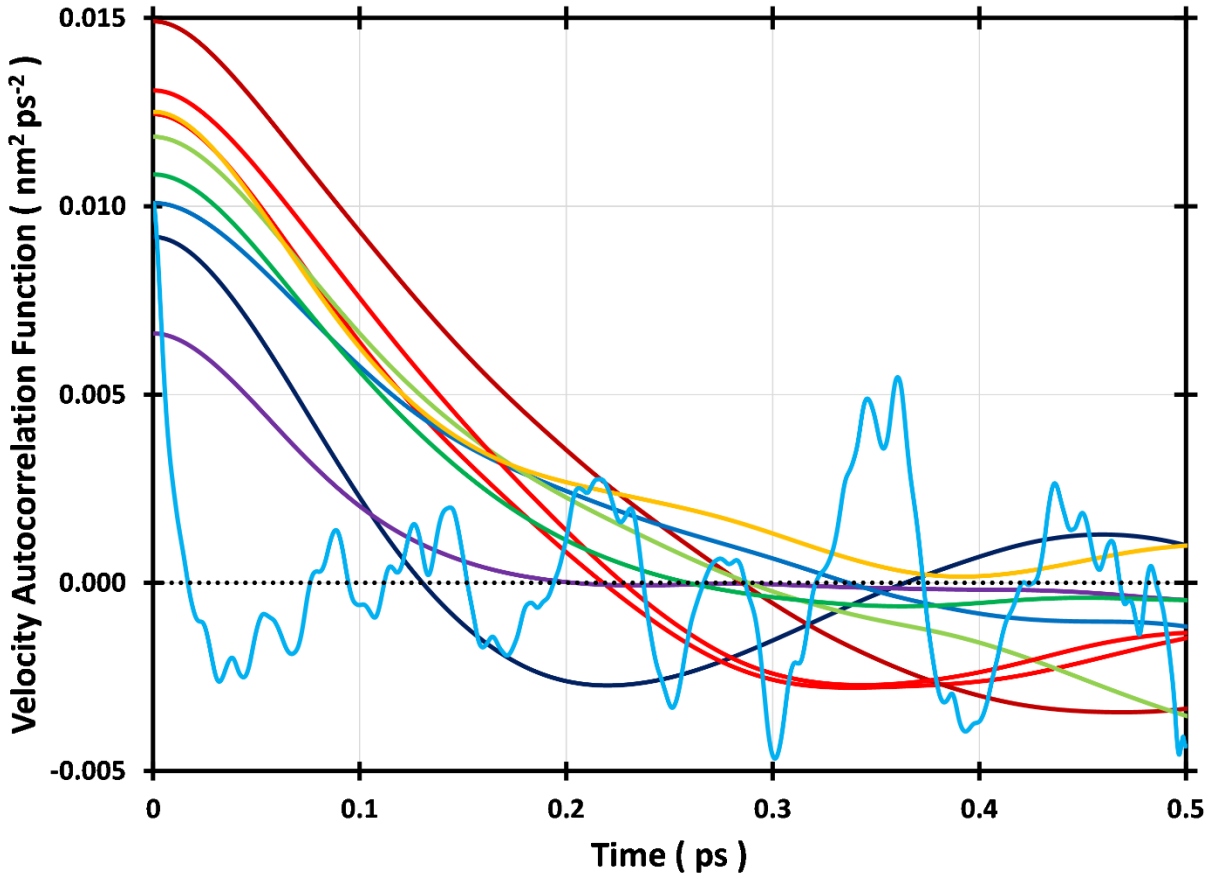


Figure 4.5. Autocorrelation functions for *z*-velocities for all ten trajectories from a 20 ps free, unconstrained GROMACS simulation starting at *z* = 1.0 nm.

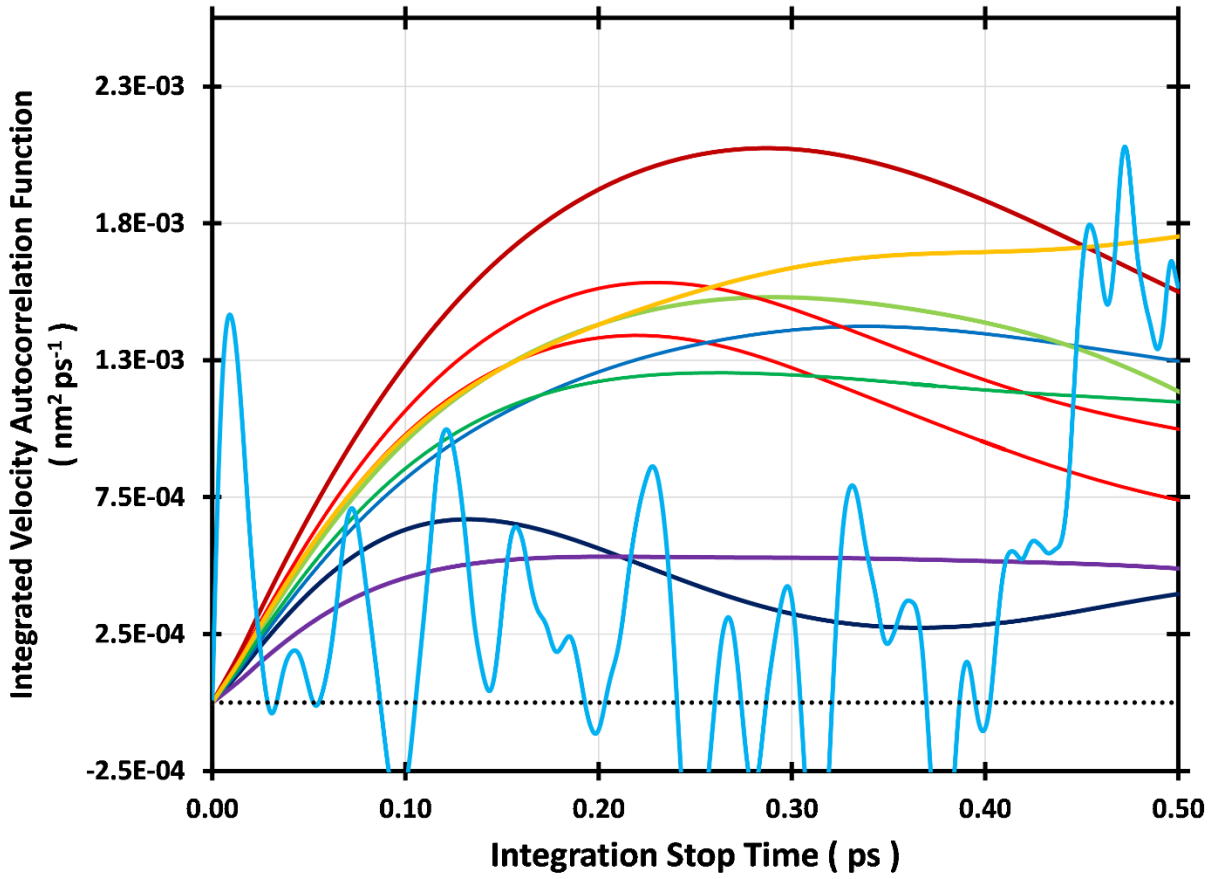


Figure 4.6. Integrated *z*-velocity autocorrelation functions for all ten trajectories from a 20 ps free, unconstrained GROMACS simulation starting at z = 1.0 nm. Velocity autocorrelation functions are integrated from 0 ps until the indicated integration stop time.

Lateral diffusion constants for either the permeant or lipid molecules within the bilayer are often reported, and the 2D Green-Kubo expression in Equation 4.9 seemed to be a natural fit. Unfortunately, these results for the phenylalanine dipeptide suffer from many of the same issues as the previous ones in the *z*-direction. Results from Equation 4.9 are plotted below in Figure 4.7. As before, average diffusion constants vary from around 1 to 15 × 10^{-10} m^2s^{-1}. It is typically reported in the literature that diffusion constants in lipid membrane systems are

independent of the direction. Again, as before, the variation in the data is very large and demonstrates a pronounced, negative autocorrelation, particularly at smaller values of *z*.

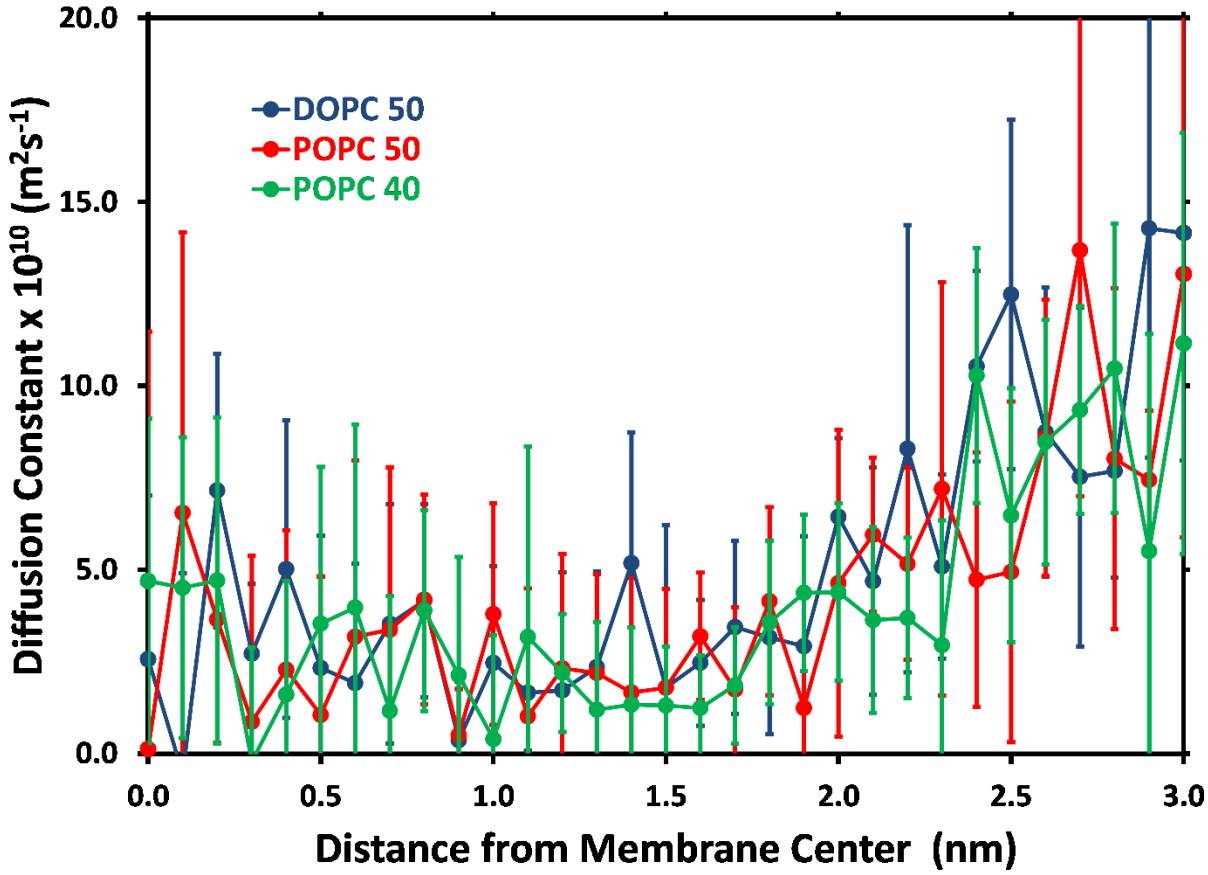


Figure 4.7. $D_{xy}(z)$ calculated from a 2D Green-Kubo expression using velocity data from unconstrained GROMACS simulations. Diffusion constants are for the passive permeation of the phenylalanine dipeptide through DOPC 50 (blue), POPC 50 (red), and POPC 40 (green) lipid bilayer systems. Error bars represent a 95% confidence interval around the mean value from 10 separate trajectories of 20 ps.

Why limit ourselves to only one or two dimensions? The data displayed in Figure 4.8 is from Equation 4.10 and represent 3D diffusion constants as a function of *z*. The same range of values and general trends were observed for $D_{3d}(z)$ as was observed for $D_{xy}(z)$ and $D_z(z)$ in the preceding paragraphs. This isotropic nature of translational diffusion in lipid bilayers makes

sense because, in the membrane interior, the permeant is confined by the surrounding lipid molecules in all directions. If it travels in either the *x*, *y*, or *z* direction, it will interact and collide with a lipid molecule.

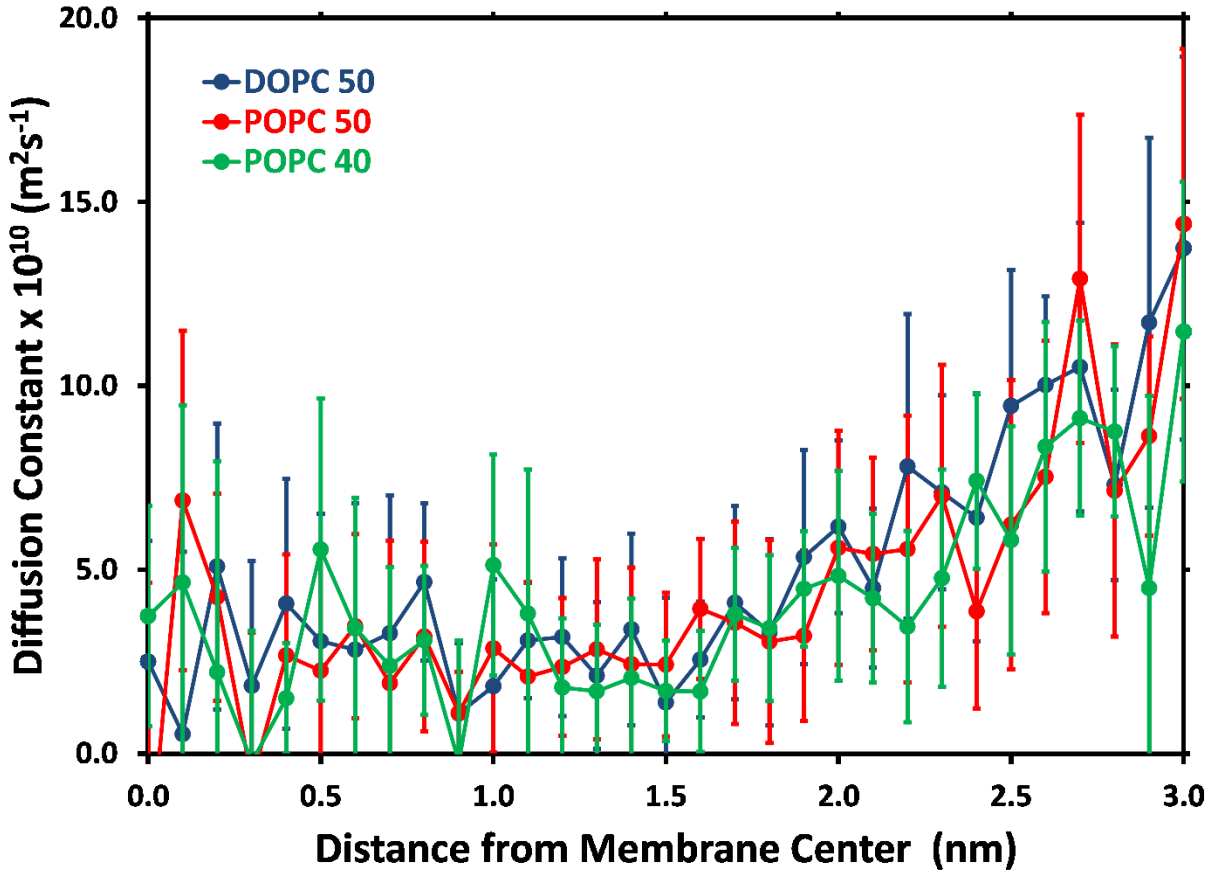


Figure 4.8. $D_{3d}(z)$ calculated from a three dimensional Green-Kubo expression using velocity data from unconstrained GROMACS simulations. Diffusion constants are for the passive permeation of the phenylalanine dipeptide through DOPC 50 (blue), POPC 50 (red), and POPC 40 (green) lipid bilayer systems. Error bars represent a 95% confidence interval around the mean value from 10 separate trajectories of 20 ps.

At the time, the presence of negative, integrated autocorrelation functions was surprising. The results in Figure 4.8 were then redone by using the built in GROMACS g_velacc

command to calculate autocorrelation functions. These results are displayed in Figure 4.9. On average, diffusion constants were slightly higher than when using our Fortran autocorrelation program; however, the results for POPC 50 in the aqueous region were markedly smaller using g_velacc. After studying the source C code for g_velacc, it is not clear as to why there should be a difference between these two figures. It is, however, reassuring that both programs yield diffusion constants around the same values of 1 to 15 × 10^{-10} m^2s^{-1} and both yield results with significant anti-correlations.

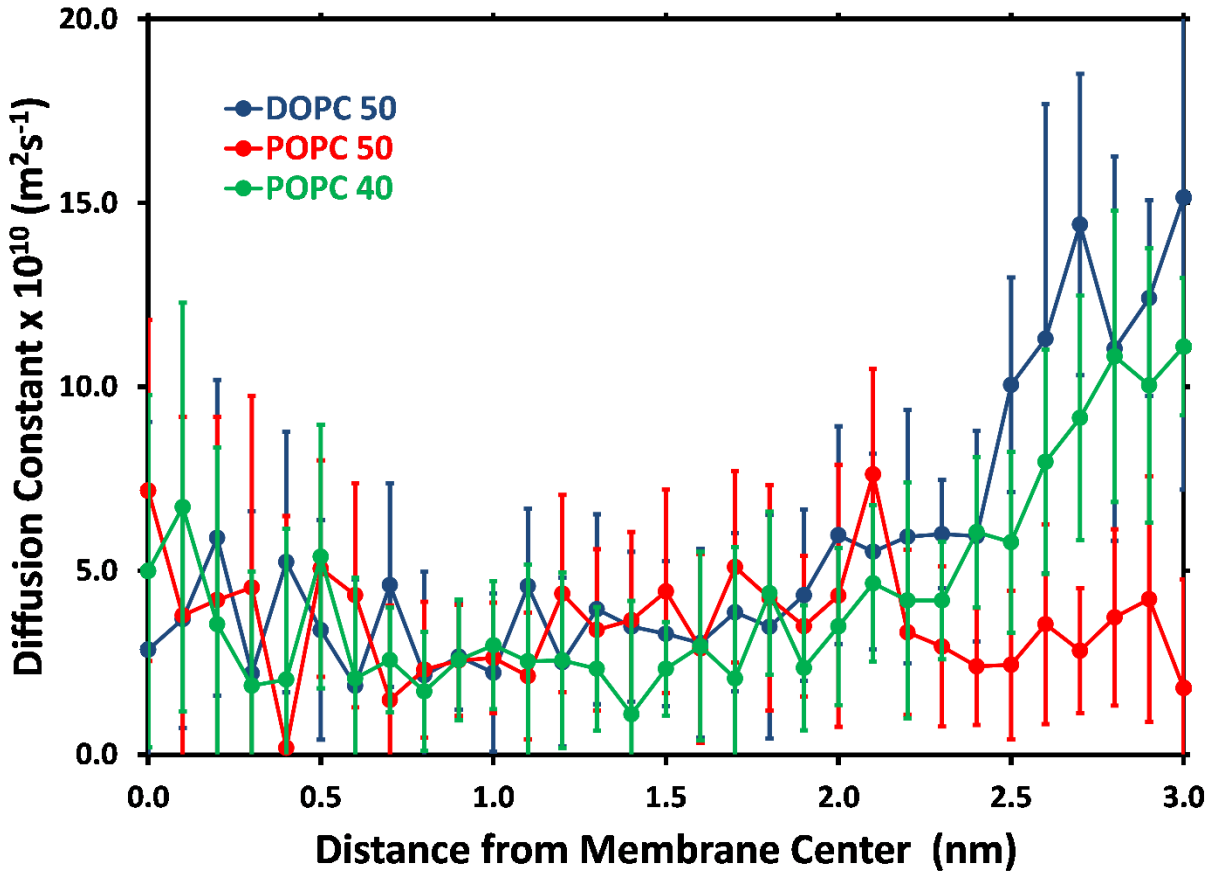


Figure 4.9. *$D_{3d}(z)$* calculated from a three dimensional Green-Kubo expression using velocity data from unconstrained GROMACS simulations. Velocity autocorrelation functions were calculated with the GROMACS g_velacc tool. Diffusion constants are for the passive permeation of the phenylalanine dipeptide through DOPC 50 (blue), POPC 50 (red), and POPC 40 (green) lipid bilayer systems. Error bars represent a 95% confidence interval around the mean value from 10 separate trajectories of 20 ps.

Finally, *$D_{xy}(z)$* diffusion constants were calculated with Equation 4.9 with data from our custom build of CHARMM v. 38 that constrains the z position of the permeant. These results are portrayed in Figure 4.10, as shown below. The difference in terms of the quality of the data is like night and day, as compared to the free GROMACS results depicted in Figure 4.7. Diffusion constants start around 20×10^{-10} m^2s^{-1} before gradually decreasing to around 10×10^{-10}

m^2s^{-1} in the interfacial region before gradually increasing to around 17×10^{-10} m^2s^{-1} at the center of the bilayer. Unlike the previous applications of this method, these diffusion constants are well behaved and possess reasonable uncertainty ranges. The primary difference between this data and that in Figure 4.7 is that these simulations were conducted with a *z*-constrained CHARMM simulation. This suggests that the free movement of the permeant in the *z*-direction may be causing many of the issues witnessed with the previous data.

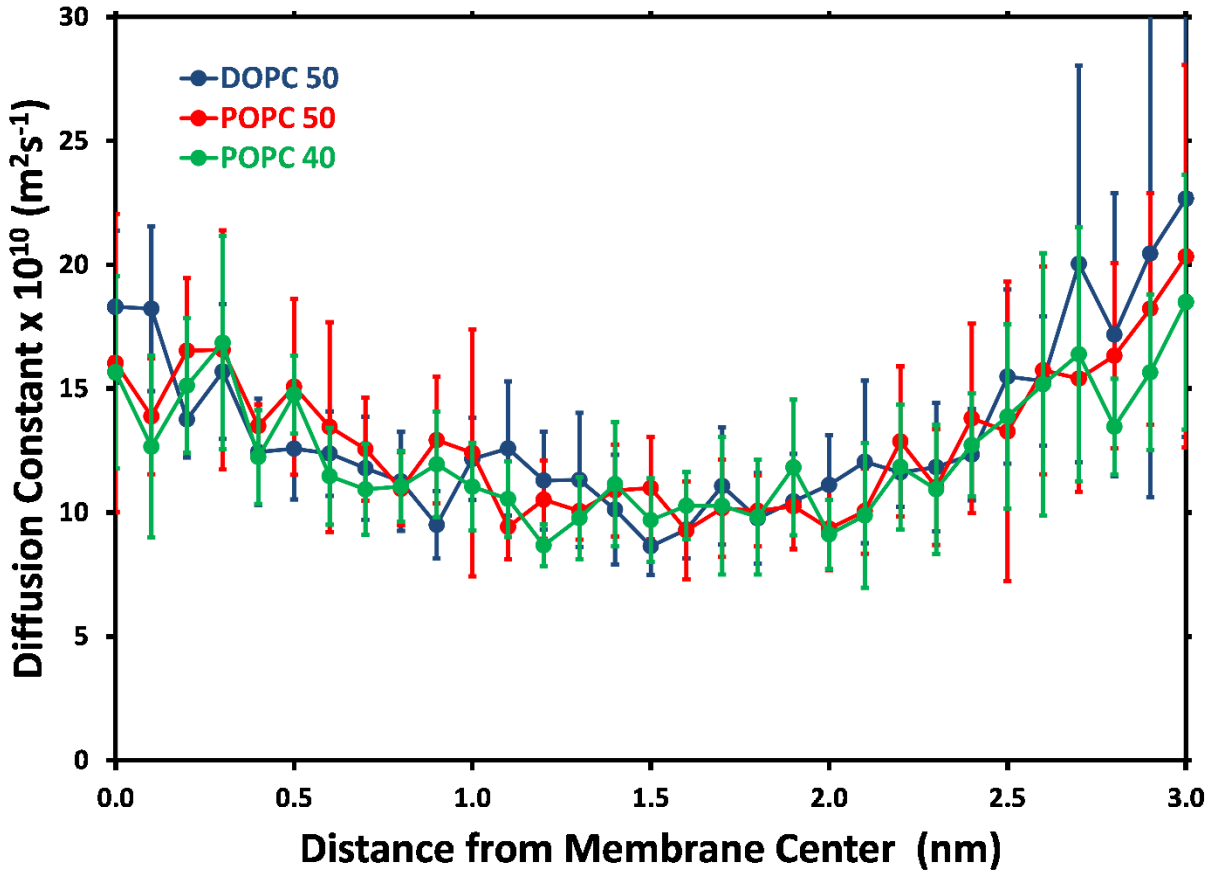


Figure 4.10. $D_z(z)$ calculated from a 2D Green-Kubo expression using velocity data from *z*-constrained CHARMM simulations. Diffusion constants are for the passive permeation of the phenylalanine dipeptide through DOPC 50 (blue), POPC 50 (red), and POPC 40 (green) lipid bilayer systems. Error bars represent a 95% confidence interval around the mean value from 10 separate trajectories of 20 ps.

Prior to these calculations, a simulation length of 20 ps was chosen due to how far the phenylalanine dipeptide traveled in the *z*-direction. Several free simulations were run at *z* = 0, 0.5, 1.0, 1.5, and 2.0 nm, and the mean displacement of the phenylalanine dipeptide was tracked as a function of time. To reduce the overlap of *z*-position data between successive runs, a time of 20 ps was selected because the phenylalanine dipeptide rarely traveled, on average, beyond 0.1 nm during this time frame. Bassolino-Klimas *et. al.* conducted a more rigorous investigation into this drift and found that permeant distances in the *z*-direction did not exceed 0.16 nm over the course of an entire 1 ns unconstrained simulation run.[65] We assumed that this position drift would not have a substantial effect on diffusion constant data; however, it may explain the differences between Figure 4.7 and Figure 4.10. Alternatively, drift in the *z*-direction may reduce lateral diffusion through the coupling of transverse and lateral motions, such as the molecule attempting a 'corkscrew' motion only to be blocked by a transverse (*z*) movement which causes an anti-correlation within the lateral velocity data.

4.4.3. – Einstein Relations.

Einstein relations are the most popular method in passive permeation studies for determining diffusion constants. They are typically used, however, only in the center of the bilayer, at the interface, or in the aqueous region, likely due to the drift of the permeant in free molecular dynamics simulations, as discussed in more detail at the end of section 4.4.2. One of the greatest difficulties is determining the region of best fit because mean square displacement (MSD) plots contain a substantial amount of information concerning the motion of the permeant. In Figure 4.11, shown below, the MSDs for 10 trajectories at *z* = 0 for POPC 50 are

plotted. From 0 to 0.25 ps, the ballistic regime can be seen. Until 1 ps, a roughly linear relationship between the MSD and time develops before one of two things happen: 1.) a somewhat linear relationship exists until around 5 ps, as seen with the purple and blue trajectories, or 2.) the MSD begins to oscillate wildly with respect to time, as seen by the red trajectories. In general, a diffusive regime exists for a small period of time before sub-diffusion or super-diffusion occurs indicating the complicated motion of the permeant within the *z*-direction. Obviously, this complicates the calculation of diffusion constants. We chose to conduct our linear fit between 1 and 5 ps where most trajectories possess a somewhat linear relationship. However, as seen in Figure 4.12, some linear fits were better than others. Linear fits were also conducted between 0.25 and 0.5 ps and between 0.5 and 1.0 ps but were no more reliable than the 1 to 5 ps range.

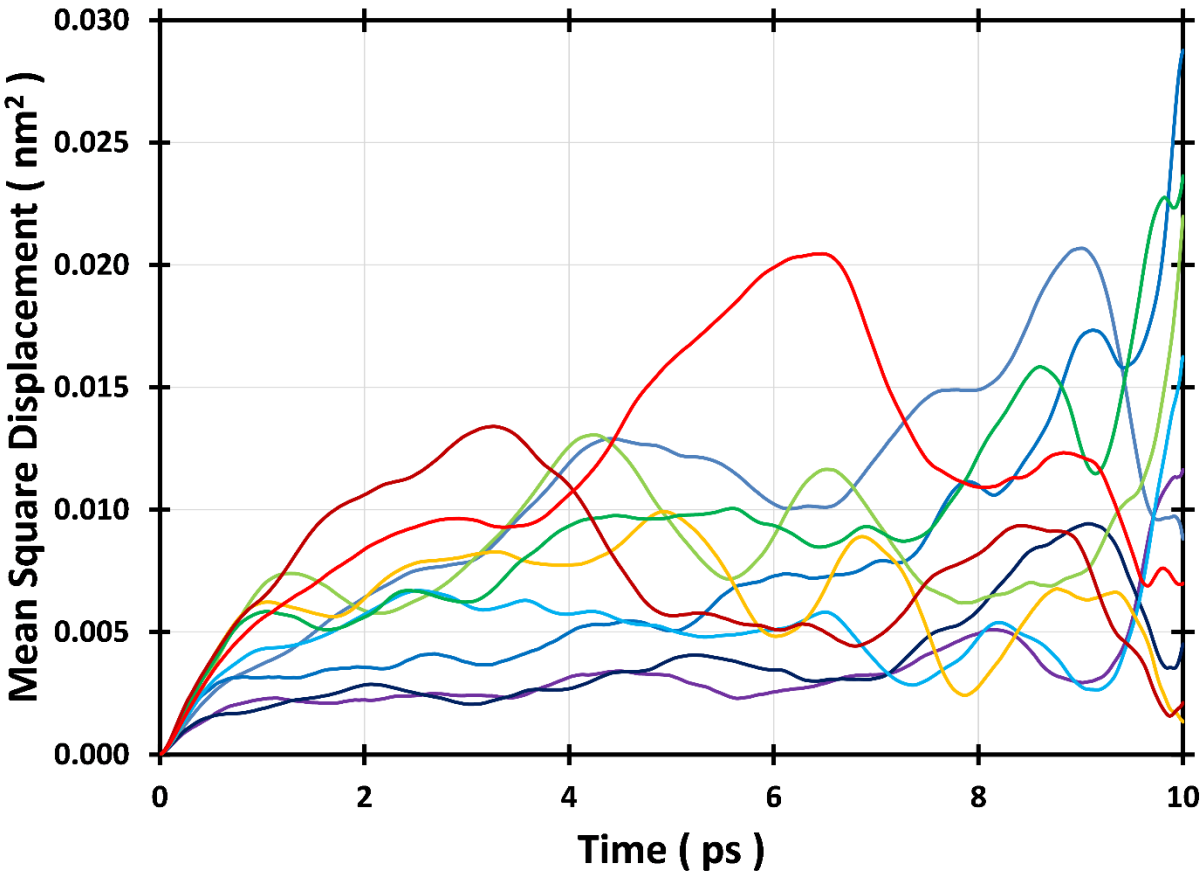


Figure 4.11. The mean square displacement for the phenylalanine dipeptide is depicted from a free GROMACS simulation at z = 0 for the POPC 50 system. Data has been averaged over a 20 ps trajectory at different time origins for a single dipeptide.

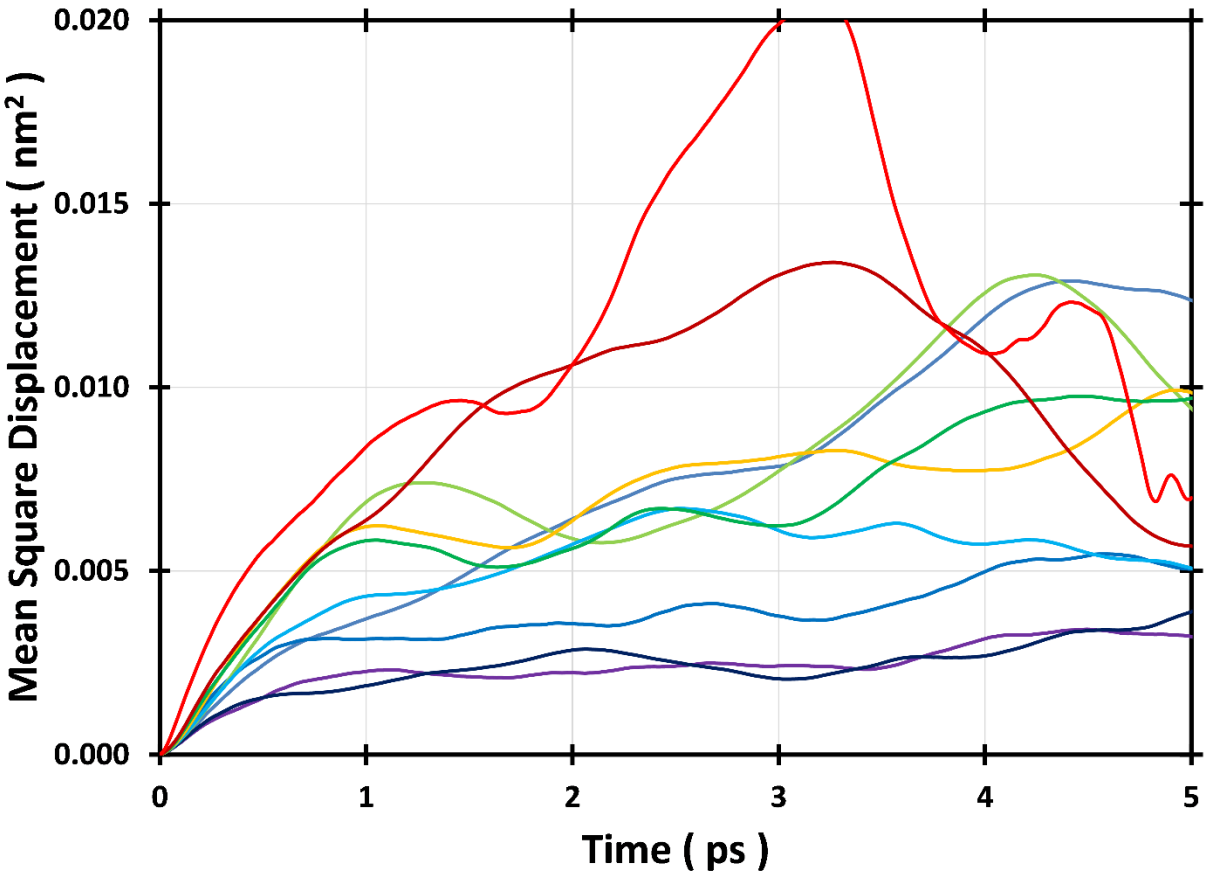


Figure 4.12. The mean square displacement for the phenylalanine dipeptide is depicted for a shorter time period from a free GROMACS simulation at z = 0 for the POPC 50 system. This is the same data as shown in Figure 4.11.

Diffusion constants using Equation 4.3 are plotted in Figure 4.13, as seen below. Similar trends are seen in this figure as were seen in Figure 4.4 which depicted the same $D_z(z)$ as calculated from a Green-Kubo expression. Specifically, diffusion constants start out at 15×10^{-10} m^2s^{-1}, a little higher than in Figure 4.4, before decreasing to around 1 to 5 × 10^{-10} m^2s^{-1} as the permeant approaches the middle of the lipid bilayer. Thankfully, the variability in the data is slightly less than with Green-Kubo, likely due to the averaging effects of the linear fitting process. However, occasionally, negative diffusion constants appear due to negative slopes

within the fitted region, as seen from Figures 4.9 and 4.10. These negative slopes are likely due to the same backscattering effect that was witnessed within the previous section: the permeant moves forward, collides with another molecule such as a lipid, and proceeds in the opposite direction.

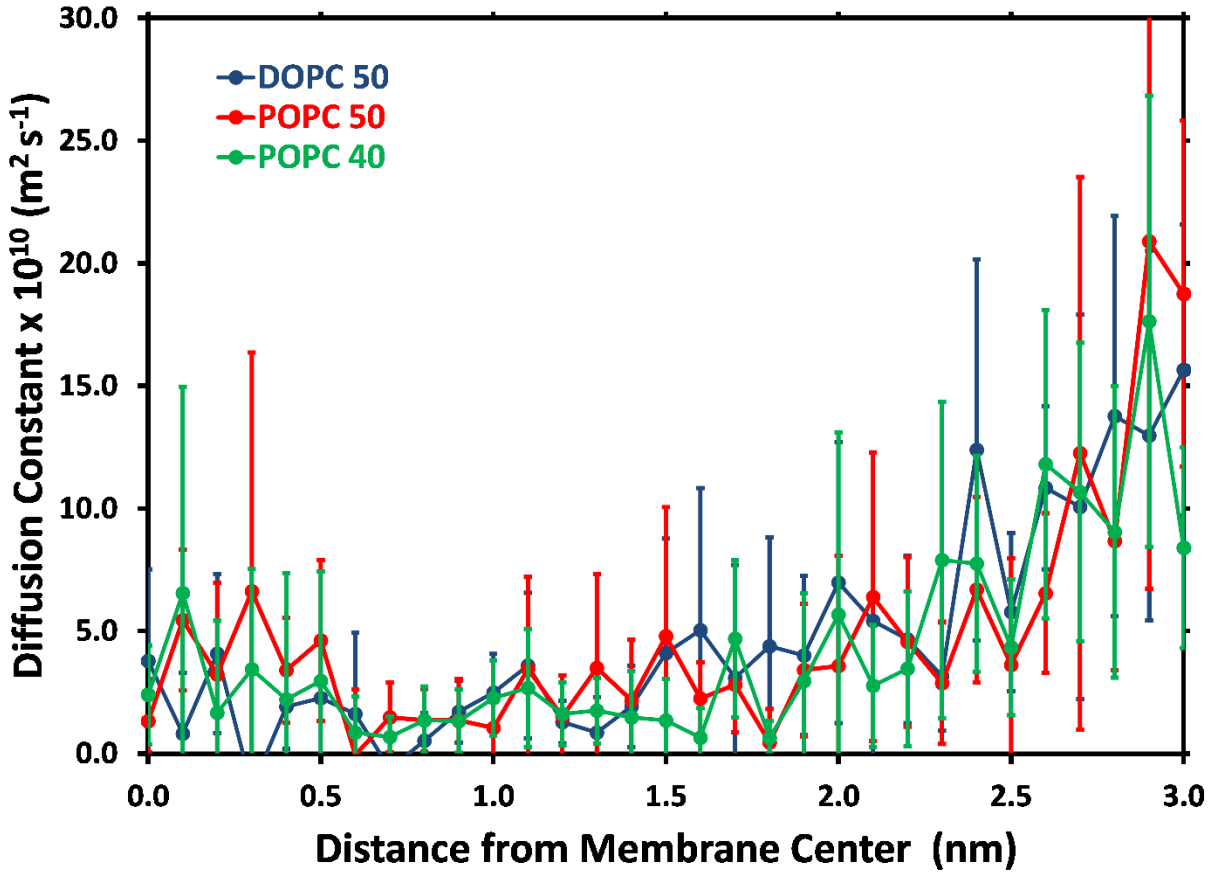


Figure 4.13. *$D_z(z)$* calculated from an Einstein relation using position data from an unconstrained, free GROMACS simulation. Diffusion constants are for the passive permeation of the phenylalanine dipeptide through DOPC 50 (blue), POPC 50 (red), and POPC 40 (green) lipid bilayer systems. Error bars represent a 95% confidence interval around the mean value from 10 separate trajectories.

4.4.4. – The Hummer Displacement Method.

The Hummer displacement method is an intriguing approach that utilized position autocorrelation data from constrained umbrella sampling simulations. In recent years, it has gained popularity as an alternative to the Fluctuation-Dissipation theorem. Results utilizing Equation 4.6 are plotted below, on a logarithmic scale, in Figure 4.14. Diffusion constants from this method spanned approximately five orders of magnitude and are clearly nonsensical. In order to display them, a logarithmic scale had to be used; and, as a result, error bars that extended below zero cannot be displayed. Perhaps, this is the unusual oscillatory behavior that Vorobyov *et. al.* observed in their application of the method. Due to the numerous and complex assumptions on which this method is based, it is doubtful that the approach is applicable to membrane permeation simulations. Therefore, it is difficult to either troubleshoot or explain the following results.

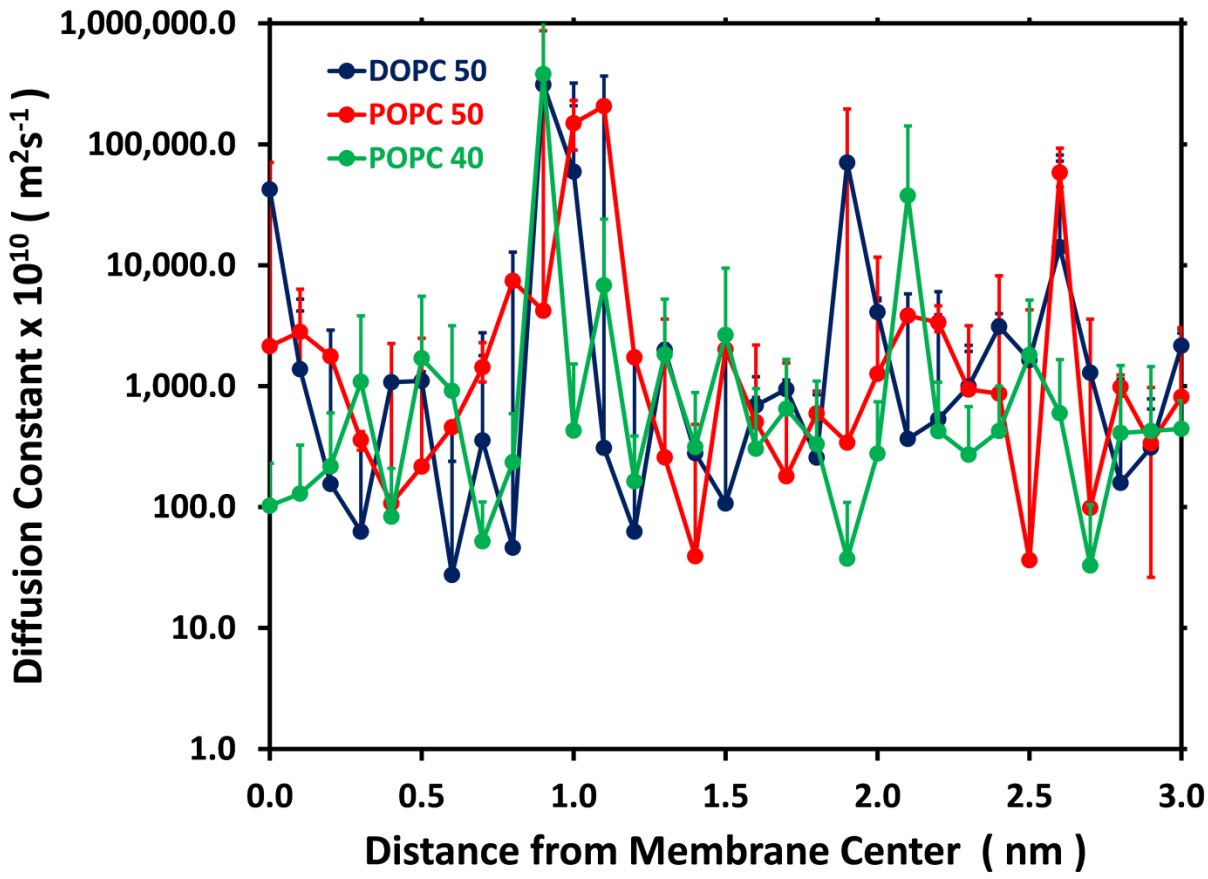


Figure 4.14. $D_z(z)$ calculated from the Hummer Displacement method using position data from umbrella sampling GROMACS simulations. Diffusion constants are for the passive permeation of the phenylalanine dipeptide through DOPC 50 (blue), POPC 50 (red), and POPC 40 (green) lipid bilayer systems. Error bars represent a 95% confidence interval around the mean value from 5 contiguous trajectories of 5 ns. Error bars resulting in negative diffusion constants are not shown on this logarithmic plot.

4.4.5. – Numerical Approximation to the Smoluchowski Equation.

Diffusion constants using Equation 4.12 are plotted in Figure 4.15, as seen below. Diffusion constants start around 5.4×10^{-10} m^2s^{-1}, decrease slightly to around 4.8×10^{-10} m^2s^{-1}, before then increasing slightly to around 5.5×10^{-10} m^2s^{-1}; however, only a few of these changes are statistically significant. With the exceptions of a large uncertainty for POPC 40 at z = 1.6 nm

and a slight decrease in DOPC 50 data at z = 1.7 nm, this method is remarkably stable, as compared to the autocorrelation and MSD methods. This is likely due to the large timescale of 12.5 ns per data analysis and the averaging effect of the binning process. As *z(t)* is binned into a relatively simple histogram of only four bins, all of the intricate permeant motions and interactions, as discussed throughout the rest of this Chapter, are averaged into a single, bulk property: does the permeant cross over into another position bin or not? If so, how long does it reside in a single bin? This loss of molecular granularity likely increases the precision of the data and provides a more accurate, long-term estimation of the diffusion constant. For these reasons, we prefer this method.

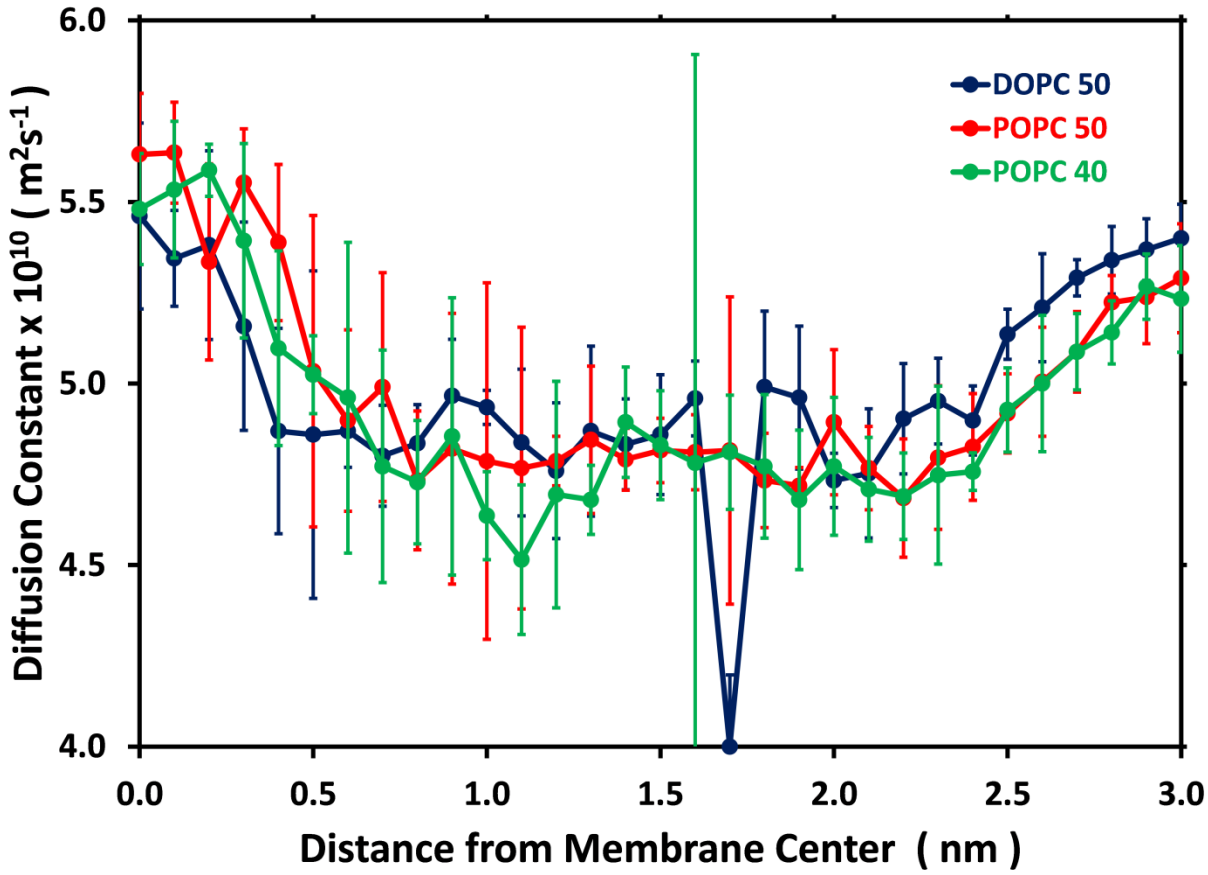


Figure 4.15. $D_z(z)$ calculated from a numerical solution to the Smoluchowski equation from position data from umbrella sampling GROMACS simulations. Diffusion constants are for the passive permeation of the phenylalanine dipeptide through DOPC 50 (blue), POPC 50 (red), and POPC 40 (green) lipid bilayer systems. Error bars represent a 95% confidence interval around the mean value from 4 contiguous trajectories of 12.5 ns each.

4.5 Conclusions.

Diffusion constants were calculated from a variety of different methods for many different cases. Specifically, five different methods were used: 1.) the Fluctuation Dissipation theorem, 2.) Green-Kubo expressions, 3.) Einstein relations, 4.) the Hummer Displacement method, and 5.) a numerical approximation of the Smoluchowski equation.

For Method 1, $D_z(z)$ was calculated with: umbrella sampling pulling forces from GROMACS; *z*-constrained forces with CHARMM; and unconstrained forces from CHARMM. The umbrella sampling pulling forces, although commonly used, presented unusual results that have not been shown in other studies. Although the pulling forces are related to the displacement of the permeant from the center of the umbrella window, which are, in turn, related in some manner to the actual forces experienced by the phenylalanine dipeptide, it is unclear as to what effect they may have on the force fluctuation autocorrelation functions. CHARMM forces from the *z*-constrained simulations, however, did not present such an artifact and produced results similar to previous studies. However, the presence of a constraint likely increases the correlation of data and lowers calculated diffusion constants, as seen in similar calculations from unconstrained CHARMM simulations.

For Method 2: $D_z(z)$ was calculated with free GROMACS velocities; $D_{xy}(z)$ was calculated with free GROMACS velocities and *z*-constrained CHARMM velocities; $D_{3d}(z)$ was calculated with free GROMACS velocities that were entered into either our autocorrelation program or the GROMACS tool g_velacc. Although the calculation of data was relatively easy from the rapid convergence and decay of the associated autocorrelation functions, this method was extremely sensitive to the motions of the permeant and often resulted in wildly fluctuating, and, at times, negative, diffusion constants. However, once the permeant was constrained in terms of the *z*-coordinate, stable and reasonable results for $D_{xy}(z)$ were obtained, implying that the free diffusion of the permeant along the *z*-axis is likely an issue in all of these calculations.

For Method 3: $D_z(z)$ was calculated from free GROMACS MSD data. Unfortunately, the use of this method was problematic due to the non-diffusive nature of the MSD versus time data. Plots of the mean square displacement clearly indicated a wide variety of unusual molecular motions, such as oscillatory behavior and negative slopes in addition to sub-diffusion and super-diffusion. In general, this method was a little more reliable than the Green-Kubo data, but still exhibited many of the same problems.

Methods 4 and 5 were only used to calculate $D_z(z)$ values. The Hummer Displacement method presents an intriguing theoretical relationship between diffusion constants and position autocorrelation functions. However, in practice, it simply does not work. Conversely, Method 5, a numerical approximation to the Smolochowski equation presented the most reliable method for calculating diffusion constants, in large part due to its ability to average out the complicated permeant motions that present difficultly for the autocorrelation based methods.

Unfortunately, however, a great many questions still remain. For example, how do other researchers use these methods to arrive at reasonable results? They likely use very short time scales to avoid many of the issues that we witnessed and only pick trajectories that are well behaved. Conversely, some articles use timescales that are extremely long (> 100 ns). Over which, many of the transient molecular motions witnessed earlier are averaged out and do not impact final results. This then begs the question: what is the optimal time scale for each of these methods? Ultimately, the above may simply be a result of poor statistical sampling because we are only averaging over a single permeant molecule. Many articles overcome this

by replicating multiple permeants within a single simulation cell; but, this has been shown to change the average result.[7] Finally, we think it is important to consider if diffusive motion is simply not relevant. After all, which parts of a passive membrane permeation process are best described by random fluctuations and simple walks in Cartesian dimensions?[66-69]

However, this was not a comprehensive analysis of all of these methods. In general, we believe that most journal articles are able to make many of the more problematic methods work by either increasing statistical sampling or by adjusting the time scale of their simulations. For example, all of the analyses in this work can be repeated with hundreds of separate trajectories over smaller time scales. These trajectories can then be averaged either at the end, as was done in this work, or at the beginning of data analysis, as is often done in the literature. Additionally, a variety of different time scales can be explored, too. For example, Einstein relations appear to work best when the MSD is fit on a nanosecond time scale; whereas, Green-Kubo relations are best used over very short, sub-picosecond time scales. A thorough parametric analysis of the effect of the simulation time scale may provide more optimal parameters for the data analysis. Many of the other methods, such as the application of the Fluctuation-Dissipation theorem to pulling forces and the Hummer-Displacement method, are hindered by the core assumptions behind their theoretical derivations. These assumptions should be tested and verified in order to adjust parameters used for data analysis accordingly, such as the length of the simulation. Finally, an exploration of many of the new methods mentioned in the previous paragraph, such as rotational coupling and modeling diffusion constants in terms of rate expressions and memory functions, provides many exciting avenues for re-examining, if not re-defining, this complicated subject.

4.6 References.

1. Marrink, S.; Berendsen, H., Simulation of Water Transport Through a Lipid-Membrane. *Journal of Physical Chemistry* **1994,** *98* (15), 4155-4168.

2. Marrink, S. J.; Berendsen, H. J. C., Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations. *The Journal of Physical Chemistry* **1996,** *100* (41), 16729-16738.

3. Cardenas, A. E.; Jas, G. S.; DeLeon, K. Y.; Hegefeld, W. A.; Kuczera, K.; Elber, R., Unassisted Transport of N-Acetyl-L-tryptophanamide through Membrane: Experiment and Simulation of Kinetics. *The Journal of Physical Chemistry B* **2012,** *116* (9), 2739-2750.

4. Diamond, J. M.; Katz, Y., Interpretation of Nonelectrolyte Partition-Coefficients between Dimyristoyl Lecithin and Water. *Journal of Membrane Biology* **1974,** *17* (2), 121-154.

5. Schulten, K., Dynamics of reactions involving diffusive barrier crossing. *The Journal of Chemical Physics* **1981,** *74* (8), 4426.

6. Ulander, J.; Haymet, A. D. J., Permeation across hydrated DPPC lipid bilayers: Simulation of the titrable amphiphilic drug valproic acid. *Biophys. J.* **2003,** *85* (6), 3475-3484.

7. Yousefpour, A.; Amjad Iranagh, S.; Nademi, Y.; Modarress, H., Molecular dynamics simulation of nonsteroidal antiinflammatory drugs, naproxen and relafen, in a lipid bilayer membrane. *International Journal of Quantum Chemistry* **2013,** *113* (15), 1919-1930.

8. Hansen, F. Y.; Peters, G. H.; Taub, H.; Miskowiec, A., Diffusion of water and selected atoms in DMPC lipid bilayer membranes. *The Journal of Chemical Physics* **2012,** *137* (20), 204910.

9. Nademi, Y.; Iranagh, S. A.; Yousefpour, A.; Mousavi, S. Z.; Modarress, H., Molecular dynamics simulations and free energy profile of Paracetamol in DPPC and DMPC lipid bilayers. *Journal of Chemical Sciences* **2014,** *126* (3), 637-647.

10. Cascales, J. J. L.; Huertas, M. L.; de la Torre, J. G., Molecular dynamics simulation of a dye molecule in the interior of a bilayer: 1,6-diphenyl-1,3,5-hexatriene in dipalmitoylphosphatidylcholine. *Biophysical Chemistry* **1997,** *69* (1), 1-8.

11. Hoopes, M. I.; Noro, M. G.; Longo, M. L.; Faller, R., Bilayer Structure and Lipid Dynamics in a Model Stratum Corneum with Oleic Acid. *The Journal of Physical Chemistry B* **2011,** *115* (12), 3164-3171.

12. Porasso, R. D.; Drew Bennett, W. F.; Oliveira-Costa, S. D.; López Cascales, J. J., Study of the Benzocaine Transfer from Aqueous Solution to the Interior of a Biological Membrane. *The Journal of Physical Chemistry B* **2009,** *113* (29), 9988-9994.

13. Bassolino-Klimas, D.; Alper, H. E.; Stouch, T. R., Solute diffusion in lipid bilayer membranes: an atomic level study by molecular dynamics simulation. *Biochemistry* **1993,** *32* (47), 12624-12637.

14. Söderhäll, J. A.; Laaksonen, A., Molecular Dynamics Simulations of Ubiquinone inside a Lipid Bilayer. *The Journal of Physical Chemistry B* **2001,** *105* (38), 9308-9315.

15. Alper, H. E.; Stouch, T. R., Orientation and Diffusion of a Drug Analog in Biomembranes: Molecular Dynamics Simulations. *The Journal of Physical Chemistry* **1995,** *99* (15), 5724-5731.

16. dos Santos, D. J. V. A.; Eriksson, L. A., Permeability of psoralen derivatives in lipid membranes. *Biophys. J.* **2006,** *91* (7), 2464-2474.

17. Khuntawee, W.; Wolschann, P.; Rungrotmongkol, T.; Wong-ekkabut, J.; Hannongbua, S., Molecular Dynamics Simulations of the Interaction of Beta Cyclodextrin with a Lipid Bilayer. *Journal of Chemical Information and Modeling* **2015,** *55* (9), 1894-1902.

18. Khajeh, A.; Modarress, H., Effect of cholesterol on behavior of 5-fluorouracil (5-FU) in a DMPC lipid bilayer, a molecular dynamics study. *Biophysical Chemistry* **2014,** *187-188*, 43-50.

19. Böckmann, R. A.; Hac, A.; Heimburg, T.; Grubmüller, H., Effect of sodium chloride on a lipid bilayer. *Biophys. J.* **2003,** *85* (3), 1647-1655.

20. Babakhani, A.; Gorfe, A. A.; Gullingsrud, J.; Kim, J. E.; Andrew McCammon, J., Peptide insertion, positioning, and stabilization in a membrane: Insight from an all-atom molecular dynamics simulation. *Biopolymers* **2007,** *85* (5-6), 490-497.

21. Wei, T.; Huang, T.; Qiao, B.; Zhang, M.; Ma, H.; Zhang, L., Structures, Dynamics, and Water Permeation Free Energy across Bilayers of Lipid A and Its Analog Studied with Molecular Dynamics Simulation. *The Journal of Physical Chemistry B* **2014,** *118* (46), 13202-13209.

22. Weiss, M.; Elsner, M.; Kartberg, F.; Nilsson, T., Anomalous Subdiffusion Is a Measure for Cytoplasmic Crowding in Living Cells. *Biophys. J.* **2004,** *87* (5), 3518-3524.

23. Trovato, F.; Tozzini, V., Diffusion within the Cytoplasm: A Mesoscale Model of Interacting Macromolecules. *Biophys. J.* **2014,** *107* (11), 2579-2591.

24. Metzler, R.; Jeon, J. H.; Cherstvy, A. G., Non-Brownian diffusion in lipid membranes: Experiments and simulations. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2016,** *1858* (10), 2451-2467.

25. Balbo, J.; Mereghetti, P.; Herten, D.-P.; Wade, Rebecca C., The Shape of Protein Crowders is a Major Determinant of Protein Diffusion. *Biophys. J.* **2013,** *104* (7), 1576-1584.

26. Shinoda, W.; Okazaki, S., A Voronoi analysis of lipid area fluctuation in a bilayer. *The Journal of Chemical Physics* **1998,** *109* (4), 1517.

27. Kubo, R., Fluctuation-Dissipation Theorem. *Reports on Progress in Physics* **1966,** *29*, 255-&.

28. Cardenas, A. E.; Shrestha, R.; Webb, L. J.; Elber, R., Membrane Permeation of a Peptide: It Is Better to be Positive. *The Journal of Physical Chemistry B* **2015,** *119* (21), 6412-6420.

29. Wei, C.; Pohorille, A., Flip-Flop of Oleic Acid in a Phospholipid Membrane: Rate and Mechanism. *The Journal of Physical Chemistry B* **2014,** *118* (45), 12919-12926.

30. Shinoda, W.; Mikami, M.; Baba, T.; Hato, M., Molecular dynamics study on the effects of chain branching on the physical properties of lipid bilayers: 2. Permeability. *The Journal of Physical Chemistry B* **2004,** *108* (26), 9346-9356.

31. Sugii, T.; Takagi, S.; Matsumoto, Y., A molecular-dynamics study of lipid bilayers: Effects of the hydrocarbon chain length on permeability. *The Journal of Chemical Physics* **2005,** *123* (18), 184714.

32. Orsi, M.; Sanderson, W. E.; Essex, J. W., Permeability of Small Molecules through a Lipid Bilayer: A Multiscale Simulation Study. *The Journal of Physical Chemistry B* **2009,** *113* (35), 12019-12029.

33. Orsi, M.; Essex, J. W., Permeability of drugs and hormones through a lipid bilayer: insights from dual-resolution molecular dynamics. *Soft Matter* **2010,** *6* (16), 3797.

34. Fiedler, S. L.; Violi, A., Simulation of Nanoparticle Permeation through a Lipid Membrane. *Biophys. J.* **2010,** *99* (1), 144-152.

35. Woolf, T. B.; Roux, B., Conformational Flexibility of o-Phosphorylcholine and o-Phosphorylethanolamine: A Molecular Dynamics Study of Solvation Effects. *Journal of the American Chemical Society* **1994,** *116* (13), 5916-5926.

36. Hummer, G., Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New Journal of Physics* **2005,** *7*, 34-34.

37. Riahi, S.; Rowley, C. N., Why Can Hydrogen Sulfide Permeate Cell Membranes? *Journal of the American Chemical Society* **2014,** *136* (43), 15111-15113.

38. Vorobyov, I.; Olson, Timothy E.; Kim, Jung H.; Koeppe, Roger E.; Andersen, Olaf S.; Allen, Toby W., Ion-Induced Defect Permeation of Lipid Membranes. *Biophys. J.* **2014,** *106* (3), 586-597.

39. Issack, B. B.; Peslherbe, G. H., Effects of Cholesterol on the Thermodynamics and Kinetics of Passive Transport of Water through Lipid Membranes. *The Journal of Physical Chemistry B* **2015,** *119* (29), 9391-9400.

40. Carpenter, Timothy S.; Kirshner, Daniel A.; Lau, Edmond Y.; Wong, Sergio E.; Nilmeier, Jerome P.; Lightstone, Felice C., A Method to Predict Blood-Brain Barrier Permeability of Drug-Like Compounds Using Molecular Dynamics Simulations. *Biophys. J.* **2014,** *107* (3), 630-641.

41. Comer, J.; Schulten, K.; Chipot, C., Diffusive Models of Membrane Permeation with Explicit Orientational Freedom. *Journal of Chemical Theory and Computation* **2014,** *10* (7), 2710-2718.

42. Holland, B. W.; Gray, C. G.; Tomberli, B., Calculating diffusion and permeability coefficients with the oscillating forward-reverse method. *Physical Review E* **2012,** *86* (3).

43. Zusman, L. D., Outer-Sphere Eelectron-Transfer in Polar-Solvents. *Chemical Physics* **1980,** *49* (2), 295-304.

44. Bicout, D. J.; Szabo, A., Electron transfer reaction dynamics in non-Debye solvents. *J. Chem. Phys.* **1998,** *109* (6), 2325-2338.

45. Petrache, H. I.; Tristram-Nagle, S.; Gawrisch, K.; Harries, D.; Parsegian, V. A.; Nagle, J. F., Structure and fluctuations of charged phosphatidylserine bilayers in the absence of salt. *Biophys. J.* **2004,** *86* (3), 1574-1586.

46. Liu, Y.; Nagle, J. F., Diffuse scattering provides material parameters and electron density profiles of biomembranes. *Physical Review E* **2004,** *69* (4).

47. Lee, B. L.; Kuczera, K.; Middaugh, C. R.; Jas, G. S., Permeation of the three aromatic dipeptides through lipid bilayers: Experimental and computational study. *J. Chem. Phys.* **2016,** *144* (24), 14.

48. Leftin, A.; Molugu, T. R.; Job, C.; Beyer, K.; Brown, M. F., Area per lipid and cholesterol interactions in membranes from separated local-field (13)C NMR spectroscopy. *Biophys. J.* **2014,** *107* (10), 2274-2286.

49. Jo, S.; Kim, T.; Im, W., Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS ONE* **2007,** *2* (9), e880.

50. Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W., CHARMM-GUI *Membrane Builder* toward realistic biological membrane simulations. *Journal of Computational Chemistry* **2014,** *35* (27), 1997-2004.

51. Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W., CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **2009,** *97* (1), 50-58.

52. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013,** *29* (7), 845-854.

53. Pastor, R. W.; MacKerell, A. D., Development of the CHARMM Force Field for Lipids. *The Journal of Physical Chemistry Letters* **2011,** *2* (13), 1526-1532.

54. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **2009**, NA-NA.

55. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926.

56. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995,** *103* (19), 8577.

57. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007,** *126* (1), 014101.

58. Allen, M. P.; Tildesley, D. J., *Computer simulation of liquids*. Clarendon Press: 1989; p 385.

59. Torrie, G.; Valleau, J., Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *Journal of Computational Physics* **1977,** *23* (2), 187-199.

60. Kästner, J., Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011,** *1* (6), 932-942.

61. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **2009,** *30* (10), 1545-1614.

62. Nosé, S., A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **1984,** *81* (1), 511.

63. Hoover, W., Canonical Dynamics - Equilibrium Phase-Space Distributions. *Physical Review A* **1985,** *31* (3), 1695-1697.

64. Stevens, M. J., Coarse-grained simulations of lipid bilayers. *The Journal of Chemical Physics* **2004,** *121* (23), 11942.

65. Bassolino-Klimas, D.; Alper, H. E.; Stouch, T. R., Mechanism of solute diffusion through lipid bilayer membranes by molecular dynamics simulation. *Journal of the American Chemical Society* **1995,** *117* (14), 4118-4129.

66. Cardenas, A. E.; Elber, R., Modeling kinetics and equilibrium of membranes with fields: Milestoning analysis and implication to permeation. *The Journal of Chemical Physics* **2014,** *141* (5), 054101.

67. Parisio, G.; Stocchero, M.; Ferrarini, A., Passive Membrane Permeability: Beyond the Standard Solubility-Diffusion Model. *Journal of Chemical Theory and Computation* **2013,** *9* (12), 5236-5246.

68. Berezhkovskii, A. M.; Makarov, D. E., Communication: Coordinate-dependent diffusivity from single molecule trajectories. *J. Chem. Phys.* **2017,** *147* (20), 3.

69. Roos, M.; Ott, M.; Hofmann, M.; Link, S.; Rössler, E.; Balbach, J.; Krushelnitsky, A.; Saalwächter, K., Coupling and Decoupling of Rotational and Translational Diffusion of Proteins under Crowding Conditions. *Journal of the American Chemical Society* **2016,** *138* (32), 10365-10372.

Chapter 5

Permeation of a Small, Helical Peptide: *wh5*

5.1. Overview:

The passive permeation of *wh5* through a lipid bilayer is explored in this chapter by conducting a variety of different analyses, whose results are cataloged below. The potential of mean force indicates large membrane deformations as the peptide permeates the membrane; however, it still exhibits the same pattern of an interfacial free energy minimum and a maximum at the center of the lipid bilayer. The extent of the membrane deformation is also indicated by the hydrogen bonding pairs within 0.35 nm analysis, where a large number of hydrogen bonding pairs are present around *wh5* even at the center of the bilayer, where few would be expected due to the presence of the acyl lipid tails. This indicates that a large solvation shell is being carried along by *wh5*. All of the remaining structural analyses – alpha carbon root mean square distance, end-to-end distance, tryptophan and histidine sidechain separation distance, the radius of gyration, and the O to N distance in hydrogen bonds 1, 2, and 3 – all indicate that the helix is unravelling around z = 2.5 nm with some fluctuations in the center of the membrane and farther out in the aqueous region. However, this is not seen at z = 3.5 nm, where experimental results predict that the wh5 helix forms every 2 – 7 ns. The radius of gyration and sidechain separation distance analyses also indicate an interesting pattern from z = 1.0 nm to z = 0.5 nm that warrants further investigation.

5.2 Introduction:

wh5 is a small peptide comprised of 5 amino acids: tryptophan, whose single letter abbreviation is a '*w*'; three alanine residues; and histidine, whose single letter abbreviation is an '*h*'. Figure 5.1, shown at the end of this section, contains a drawing of *wh5*. *wh5* is one of

the smallest peptides that is capable of forming an alpha helix within itself.[1-2] This interesting property is due to the formation of three intramolecular hydrogen bonds, as shown in Figure 5.2. The first hydrogen bond is formed between the oxygen atom on the acetylated N-terminus and the N-terminus of the second alanine residue. The second hydrogen bond is formed between the oxygen atom on the C-terminus of the tryptophan residue and the N-terminus of the histidine residue. The third hydrogen bond is formed between the oxygen atom on the C-terminus of the first alanine residue and the N-terminus of the amidated histidine residue. Given these important structural features, we plan on examining the distances between these atoms, along with many of our analyses from Chapters 2 and 3, on the passive permeation of wh5 through a lipid bilayer.

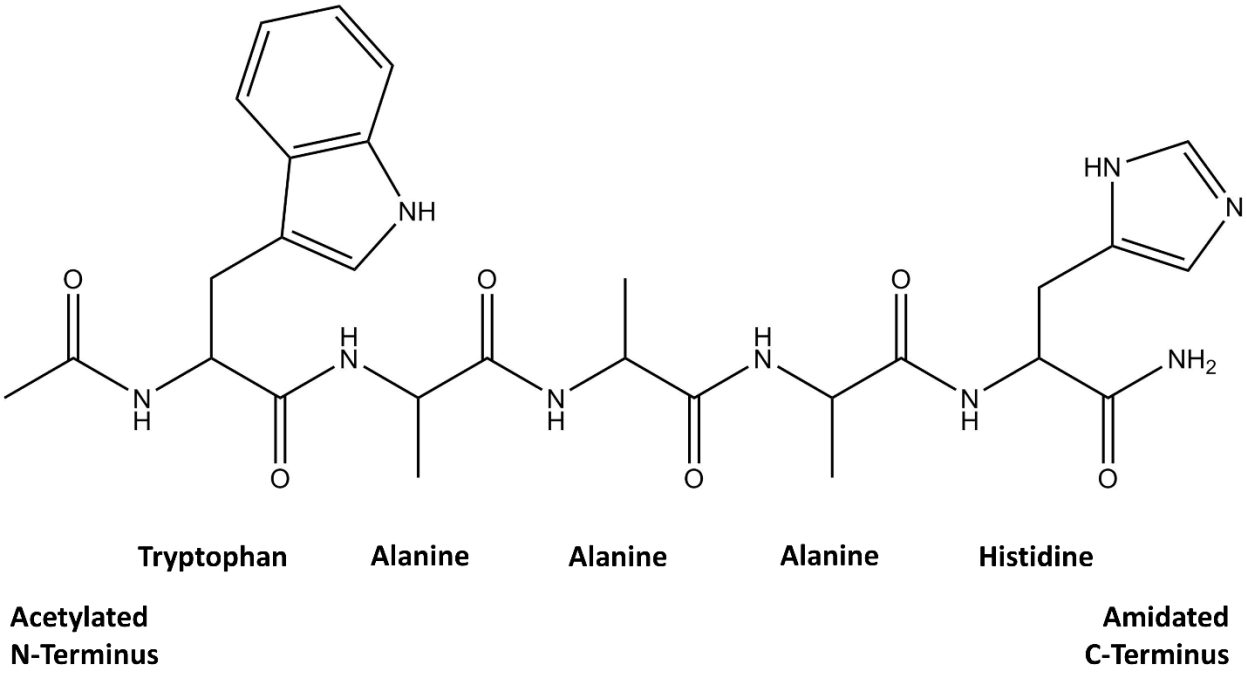


Figure 5.1. A schematic diagram of the peptide *wh5*. *wh5* consists of tryptophan, three alanines, and a histidine. Note that the N-terminus (left) has been acetylated and the C-terminus (right) has been amidated.

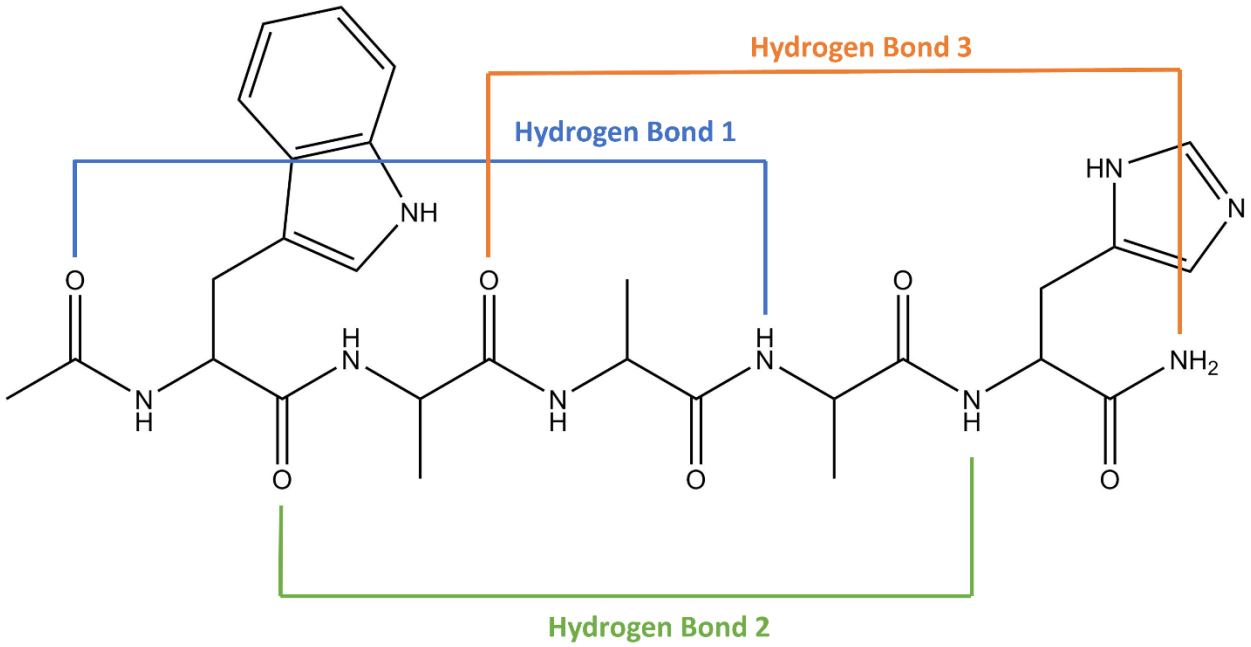


Figure 5.2. The locations of hydrogen bonds 1, 2, and 3 are displayed on *wh5*. These hydrogen bonding sites are responsible for the ability of *wh5* to form an alpha helix within itself.

5.3. Methods:

The physical setup of our simulation window followed a procedure similar to our previous work, as seen in Chapters 2 and 3.[3] Our simulation windows consisted of one *wh5* molecule, also known as Ac-WAAAH-NH2; water; and a phospholipid bilayer consisting 1,3-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC). Specifically, the system contained the following molecules: 1 *wh5*, 50 DOPC lipids, 2991 TIP3P waters, 7 sodium atoms, and 8 chlorine atoms, for a grand total of 15,822 atoms. Sodium and chlorine atoms were added to maintain a physiologically relevant ionic strength. The system was confined to a tetragonal box of dimensions: 4.12 × 4.12 × 9.03 nm along the *x*, *y*, and *z* axes, respectively. The *x* and *y* axes of the simulation box are parallel to the plane of the lipid bilayer, and the *z* axis is defined as being

perpendicular to the plane of the lipid bilayer. The initial coordinates for all systems, including the phospholipid bilayer, were determined by using CHARMM-GUI.[4-6]

Molecular dynamics simulations were carried out with GROMACS 4.5.4. Forces for DOPC and *wh5* were both represented by the version 36 CHARMM force field.[7-8] Water was represented by using the TIP3P model.[9] Direct electrostatic interactions were truncated at 0.13 nm, and a long range-correction using the Particle Mesh Ewald method was used.[10] Van der Waals interactions were similarly truncated at 1.2 nm; however, a force switching function was used to smooth the transition between 1.0 and 1.2 nm. A canonical ensemble was used; and, therefore, the number of system particles, the system volume, and temperature were constant. Temperature was held constant at 300K by using velocity rescaling.[11] Initial velocities were sampled from a Maxwell distribution at 300K. The default leap-frog algorithm was used to integrate the relevant equations of motion with a time step of 2 fs. Periodic boundary conditions in the *x*, *y*, and *z* simulation box directions were used. Constraints on all bonds involving hydrogen atoms were enforced by using the LINCS algorithm.[12] Position, energy, and other simulation data were saved every 500 time steps and once every picosecond.

Simulations began by equilibrating each system, as setup by CHARMM-GUI, over 500 ps intervals with the parameters described in the previous paragraph. Once equilibrated, an unrestrained simulation was run for at least 50 ns, and a system image was obtained at z = 1.6 nm. This image was then used as the starting point for further umbrella sampling windows in both positive and negative z directions. Umbrella sampling windows were run at z = 0.0 nm, 0.07 nm, 0.15 nm, 0.23 nm, 0.31 nm, 0.39 nm, 0.48 nm, 0.56 nm, 0.65 nm, 0.74 nm, 0.83 nm,

0.92 nm, and 1.01 nm. From z = 1.1 nm to z = 3.5 nm, windows were run in 0.1 nm increments. More windows were run near the center of the lipid bilayer in order to increase statistical overlap for the potential of mean force calculations. A harmonic restraining potential of 3000 $kJmol^{-1}nm^{-1}$, located at the center of each umbrella sampling window, was applied to the center of mass of the *wh5*.[13-14] Each window ran for 100 ns. The potential of mean force (PMF) was then calculated by using the weighted histogram analysis method (WHAM).[15-17]

Additional structural information was extracted from each umbrella sampling window by using the appropriate GROMACS tool. Additional data analysis, such as the calculation of averages, statistical uncertainties, and histogram binning was carried out by using manually written FORTRAN programs or shell scripts, as needed.

5.4. Results and Discussion

The potential of mean force represents the energy required to pull the permeant along the z-axis of the simulation box. The PMF for *wh5* is plotted below in Figure 5.3. *wh5* possesses a minimum of -18.9 $kJmol^{-1}$ at z = 1.66 nm and a maximum at z = 0.0 nm of 73.5 $kJmol^{-1}$. Both values are relative to the value of the PMF in the aqueous region which was set to 0. *wh5* also possesses a remarkably broad and flat PMF with a relatively linear increase towards the center of the membrane which is indicative of large membrane deformations.[18]

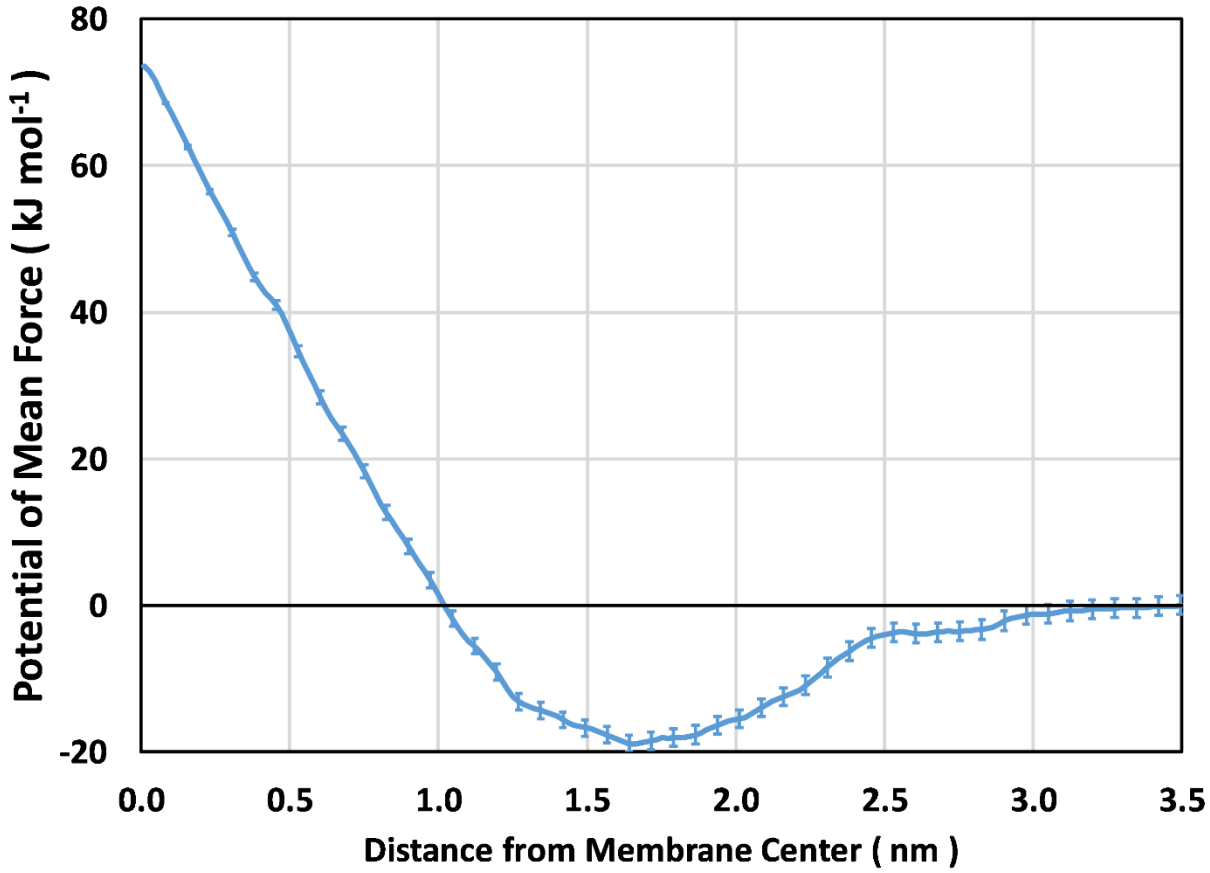


Figure 5.3. The potential of mean force for *wh5*. Error bars represent the standard error at a 95% level of confidence, as calculated by the Bootstrap method.

The root mean square deviation (RMSD) is a measure of how different the desired molecule's structure is over the course of the simulation to a reference structure. Below, the RMSD for all of the alpha carbon atoms in *wh5* is plotted in Figure 5.4. The greater the RMSD, the more the structure of *wh5* varies from its reference structure. In this case, the reference structure is the initial starting structure for each umbrella sampling window. For most of the umbrella sampling windows, the RMSD is relatively small at around 0.05 nm. However, at a few positions near the center of the membrane (z = 0.23 and z = 0.39 nm) and within the aqueous

region (z = 2.4, 2.5, 2.6, and 3.1 nm) the RMSD was several times larger, around 0.10 to 0.23 nm, indicating regions of high backbone mobility.

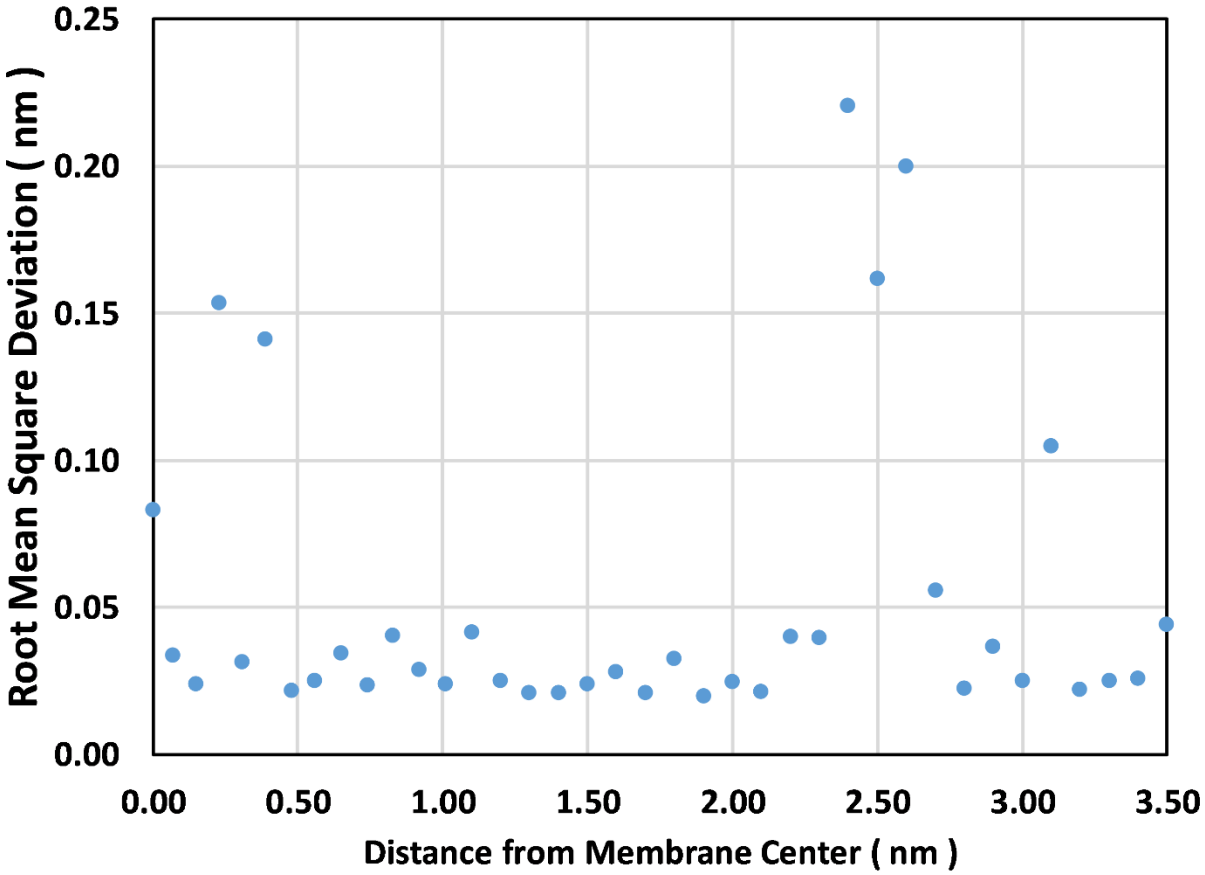


Figure 5.4. The root mean square deviation of the alpha carbons of *wh5*. The greater the value, the more the alpha carbons deviate from their starting structure over the course of an umbrella sampling window. Note: the standard error of the mean value at a 95% confidence level is too small to be seen on this plot and was typically around 100 times smaller than the plotted average.

The end to end distance provides additional information concerning the peptide conformation. Below, in Figure 5.5, the end-to-end distance between the first carbon of the peptide, within the acetylated N-terminus, and the alpha carbon of the distal histidine residue are plotted. As before, with the alpha carbon RMSD analysis, there are two regions of interest:

near the interface, the distance increases slightly from around 0.9 nm to 0.98 and 1.09 nm, respectively, at z = 0.0 and 0.23 nm; in the aqueous region, the distance substantially fluctuations between z = 2.4 and 3.1 nm between 0.9 and 1.33 nm. Both of these regions indicate that the peptide is in an elongated form.

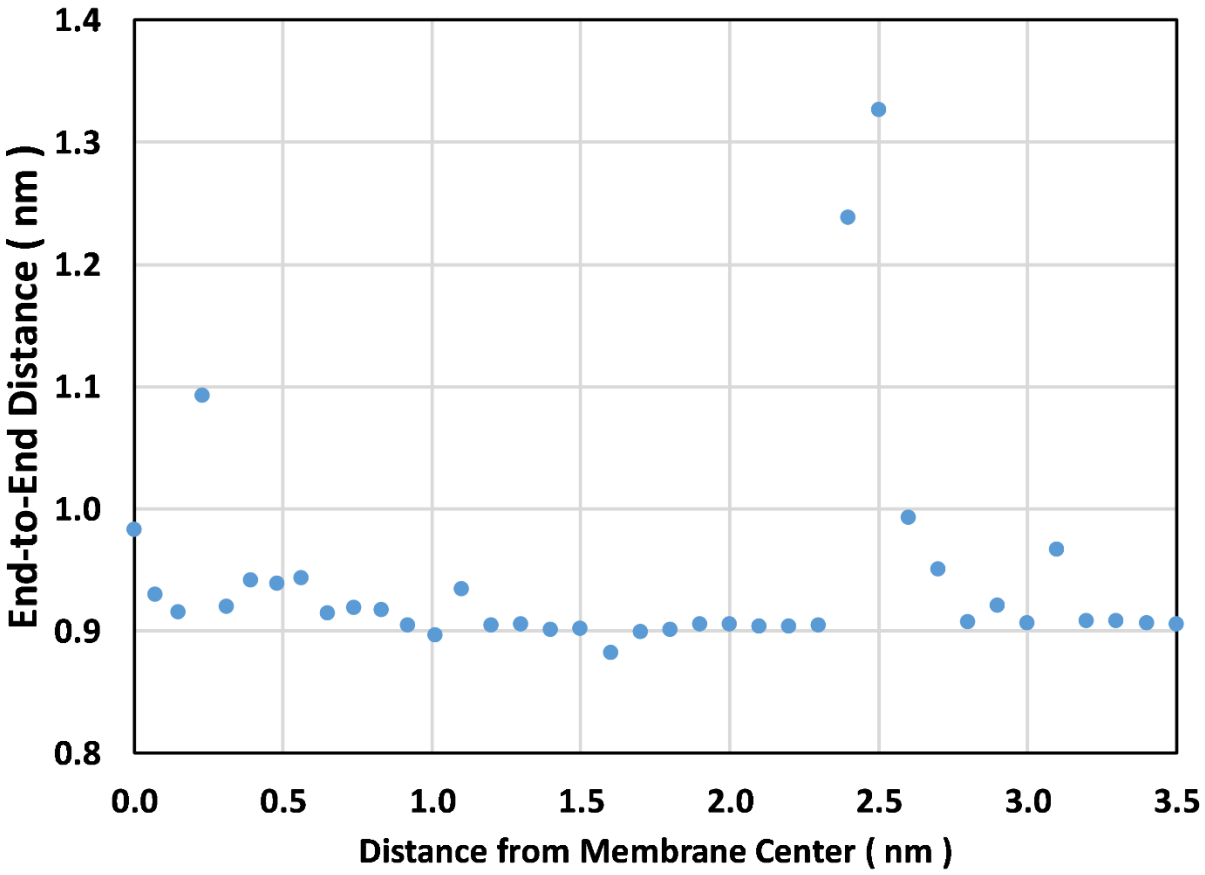


Figure 5.5. The distance between the two ends of the *wh5* molecule. Specifically, the distance from the first carbon atom of the peptide until the alpha carbon of the histidine residue on the other end. Note: the standard uncertainty range of each average was roughly one thousand times smaller than the plotted data points and cannot be seen.

Similarly, Figure 5.6 depicts the distance between the center of mass of the tryptophan and histidine side-chains. Again, there are roughly two regions of interest: between z = 2.4 nm

and z = 3.1 nm, the distance between sidechains fluctuates and reaches a relative maximum around z = 2.4 and 2.5 nm of around 1.15 nm. Similar fluctuations can also be seen between z = 0.0 and 1.0 nm. Equally interesting, the sidechain distance within the interfacial region from z = 1.1 nm and z = 2.3 nm is remarkably stable.

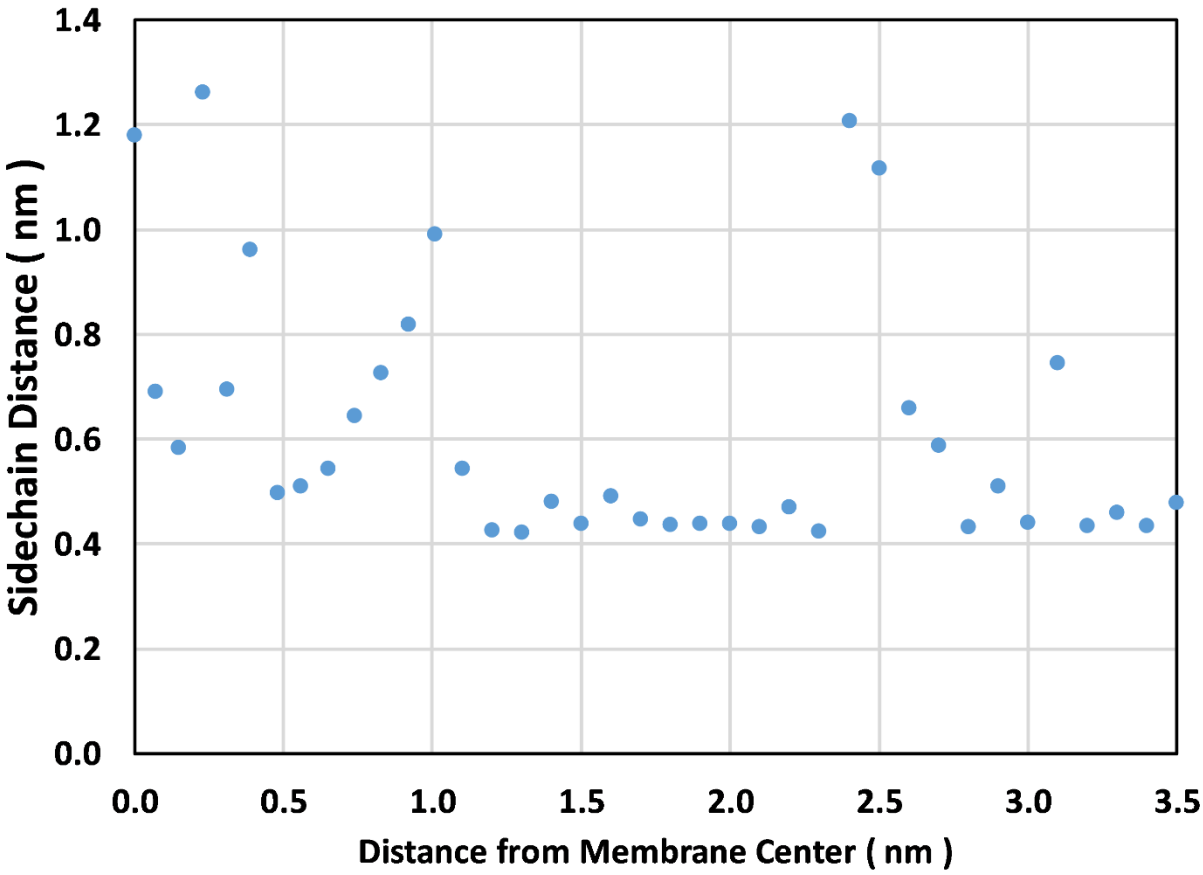


Figure 5.6. The distance between the center of mass of the tryptophan and histidine sidechains in *wh5*. Note: the standard uncertainty range of each average was roughly five hundred times smaller than the plotted data points and cannot be seen.

Below, in Figure 5.7, the radius of gyration for *wh5* is depicted. The radius of gyration is a measure of how linear and compact the peptide is. The longer and more linear the molecule is, the greater the radius of gyration. The more compact and spherical the molecule is, the

smaller the radius of gyration. The same trends and patterns as the previous structural properties are also observed here. However, the pattern between z = 0.5 nm and z = 1.0 nm that was seen in the previous paragraph is more pronounced with the radius of gyration and indicates that *wh5* is elongating at z = 1.0 nm before becoming more compact at z = 0.5 nm, after which the radius of gyration fluctuates at the center of the membrane. Due to the greater mobility of the interdigitating lipid tails at the center of the membrane, the fluctuations there make sense. However, the elongation of *wh5* around z = 2.5 nm, before it reaches the interfacial region, is interesting and warrants additional investigation.

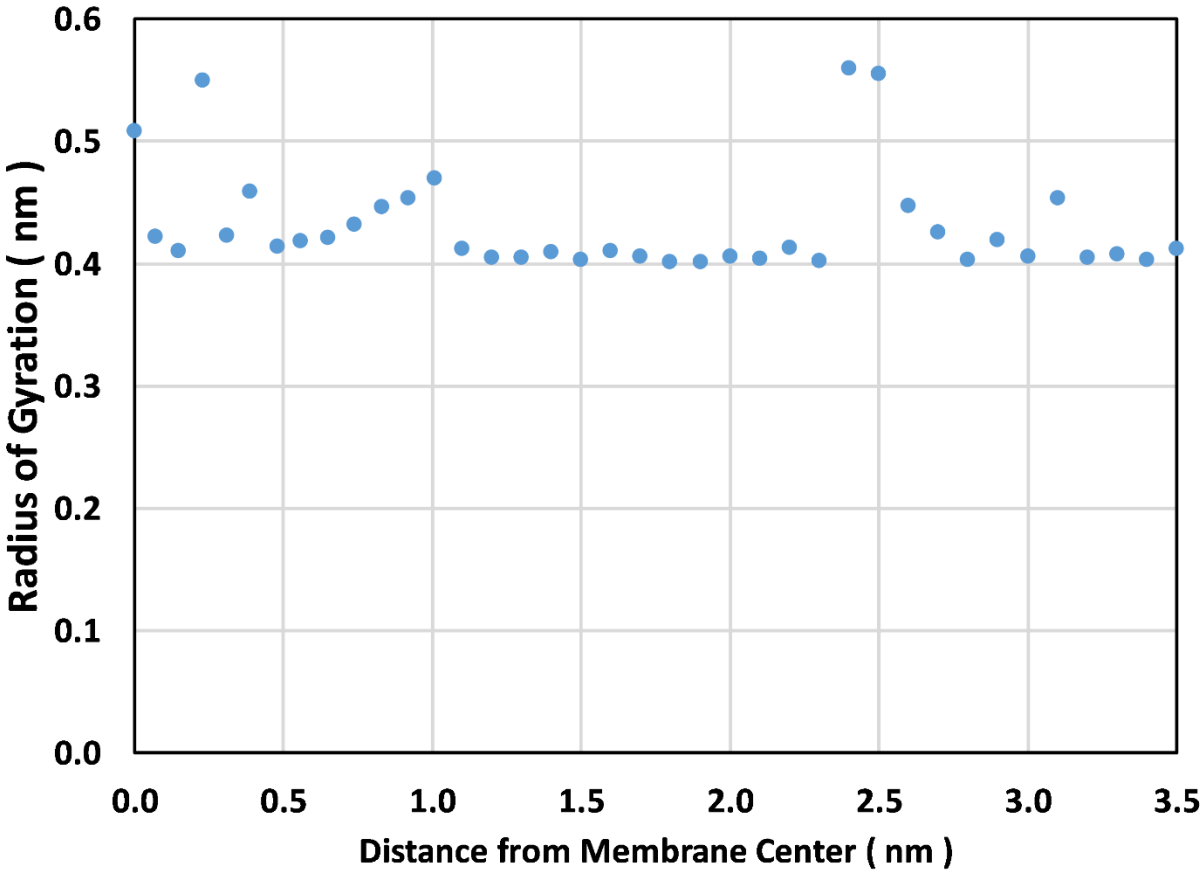


Figure 5.7. The radius of gyration for *wh5*. Note: the standard error for each average was roughly three thousand times smaller than the plotted data points and cannot be seen.

Figure 5.8, shown below, begins our exploration into hydrogen bonding by plotting all of the hydrogen bond pairs within 0.35 nm of *wh5*. Again, a similar pattern emerges where there are approximately 5 to 6 fewer hydrogen bonding pairs at z = 2.4 nm and z = 2.5 nm, as compared to other regions. There is also a slight decrease near the center of the membrane. This indicates, that as *wh5* elongates in this region, it breaks hydrogen bonds either within itself or with its surrounding solvation shell. Remarkably, a large number of hydrogen bonding pairs are present as *wh5* permeates the membrane, indicating that it is likely dragging along a large number of water molecules throughout this process.

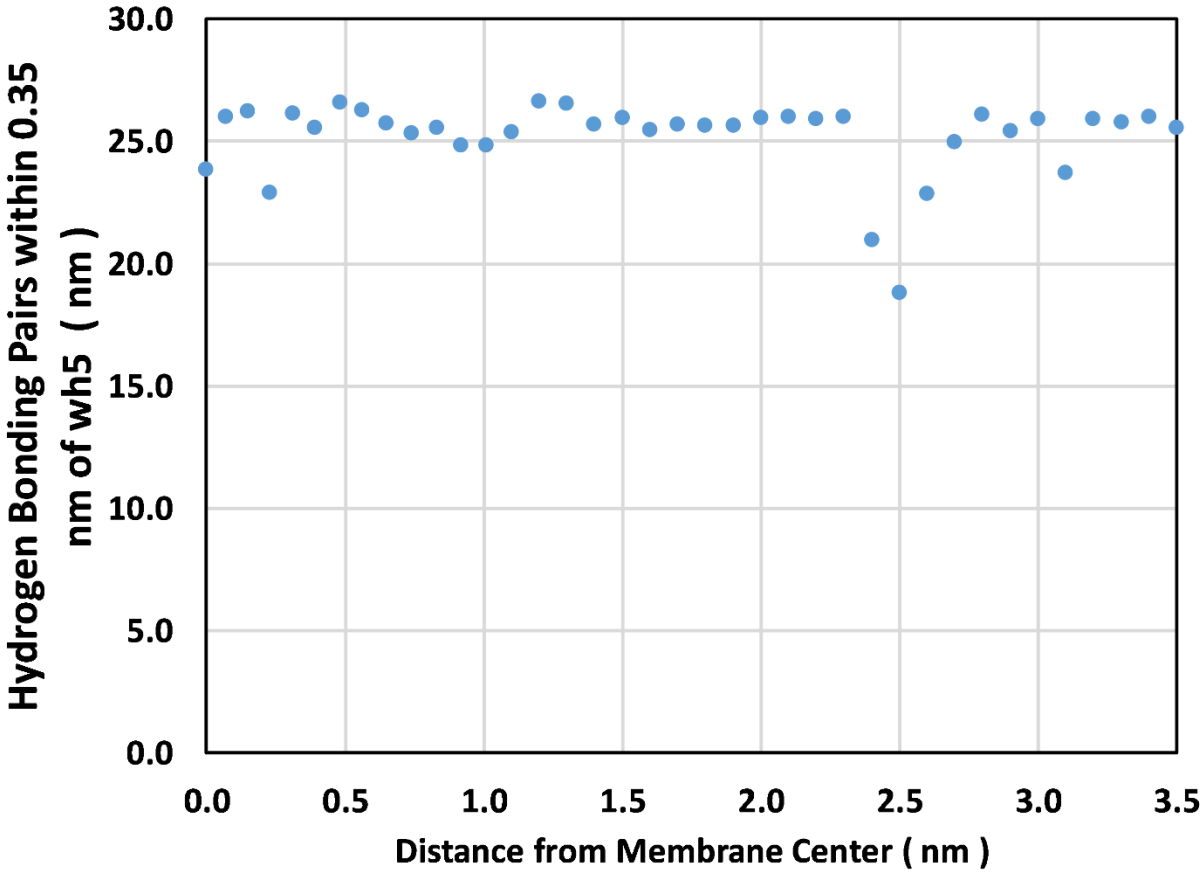


Figure 5.8. The number of hydrogen bond pairs within 0.35 nm of *wh5*. Note: the standard error for each average was roughly three hundred times smaller than the plotted data points and cannot be seen.

However, the previous figure plots all hydrogen bonding pairs. In the next figure, the distance between the oxygen and nitrogen atoms involved in the three hydrogen bonds that form the *wh5* alpha helix are plotted. Hydrogen bonds 1, 2, and 3, are the same as those presented in Figure 5.2 in the Introduction. For the most part, these three hydrogen bonds follow the same patterns as before. At z = 2.5 nm, the hydrogen bonds that hold the alpha helix together are broken apart and then reformed as *wh5* approaches the interface. At the center of the membrane and out in the aqueous region, these distances fluctuate and increase. Interestingly, however, experimental results indicate that the *wh5* helix forms and unravels in water every 2 to 7 ns, yet this is not seen at 3.5 nm in our simulation.[1, 19]

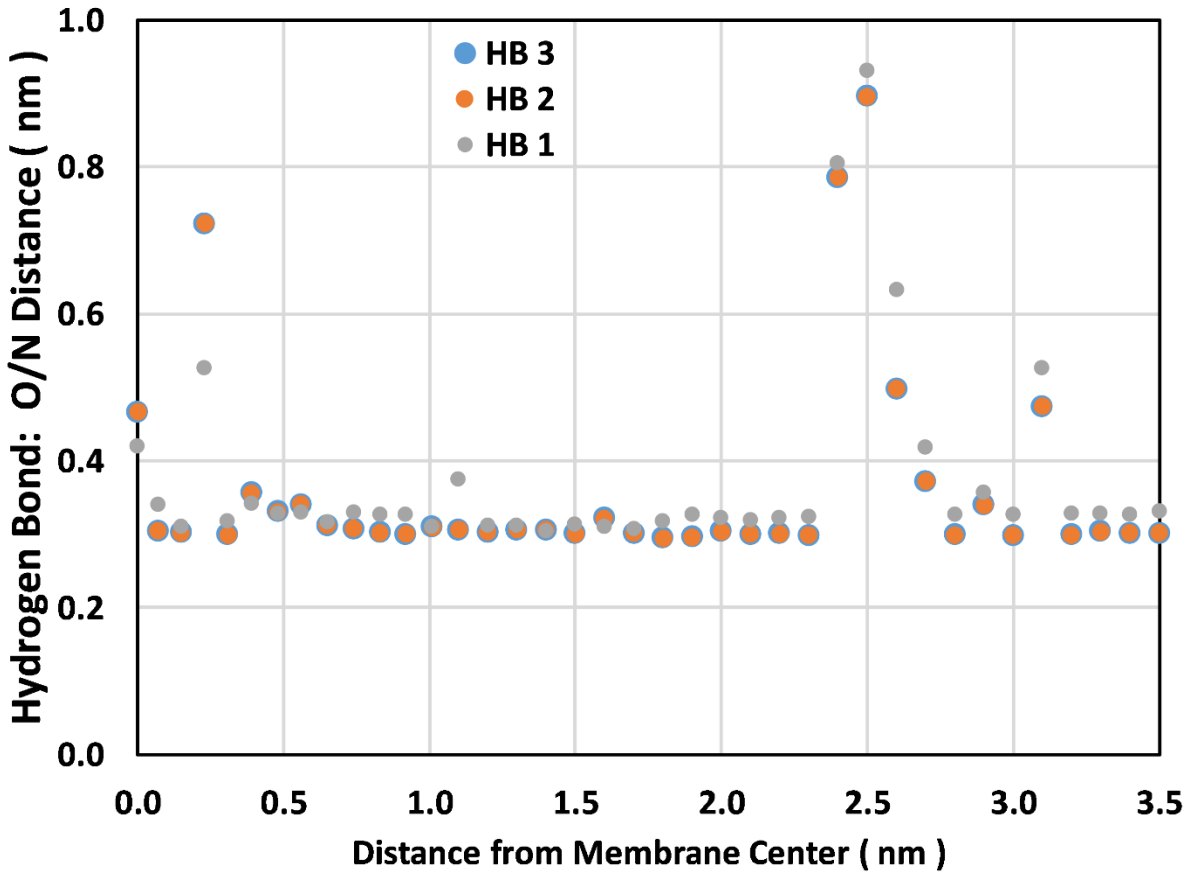


Figure 5.9. The oxygen to nitrogen distance for each of the three hydrogen bonds that hold together the *wh5* alpha helix. The small grey circles depict hydrogen bond 1; the larger orange circles depict hydrogen bond 2; the large blue circles depict hydrogen bond 3. Note: the standard error for each average was roughly fifteen hundred times smaller than the plotted data points and cannot be seen.

5.5. Conclusions

Compared to the dipeptides in Chapters 2 and 3, *wh5* is much larger and is capable of forming an alpha helix through three, strategic hydrogen bonds. The potential of mean force indicates large membrane deformations as the peptide permeates the membrane; however, it still exhibits the same pattern of an interfacial free energy minimum and a maximum at the

center of the lipid bilayer. The extent of the membrane deformation is also indicated by the hydrogen bonding pairs within 0.35 nm analysis, where a large number of hydrogen bonding pairs are present around *wh5* even at the center of the bilayer, where few would be expected. This indicates that a large solvation shell is being carried along by *wh5*. All of the remaining structural analyses – alpha carbon root mean square distance, end-to-end distance, tryptophan and histidine sidechain separation distance, the radius of gyration, and the O to N distance in hydrogen bonds 1, 2, and 3 – all indicate that the helix is unravelling around z = 2.5 nm with some fluctuations in the center of the membrane and farther out in the aqueous region. However, this is not seen at 3.5 nm where experimental results indicate the *wh5* helix forms every 2 – 7 ns in water. The radius of gyration and sidechain separation distance analyses also indicate an interesting pattern from z = 1.0 nm to z = 0.5 nm that warrants further investigation.

One objective of this Chapter was to demonstrate how similar techniques to the previous chapters can be used to explore larger and, perhaps, more interesting molecules. Towards that end, there remains a wide variety of avenues for further investigation. For example, a clustering analysis on all of the phi, psi, chi-1, and chi-2 angles can yield additional information concerning specific conformations that are responsible for the previous data. Autocorrelation functions concerning the tryptophan sidechain orientation can be calculated to provide insight into sidechain reorientation and also provide experimentally verifiable information. Previous studies[2] have also indicated that patterns of hydrogen bond breaking and formation are responsible for the unraveling and raveling of the *wh5* alpha helix; these patterns can also be examined. Finally, autocorrelation functions can also be calculated for oxygen to nitrogen distances in order to determine the time scale of alpha helix formation.

5.6. References

1. Mohammed, O. F.; Jas, G. S.; Lin, M. M.; Zewail, A. H., Primary Peptide Folding Dynamics Observed with Ultrafast Temperature Jump. *Angewandte Chemie International Edition* **2009,** *48* (31), 5628-5632.

2. Jas, G. S.; Hegefeld, W. A.; Májek, P.; Kuczera, K.; Elber, R., Experiments and Comprehensive Simulations of the Formation of a Helical Turn. *The Journal of Physical Chemistry B* **2012,** *116* (23), 6598-6610.

3. Lee, B. L.; Kuczera, K.; Middaugh, C. R.; Jas, G. S., Permeation of the three aromatic dipeptides through lipid bilayers: Experimental and computational study. *J. Chem. Phys.* **2016,** *144* (24), 14.

4. Jo, S.; Kim, T.; Im, W., Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS ONE* **2007,** *2* (9), e880.

5. Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W., CHARMM-GUI *Membrane Builder* toward realistic biological membrane simulations. *Journal of Computational Chemistry* **2014,** *35* (27), 1997-2004.

6. Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W., CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **2009,** *97* (1), 50-58.

7. Pastor, R. W.; MacKerell, A. D., Development of the CHARMM Force Field for Lipids. *The Journal of Physical Chemistry Letters* **2011,** *2* (13), 1526-1532.

8. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **2009**, NA-NA.

9. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926.

10. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995,** *103* (19), 8577.

11. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007,** *126* (1), 014101.

12. Allen, M. P.; Tildesley, D. J., *Computer simulation of liquids*. Clarendon Press: 1989; p 385.

13. Torrie, G.; Valleau, J., Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *Journal of Computational Physics* **1977,** *23* (2), 187-199.

14. Kästner, J., Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011,** *1* (6), 932-942.

15. Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. the Method. *Journal of Computational Chemistry* **1992,** *13* (8), 1011-1021.

16. Souaille, M.; Roux, B., Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer Physics Communications* **2001,** *135* (1), 40-57.

17. Roux, B., The calculation of the potential of mean force using computer-simulations. *Computer Physics Communications* **1995,** *91* (1-3), 275-282.

18. Nitschke, N.; Atkovska, K.; Hub, J. S., Accelerating potential of mean force calculations for lipid membrane permeation: System size, reaction coordinate, solute-solute distance, and cutoffs. *J. Chem. Phys.* **2016,** *145* (12), 9.

19. Lin, M. M.; Mohammed, O. F.; Jas, G. S.; Zewail, A. H., Speed limit of protein folding evidenced in secondary structure dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (40), 16622-16627.

Chapter 6

Conclusions and Future Directions

6.1 Conclusions:

At the present moment, the field of passive, model membrane permeation by small peptides and other small molecules is rather fragmented. Few studies offer much continuity with the previous body of work; and, when they do, they are typically continuations of an individual research group's work. Alternatively, most studies focus on a very specific goal, such as trying to predict blood/brain permeability[1] or investigating the effect of cholesterol on membrane permeation.[2] In this work, we focus on filling out one of the more fundamental bodies of membrane permeation knowledge: what happens to amino acids when they traverse a lipid bilayer? There has already been substantial work done on the charged amino acids arginine and lysine,[3-6] but only one work looks at aromatic amino acids analogues and presents a rather cursory analysis.[7] We, therefore, chose to conduct a comprehensive examination of the resonant amino acids – phenylalanine, tyrosine, and tryptophan – due to their relative ease of lipid bilayer permeation in their amidated and acetylated forms and prior experimental studies. This analysis was then furthered by exploring the effect of system size and lipid type on phenylalanine permeation and by conducting a review of diffusion constant methods – one of the more difficult and most ambiguous analyses. Our work was then broadened towards the analysis of a larger peptide, *wh5*, which is capable of forming tertiary structure. In this section, we summarize our key findings from each chapter within the broader context of general membrane permeation.

In Chapter 2, we conducted a permeation analysis on the acetylated and amidated forms of phenylalanine, tyrosine, and tryptophan. The potential of mean force for all three

peptides followed the same pattern that has been observed for other amphiphilic compounds: an interfacial free energy minimum followed by a free energy maximum at the center of the lipid bilayer. Diffusion constants, as determined by a numerical approximation to the Smoluchowksi equation, were slightly smaller than experimental values in the aqueous region and, surprisingly, do not demonstrate a dependence on membrane insertion depth. The permeation time was much smaller and the permeability coefficients were much larger than experimental values, likely due to underestimating the free energy barriers in non-translational motion. This was further explored by examining the rotation of the sidechain and the entire molecule, which indicated substantial barriers to rotational motion that have been observed in many previous studies. Interestingly, each peptide adopted different rotational conformations in the aqueous, interfacial, and center regions of the model membrane. In the aqueous region, all peptides were in either alpha-helix or beta-sheet conformers; however, in the center of the membrane, conformations that minimized the free energy by bringing the terminal oxygen and nitrogen atoms closer together were adopted. All of this indicates that the translational permeation of the dipeptide involves significant rotational and conformational changes. In all cases, particularly with tryptophan, each dipeptide held onto a small solvation shell within the center of the membrane, which is a common observation for more polar and charged amino acids. In general, the permeation of the resonant dipeptides is similar to other amphiphilic molecules but encounters significant rotational barriers and undergoes conformational changes that adapt to the hydrophobic environment of the membrane interior.

In Chapter 3, we examined the effect of changes in system size and lipid type as applied to the phenylalanine dipeptide. Specifically, the analyses in Chapter 2 were repeated for the

permeation of the phenylalanine dipeptide in a 50 vs. 40 POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine) lipid system and in a 50 POPC vs. 50 DOPC (1,2-dioleoyl-*sn*-glycero-3-phospocholine) lipid system. Again, the same general trends in terms of the potential of mean force, diffusion constants, rotational barriers, and other conformational analyses were observed, lending additional credence to all of the previous findings in Chapter 2. However, a few small differences were noticed: the larger, POPC 50 system possess a deeper interfacial potential of mean force due to greater membrane deformations and additional *X-1* and *X-2* conformers were witnessed. This supports other studies that have found that larger systems may be required to accurately capture the potential of mean force. Similarly, the DOPC 50 system possessed a broader PMF due to greater disorder in the monounsaturated acyl tails and also accelerated the depth at which lower energy conformers were adopted. In both analyses, the relationship between system size and lipid type on rotational dynamics has not been investigated before, and the appearance of additional sidechain conformers in larger systems and the faster adoption of low energy conformers in DOPC may have important implications as other researchers attempt to couple rotational and translational dynamics to more accurately determine the kinetic properties of membrane permeation.

In terms of kinetic and energetic analyses, the potential of mean force has been thoroughly examined; however, only one study attempted a comprehensive analysis on diffusion constant data and focused solely on water in previous studies that differed dramatically in terms of temperature, force field, lipid type, and other molecular dynamics simulation parameters. We attempted to fill this void by examining five different methods towards the diffusion of the phenylalanine dipeptide in POPC 40, POPC 50, and DOPC 50

systems. The Fluctuation-Dissipation theorem is often applied to umbrella sampling pulling forces and produced results atypical of all other methods. However, when applied to non-constrained force data, this method produces acceptable results. Green-Kubo expressions for calculating diffusion constants provided the fastest calculation method but were hindered by a large degree of uncertainty, likely due to backscattering. Einstein relations are the second most common method for calculating diffusion constants but are highly dependent upon subjective stability regions to conduct a linear fit. The Hummer Displacement method uses position autocorrelation data but does not yield sensical results due to a wide variety of complicated assumptions in its derivation. A numerical approximation to the Smoluchowski equation is our preferred method due to results that are consistently more precise and more easily analyzable than the other methods. To our knowledge, no one has conducted such a survey of diffusion constant methods applied to passive membrane permeation, and our results emphasize that greater care, thought, and documentation should be taken in this field when calculating diffusion data.

After the diffusion digression, we then turned our attention towards applying our toolset of passive permeation analyses to *wh5*, a small peptide capable of forming alpha helices. The potential of mean force indicated large membrane deformations and was very broad; however, it still exhibits the same pattern of an interfacial free energy minimum and a maximum at the center of the lipid bilayer. A wide variety of structural analyses then indicated unusual activity, such as the unraveling of the peptide helix, around 2.5 nm and near the center of the lipid bilayer. These results were rather surprising because experimental results in water indicate that the helix folds and unfolds within 2 to 7 ns.[8-9] The presence of the lipid bilayer is

influencing the tertiary structure of the peptide, and further analysis may yield important insights concerning the formation of helices within more complicated biology systems.

6.2 Future Directions:

6.2.1. – Near Future.

Chapters 2 and 3 provide a detailed study of the aromatic dipeptides. These studies can be easily expanded to include other amino acids and may provide information that can be used to predict the behavior of more complicated peptides, such as *wh5*. Many of the individual data analyses can also be improved. For example, the potential of mean force can be more easily and accurately calculated using milestoning[10] or transition-tempered metadynamics.[11] All of the sidechain and insertion vector analyses could benefit from longer simulation times in order to better investigate the role that rotational motion plays in permeation processes. Mixtures of lipids and the addition of cholesterol can also be examined in order to better simulate real cell membranes, which possess intriguing phase behavior.[12-13] Finally, lipid bilayers, by themselves, also possess interesting properties that can be investigated. For example, membrane fluctuations, order parameters, acyl chain rotations, and transient void spaces can be simultaneously studied.[12-14]

Chapter 4 examined existing diffusion constant methods, as applied to passive membrane permeation. Unfortunately, passive membrane permeation is an inherently non-Markovian process that does not follow a series of random motions. The permeation of molecules involves an intricate interplay of rotational dynamics, void space formations, lipid

translations, lipid flips, lipid rotations, and even the temporary breaking of the membrane itself. These processes are based upon the history of intermolecular motions and direct the motion of a permeate. There have been attempts to create analytical, memory-based methods for modeling diffusion; but these, like the inhomogeneous solubility diffusion model, are relatively simplistic. Short of modeling the entire permeation process, we believe the future in terms of more accurately calculating diffusion constants lies with rate based approaches. For example, the Hummer Displacement method is actually an appendix in an article by Hummer that outlines such an approach: many short trajectories are run in order to sample the behavior of the permeate. These short trajectories are then used to create a rate based, Markovian approach where the permeate has a certain probability of undergoing a specific motion.[15] In addition, milestoning represents a similar approach that bypasses the need for diffusion constant data and directly calculates the kinetics of a permeation process, such as the mean first passage time.

Finally, Chapter 5 and the analysis of small helical peptides like *wh5* represent an invaluable opportunity to predict how helices are formed on an atomistic level. Unfortunately, as mentioned in Chapter 5, there remain a substantial number of short-term analyses that can be conducted. However, experimental results indicate that the *wh5* helix is formed every 2 to 7 ns in water,[8] but our preliminary results show no evidence of that near the edge of our simulation box. Therefore, the presence of the lipid bilayer may be influencing the conformation of *wh5* at such a distance. What then, concerning our atomistic investigations, is responsible for this behavior and how can it be used to draw conclusions about helical formations and other types of tertiary structures? The answers to such questions can then be

used to design molecules with functional behavior near cell membranes that utilize these types of tertiary structure.

6.2.2. – Far Future.

There are two primary foci for the studies conducted within this Dissertation:

1.) Describe, analyze, and interpret the basic mechanisms of small peptide permeation through lipid bilayers.

2.) Aid the intelligent design of pharmaceutical compounds.

In terms of the primary focus, these studies can be replicated to all amino acids in order to determine how they behave in lipid bilayers which model a cell membrane. These general patterns and insights can then be extrapolated to tertiary structures, like alpha helix and beta sheet formation, before then being tested on larger peptides like *wh5*. Eventually and hopefully, these bottom-up analyses will merge with insights obtained from bioinformatics and other top-down approaches that focus on the function of much larger biological mechanisms. This is where the secondary focus becomes a reality: over 70% of all modern pharmaceutical targets[16] interact or regulate the permeation of molecules across cell membranes. The more knowledge we possess concerning the operation and function of permeation process – passive, assisted, and active – the faster, more efficient, and less costly, we can devise new pharmaceutical treatments for a wide variety of illnesses and disorders.

6.3 References

1. Carpenter, Timothy S.; Kirshner, Daniel A.; Lau, Edmond Y.; Wong, Sergio E.; Nilmeier, Jerome P.; Lightstone, Felice C., A Method to Predict Blood-Brain Barrier Permeability of Drug-Like Compounds Using Molecular Dynamics Simulations. *Biophys. J.* **2014,** *107* (3), 630-641.

2. Starke-Peterkovic, T.; Turner, N.; Vitha, M. F.; Waller, M. P.; Hibbs, D. E.; Clarke, R. J., Cholesterol Effect on the Dipole Potential of Lipid Membranes. *Biophys. J.* **2006,** *90* (11), 4060-4070.

3. Hu, Y.; Ou, S. C.; Patel, S., Free Energetics of Arginine Permeation into Model DMPC Lipid Bilayers: Coupling of Effective Counterion Concentration and Lateral Bilayer Dimensions. *Journal of Physical Chemistry B* **2013,** *117* (39), 11641-11653.

4. Cardenas, A. E.; Shrestha, R.; Webb, L. J.; Elber, R., Membrane Permeation of a Peptide: It Is Better to be Positive. *Journal of Physical Chemistry B* **2015,** *119* (21), 6412-6420.

5. Neale, C.; Bennett, W. F. D.; Tieleman, D. P.; Pomes, R., Statistical Convergence of Equilibrium Properties in Simulations of Molecular Solutes Embedded in Lipid Bilayers. *Journal of Chemical Theory and Computation* **2011,** *7* (12), 4175-4188.

6. MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P., Transfer of Arginine into Lipid Bilayers Is Nonadditive. *Biophys. J.* **2011,** *101* (1), 110-117.

7. MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P., Distribution of Amino Acids in a Lipid Bilayer from Computer Simulations. *Biophys. J.* **2008,** *94* (9), 3393-3404.

8. Mohammed, O. F.; Jas, G. S.; Lin, M. M.; Zewail, A. H., Primary Peptide Folding Dynamics Observed with Ultrafast Temperature Jump. *Angewandte Chemie International Edition* **2009,** *48* (31), 5628-5632.

9. Lin, M. M.; Mohammed, O. F.; Jas, G. S.; Zewail, A. H., Speed limit of protein folding evidenced in secondary structure dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (40), 16622-16627.

10. Cardenas, A. E.; Elber, R., Markovian and Non-Markovian Modeling of Membrane Dynamics with Milestoning. *Journal of Physical Chemistry B* **2016,** *120* (33), 8208-8216.

11. Dama, J. F.; Rotskoff, G.; Parrinello, M.; Voth, G. A., Transition-Tempered Metadynamics: Robust, Convergent Metadynamics via On-the-Fly Transition Barrier Estimation. *Journal of Chemical Theory and Computation* **2014,** *10* (9), 3626-3633.

12. Baoukina, S.; Rozmanov, D.; Tieleman, D. P., Composition Fluctuations in Lipid Bilayers. *Biophys. J.* **2017,** *113* (12), 2750-2761.

13. Mills, T. T.; Toombes, G. E. S.; Tristram-Nagle, S.; Smilgies, D. M.; Feigenson, G. W.; Nagle, J. F., Order parameters and areas in fluid-phase oriented lipid membranes using wide angle x-ray scattering. *Biophys. J.* **2008,** *95* (2), 669-681.

14. Shinoda, W.; Okazaki, S., A Voronoi analysis of lipid area fluctuation in a bilayer. *The Journal of Chemical Physics* **1998,** *109* (4), 1517.

15. Hummer, G., Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New Journal of Physics* **2005,** *7*, 34-34.

16. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L., Opinion - How many drug targets are there? *Nature Reviews Drug Discovery* **2006,** *5* (12), 993-996.