7-10-2017

Advantages and Limitations of Different Resolution Force Fields to Study Membrane Active Peptides

Michael Ward
University of Connecticut - Storrs, michael.d.ward@uconn.edu

Advantages and Limitations of Different Resolution Force Fields to Study Membrane Active Peptides

Michael D. Ward

B.S. University of Connecticut, 2015

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

At the

University of Connecticut

2017

APPROVAL PAGE

Masters of Science Thesis

Advantages and Limitations of Different Resolution Force Fields to Study Membrane Active Peptides

Presented by

Michael D. Ward, B.S.

Major Advisor_____
Eric R. May, Ph.D.

Associate Advisor_____
Nathan N. Alder, Ph.D.

Associate Advisor_____
José A. Gascón, Ph.D.

University of Connecticut

2017

ACKNOWLEDGEMENTS

Computational resources have been provided through the University of Connecticut Booth Engineering Center for Advanced Technology. Financial support has been provided from NIH grants awarded to Eric May (grant numbers: K22-AI099163, R35-GM119762).

Table of Contents

Chapter 1: Introduction….....1

1.1 Membrane Active Peptides……...1

1.2 Molecular Dynamics Simulations…….7

1.3 References……12

Chapter 2: Behavior of the NωV γ-peptide in a Membrane Environment……....14

2.1 Introduction……...14

2.2 Methods…….19

2.3 Results and Discussion……...22

2.4 Conclusions……..27

2.5 References……...29

Chapter 3: Evaluation of the Hybrid Resolution PACE Model for the Study of Folding, Insertion, and Pore Formation of Membrane Associated Peptides……...30

3.1 Abstract…….30

3.2 Introduction……...30

3.3 Methods…….36

3.4 Results and Discussion……...42

3.5 Conclusions…….....58

3.6 References……60

Chapter 4: Closing Remarks and Future Directions……63

Chapter 1:

Introduction

MEMBRANE ACTIVE PEPTIDES:

Peptides with activity against cell membranes (membrane active peptides, MAPs) have attracted attention as a potential therapeutic targeting and delivery device in cancer treatments and can also have a cytotoxic effect on tumor cells[1,2,3,4,5]. In nature, these peptides evolved as antimicrobial agents[6,7] or viral factors that aid in cell entry[8,9,10]. Many antimicrobial peptides, such as Melittin, are known to form membrane pores[11]. Enveloped viruses employ fusion peptides, such as Hemagglutinin from Influenza virus, which disrupt and reorganize host cell membranes to help fuse the host membrane to the viral membrane envelope[12]. Nonenveloped viruses release MAPs to disrupt membranes to help the virus escape endosomes of host cells[8]. Typically, MAPs are less than 50 amino acids in length, have helical propensity, and are amphipathic. While these structural and mechanistic details cover general cases of MAPs, the variety of membrane-peptide interactions employed is extensive. Also, these peptides can be activated by different conditions such as pH, membrane composition, concentration, receptor binding, etc. This variety is what makes MAPs promising as therapeutics. They have the potential to exclusively target and destroy specific cell types, like in tumor microenvironments where the extracellular pH is different than in normal tissue[13]. In order to develop therapeutics based on MAPs, it is important to understand the details of how they achieve membrane disruption. The rest of this section will highlight the

details for a few well-studied MAPs as well as identify MAPs that deserve more attention and how they can be studied.

Melittin

Melittin is a highly studied peptide that has become a model system to examine the characteristics and mechanisms of pore forming lytic peptides. Melittin is the main toxic component in bee venom and has been shown to have hemolytic and antimicrobial activity. Melittin is a cationic peptide 26 amino acids in length (N-GIGAVLKVLTTGLPALISWIKRKRQQ-C). Many experiments have been performed to determine the biophysical characteristics of Melittin in solution as well as interacting with lipid membranes. NMR, CD, fluorescence, and neutron diffraction experiments have all contributed to understanding secondary structure, aggregation, pore size, and peptide orientation in a membrane[14]. While there are a variety of variables (e.g. pH, membrane composition, peptide concentration, temperature) that vary these characteristics, a general mechanism for Melittin pore formation has been established. Melittin is mostly disordered in solution, but becomes helical upon membrane binding. The helix is interrupted with a small kink by a proline at residue 14. At low peptide concentrations, Melittin remains parallel to the membrane and does not induce pores[15]. However, at increased concentrations, Melittin can insert perpendicular to the membrane to form transient, toroidal pores[16]. These transient pores permit Melittin molecules to cross the membrane allowing Melittin molecules to adsorb to both sides of the membrane, which stabilizes the pores[17].

In addition to these experimental studies, Molecular Dynamics simulations (MD) have been employed to resolve molecular details inaccessible by experiment. While experiments have helped describe *what* Melittin does, MD can be used to explain *how* it works at an atomistic level. For instance, from experiments we know that Melittin folds into an α-helix on the membrane, which gives it lytic activity. MD has been used to understand *how* Melittin folds into an α-helix on the lipid bilayer[18]. In POPC bilayers, cationic Lysine and Arginine residues stabilize contact with the membrane via electrostatic interactions with anionic phosphate headgroups. The central proline causes kinked conformations that promote sampling of deeply inserted orientations. Being deeply inserted in a hydrophobic core promotes helix formation because backbone hydrogen bond formation is no longer competing with hydrogen bonds between backbone and water. One can incorporate these residue-level details of folding into a framework for the rational design of improved therapeutic MAPs[45].

Additionally, MD has been used to understand how Melittin translocates from an inactive membrane surface-bound state to a perpendicular orientation across the membrane[19]. In the membrane bound state, Melittin causes an increase in area per lipid as well as thinning of the membrane. The increase in area per lipid creates increased exposure of the hydrophobic lipid tails to polar solvent. To protect from this, additional Melittin peptides bind in the gaps, burying their hydrophobic residues into the membrane core. This has a cooperative effect leading to aggregation of membrane-bound Melittin peptides that thin the membrane further. On occasion, this perturbation induces a lipid

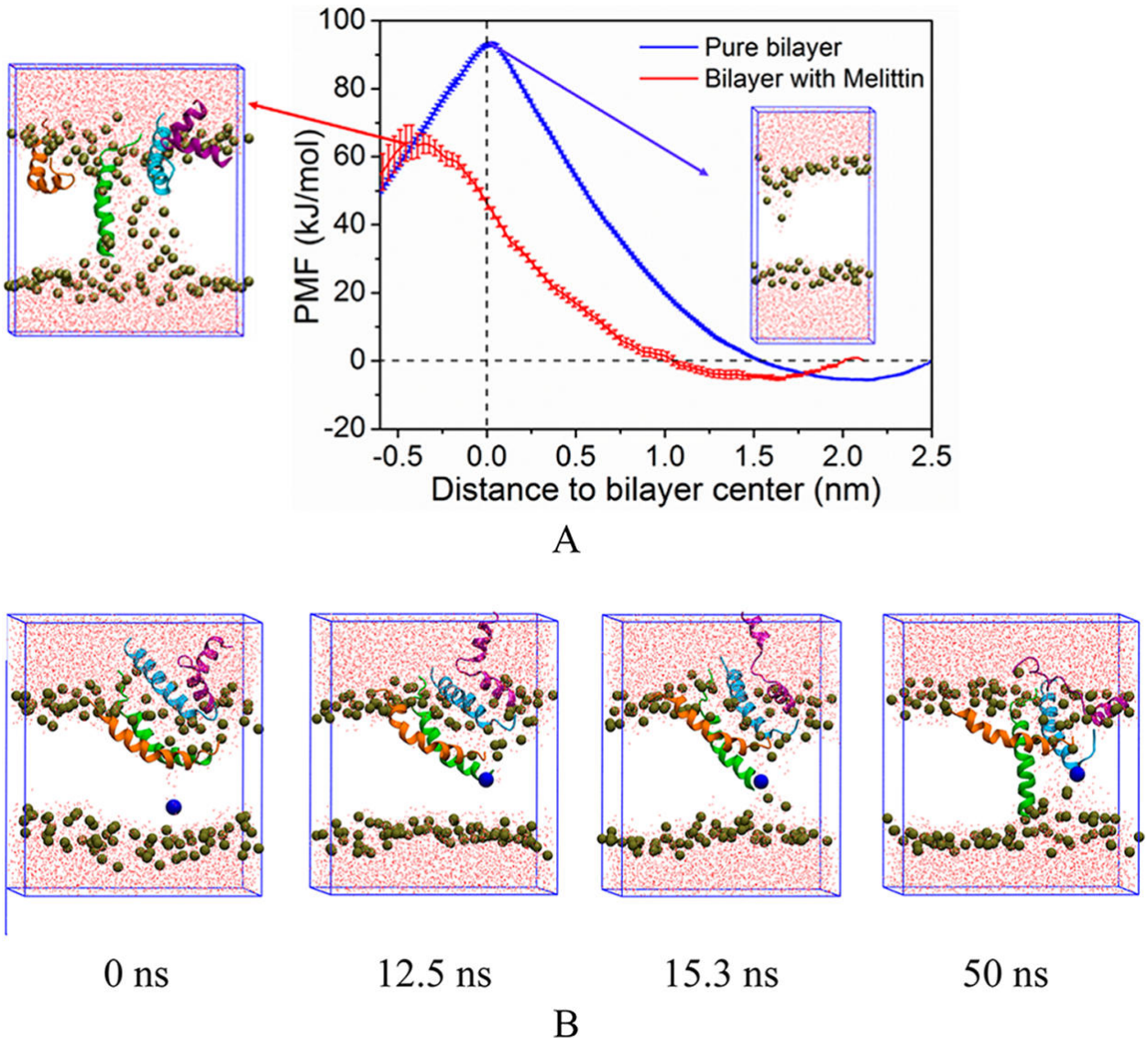

Figure 1. Lipid flip-flop nucleates Melittin pores. A) Free energy cost of lipid flip-flop is lowered by ~30kJ/mol when Melittin is present. B) Snapshots along the insertion process. Head group of first lipid flip-flop is highlighted in blue. (Adapted from D. Sun et al.)[19]

flip-flop causing polar head groups from the upper leaflet to flip and move to the lower leaflet (**Fig. 1B**). Free energy calculations using MD have shown that the energetic cost of the lipid flip-flop is lowered by ~30 kJ/mol when the lipid head group translocating across the membrane is accompanied by a Melittin peptide (**Fig. 1A**). This reduction in free energy is explained by the favorable interaction between the positively charged N-

terminus of Melittin and the negatively charged phosphate of the DPPC lipid as they cross the membrane. This event nucleates a pore as other lipids and peptides dive across the membrane following the defect.

HAfp

Hemagglutinin fusion peptide (HAfp) from the enveloped Influenza virus is an example of a well-studied MAP that is activated via a pH dependent mechanism. In order to enter host cells, hemagglutinin glycoproteins (HA) on the viral envelope bind to host cell receptors to be taken up by receptor-mediated endocytosis. HA undergoes a cleavage that leaves two subunits, HA1 and HA2. At low pH, HA2 is responsible for membrane fusion. To accomplish fusion, the newly generated N-terminus of HA2 (fusion peptide, HAfp) inserts into the endosomal membrane, while the C-terminal transmembrane domain keeps HA2 tethered to its own viral membrane. However, at neutral pH, HAfp is buried on the interior of HA2 (**Fig. 2A**). A pH drop to between 5 and 6 in the endosome causes a conformational change that exposes HAfp, allowing it to bind to the endosomal membrane to initiate the fusion process (**Fig. 2B**)[12,20].

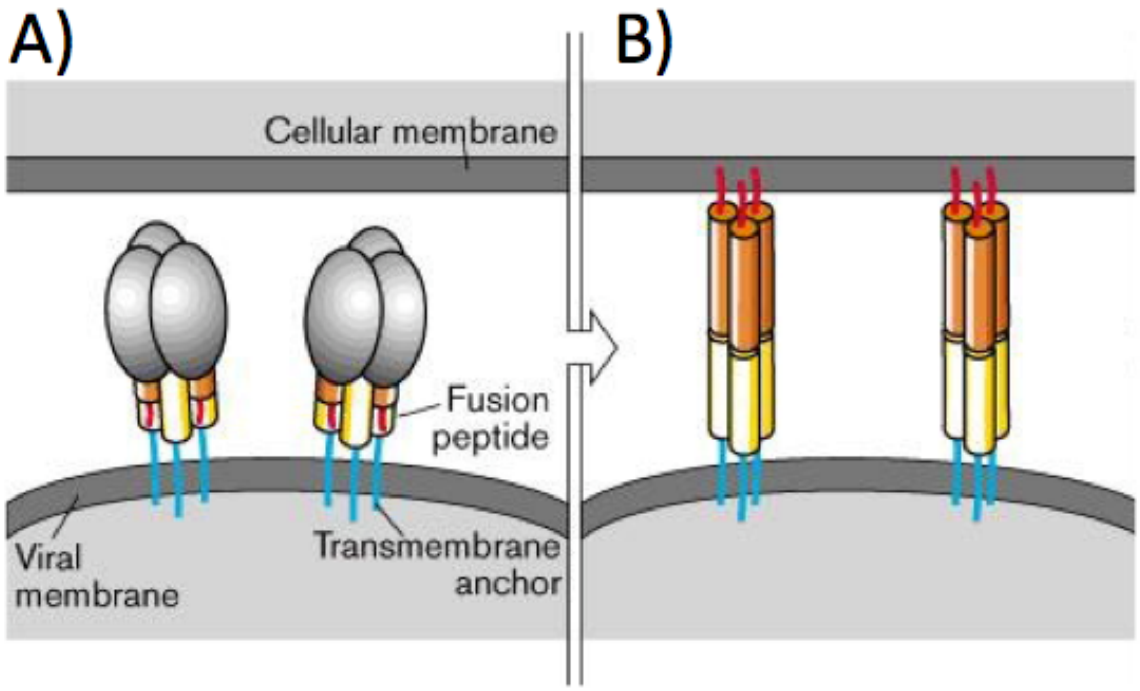

Figure 2. HAfp is exposed and inserted into host cell membranes at low pH. A) Cartoon representation before HA1 binds host cell membrane. HA1 shown in grey, HA2 shown in red and yellow. B) Inside the endosome at low pH, fusion peptide (red) is exposed and anchored in endosomal membrane. (Adapted from Hughson)[46]

Experimental and MD studies have been used to understand how HAfp-membrane interactions cause fusion and how these interactions are affected by pH[21,22]. Using NMR, Lorieau et al. determined the structure of HAfp in the presence of a micelle[21]. When membrane bound, the dominant structure of HAfp (23-residue variant) has a kink at residue 11 that promotes formation of a helical hairpin. This predominant structure does not change between pH 5.0 and pH 7.4. However, the same study notes that at the lower pH there is more conformational flexibility with ~20% of the population sampling an open (less kinked) conformation. This flexibility at low pH is in agreement with an earlier MD study using the 20-residue HAfp variant[22]. Panahi & Feig ran three separate simulations where they protonated one of three acidic residues (E11, E15, D19) to model the pH change from 7.4 to 5. Each protonation resulted in a different

dominant (highly populated) conformation that altered where the kink is and the degree to which HAfp is bent. From these simulations they also calculated the free energy of crossing an implicitly modeled membrane. They found that when E11 is protonated the peptide had the lowest free energy cost of insertion (~5kcal/mol) suggesting that at low pH this residue could become protonated and lead to the initiation of fusion.

MD studies have been a valuable tool to interpret and guide experimental studies on MAPs including both antimicrobial peptides and enveloped viruses. However, there are many MAPs from nonenveloped viruses that have interesting properties (therapeutically and biologically) known from experiment, like pH-dependent activity, which have not been well characterized to establish a structure-function relationship[10]. One example is γ-peptide from *Nudaurelia capensis ω* Virus (NωV), which has a unique activation mechanism (basic pH) for membrane lysis. Chapter 2 will focus on understanding this activation mechanism using MD. Generally, applying MD to these systems holds the promise of understanding nonenveloped viruses at a higher resolution as well as progressing the field of peptide-based therapeutics.

MOLECULAR DYNAMICS SIMULATIONS:

MD simulations have been described as a computational microscope since they provide a view of biomolecular processes with exceptionally high resolution[23]. While traditional microscopes have informed a large portion of our understanding of biology, a deep understanding of biology entails understanding the dynamics of biomolecules at the atomic scale, which is beyond the resolution limit for even super-resolution light

microscopes. Biophysical techniques have helped deepen our understanding of biology by revealing structures and dynamics that can be directly observed. High-resolution structures of biomolecules can be obtained using X-ray crystallography, Cryo-EM, and NMR. For example, static structures of inactive and active src-kinase at atomic resolution have been determined from x-ray crystallography[24,25]. Additionally, the various single molecule techniques and NMR are harnessed to better understand protein dynamics. NMR studies of src-kinase have revealed slow timescale conformational changes and describe how these changes regulate activity[26]. However, none of these biophysical techniques describe atomic details of the *pathways* of these processes. Understanding these pathways is key to understanding how a potential mutation, or the introduction of a drug, will affect a given process. MD simulations provide this information by giving a time-course trajectory of all the atoms in a system. For example, MD simulations have been used to identify pathways that predominate the transition from inactive to active in src-kinase[27]. From these pathways, Shukla et al. identified intermediates as rational states for drug design. With MD simulations, one can see how every single atomic interaction in a system evolves over time. Thus, MD has the potential to describe any biomolecular dynamics with greater spatiotemporal detail than experimental techniques.

The amount of value MD simulations can provide is directly related to two main factors: 1) Timescales accessible and 2) model (force field) accuracy. Simulations must reach timescales of biological processes in order to describe them and the descriptions from simulation are meaningless unless they accurately model the underlying physics.

Over the past decade the timescales accessible to MD simulations has increased substantially, which has increased the utility of MD in studying functionally relevant biomolecular processes. **Fig. 3** shows the timescales associated with protein motions.

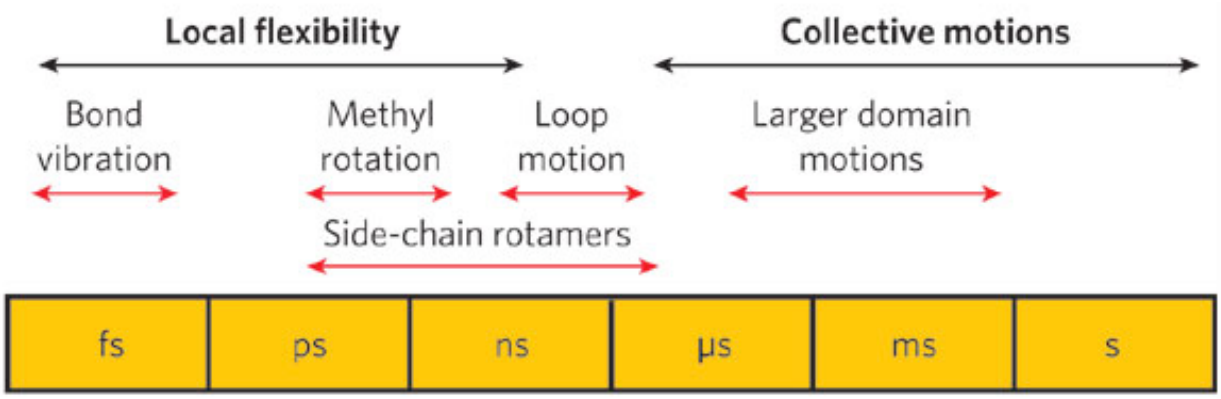

Figure 3. Timescales associated with protein dynamics. (Adapted from Henzler-Wildman & Kern)[28]

By 2007, the longest MD simulation was 2μs[23]. At this timescale, one can study local flexibility within a protein to scratch the surface of important functional dynamics like loop motions. However, this amount of simulation time is likely not enough to collect good statistics on these processes[29]; never mind provide insight on functionally important processes like allostery and folding that happen on μs-s timescales. Nowadays, state-of-the-art systems can produce hundreds of microseconds to milliseconds of data and are routinely used to study collective motions, folding and allostery[30,31]. While state-of-the-art systems have taken advantage of new technology like distributed computing[32], Graphics processing units (GPUs)[33], and specialized MD hardware[34], simulation timescales have also increased dramatically for the average user since increases in the length of simulations is intrinsically linked to computing power, which doubles in performance per dollar every ~18-24 months[35]. Many people are

familiar with this trend from Moore's law[36], the observation that the number of transistors in an integrated circuit doubles every ~18 months. However, this trend of exponential growth is a more general feature of computing systems and will likely continue after Moore's law. While MD simulation timescales are already sufficient to study functionally relevant processes, exponential growth in computing implies that MD will continue to grow as an important tool in new areas of science.

Despite that computing power has only recently been able to provide MD simulations on or approaching biological timescales, MD has had a long history of making advancements in our understanding of biomolecules. Much of these advances were enabled by methodological advances in MD enhanced sampling algorithms, which generally provide improved access to longer timescales. The general idea of enhanced sampling algorithms is to minimize redundancy to use computational power more efficiently. The dynamics of a system or process can be understood in terms of the underlying free energy landscape. Typically, the system spends most of the time in a free energy well where it is more likely for the system to redundantly sample the well than to cross a free energy barrier. However, processes of interest (e.g. inactive to active, unfolded to folded) usually necessitate crossing free energy barriers. Enhanced sampling algorithms are cleverly designed to reduce redundant sampling and facilitate free energy barrier crossing, while maintaining the ability to extract thermodynamic -- and in some cases kinetic -- properties. Enhanced sampling algorithms such as metadynamics[37], adaptive sampling[38], and replica exchange[39] have been used to "speed up" MD simulations by orders of magnitude.

In order for MD simulations to be valid and informative, they have to model biomolecules accurately. All-atom MD simulations rely on Newtonian dynamics where every atom is modeled as a point with the radius, mass, and charge dependent on atom type. Atoms interact with each other based on interaction potentials that include bonded and non-bonded terms. These energy terms form the model, or force field. The interaction potentials between certain atom types originally came from quantum mechanical calculations and condensed phase experimental data[23]. Fortunately, with access to longer timescales MD data has had more overlap with experimental data. Thus, current force fields have been improved by fitting to experimental data. Over the past decade there has been significant improvement in protein modeling. MD can properly model secondary structure, fold proteins, and reproduce NMR data[40]. Overall, there is a consensus that lipid[41] and protein dynamics are accurately modeled, while other biomolecules, such as carbohydrates and nucleic acids, need improvement.

With improvements in force field accuracy and access to longer timescales, MD has been used to make impactful discoveries. MD simulations have enhanced our understanding of conformational changes in GPCRs[42] and kinases[27]. Work done on these systems could lead to new insights into drug development. MD has also improved our understanding of membrane transport. Experimental structures show different conformations of the selectivity filter in ion channels. MD simulations have helped resolve how this filter is regulated and works to selectively conduct ions[43,44]. Lastly, MD has helped answer fundamental biophysical questions, such as “how do proteins fold?” By folding 12 structurally different proteins, Lindorff-Larsen et al., not only showed the

current force fields accurately predict the free energy minimum structures for a diverse set of proteins, but also discovered principles small proteins follow when folding[30]. As MD continues to develop, it will be applicable to larger and more diverse systems and become even more useful in understanding biological processes as well as disease states.

REFERENCES:

[1] Leuschner C., Hansel W., *Curr Pharm Des.* 2004;10(19):2299-310.
[2] Chen L. et al., *J Pharm Sci.,* 2012 Apr;101(4):1508-17.
[3] B Pennarun et al., *Cell Death and Disease*, 2013 4, e894.
[4] Q liu et al., *Cell Death Discovery*, 2016;2:160008
[5] M. Kohno et al., *Chemistry & Biology,* 2014 Nov;21(11):1522-32.
[6] M. Mahlapuu et al., *Front Cell Infect Microbiol*. 2016; 6:194.
[7] Y. Shai. *Peptide Science.* 2002 Dec; 66(4):236-48.
[8] B. Tsai. *Annu. Rev. Cell Dev. Biol., 2007,* Vol. 23:23-43.
[9] CL Moyer, GR Nemerow. *Curr Opin Virol.* 2011 Jul;1(1):44-49.
[10] Banerjee M, Johnson JE. *Curr Protein Pept Sci.* 2008 Feb;9(1):16-27.
[11] Boogart et al., *J Biol Chem,* 2008 Dec; 283(49): 33854-7.
[12] Cross KJ et al., *Protein Pept Lett.* 2009;16(7):766-78.
[13] IF Tannock and D Rotin, *Cancer Research*, 1989 Aug.;49:4373-84.
[14] H Raghuraman and A. Chattopadhyay. *Biosci Rep.* 2007;27:189-223.
[15] H.W. Huang. *Biochemistry.* 2000; 39(29):8347-8352.
[16] L Yang et al., *Biophys J*, 2001; 81(3):1475-85.
[17] MT Lee et al., *PNAS*, 2013; 110(35):14243-8.
[18] C.H. Chen et al., *BBA Biomembranes*, 2014 Sept.; 1838(9):2243-3349.
[19] D. Sun et al., *Langmuir*, 2015; 31(34):9388-9401.
[20] R. Worch. *Acta Biochim Pol.*, 2014; 61(3):421-6.
[21] J.L. Lorieau et al., *JACS*. 2011; 133:2824-7.
[22] A Panahi and M Feig. *J. Phys. Chem. B.* 2010; 114:1407-16.
[23] R. Dror et al., *Annu. Rev. Biophys.* 2012; 41:429-52.
[24] Cowan-Jacob S. et al., *Structure*. 2005; 13:861-871.
[25] W. Xu et al., *Mol. Cell.* 1999; 3:629-638.
[26] Y Xiao et al., *Acc Chem Res*. 2015 Apr; 48(4):1106-14.
[27] D Shukla et al., *Nat. Comm.* 2013; 5:3397.
[28] K Henzler-Wildman & D Kern. *Nature.* 2007; 450:964-72.
[29] G Bowman, *J. Comp. Chem.* 2016; 37(6):558-66.
[30] K Lindorff-Larsen et al., *Science*. 2011; 334(6055):517-20.

[31] J Shi et al., *Sci. Rep.* 2017; 7:44116.
[32] M Shirts and VS Pande. *Science.* 2000; 290(5498):1903-4.
[33] J Anderson et al., *J Comp. Phy.* 2008; 227(10):5342-59.
[34] D.E. Shaw et al., *D.E. Comm. of the ACM.* 2008 July; 51(7):91-7.
[35] R. Kurzweil. *The Singularity is Near*. London: Duckworth, 2005. Print.
[36] GE Moore. *Electronics.* 1965; 114-117.
[37] A. Laio and M. Parrinello. *PNAS*. 2002; 12562-12566.
[38] MI Zimmeman and GR Bowman. *JCTC.* 2015; 11(12):5747-57.
[39] Y. Sugita and Y. Okamoto*, Chem. Phys. Lett.* 1999 Nov 7;314:141-51.
[40] KA Beauchamp et al., *JCTC.* 2012; 8:1409-14.
[41] D Poger et al., *BBA*. 2016; 1585(7 Pt B):1556-65.
[42] DM Rosenbaum et al., *Nature*. 2011; 469:236-40.
[43] SY Noskov et al., *Nature.* 2004; 431:830-34.
[44] MP Harrigan et al., *Sci. Rep.* 2017; 7(1):632.
[45] G Wiedman et al. *BBA Biomemb.* 2015; 1848(4):951-7.
[46] F.M. Hughson. *Curr. Biol.* 1997; 7(9):R565-9.

Chapter 2:

Behavior of the NωV γ-peptide in a Membrane Environment

INTRODUCTION:

Entering a host cell is a key step in the infection cycle of all viruses. However, the strategies employed to penetrate a cell are distinct between enveloped and nonenveloped viruses[1,2] (**Fig. 1**). Structurally, nonenveloped viruses are minimalists, in the simplest cases sporting nothing but a protein shell (capsid) to coat their genome. Enveloped viruses also have a capsid, which is contained within a lipid bilayer, and the bilayer is decorated with surface proteins, which are critical for cell recognition and entry. Much work has been done to understand host cell penetration by enveloped viruses and a clear picture of this mechanism is emerging[3]. The mechanism is intuitive; enveloped viruses fuse their own membrane with the host cell membrane using the help of surface bound fusion proteins[4]. Once the membranes have fused, the capsid can be transported into the interior of the host cell. Without a membrane, nonenveloped viruses lack the ability to undergo such a harmonious fusion. Rather, it's thought that nonenveloped viruses take a more aggressive approach, releasing peptides and/or other lytic factors to breach the host cell membrane after being endocytosed so the capsid can access the interior of the host cell[1]. However, this process has not been well characterized and our understanding of it can be improved by studying the interaction between membrane active viral peptides and lipid bilayer membranes.

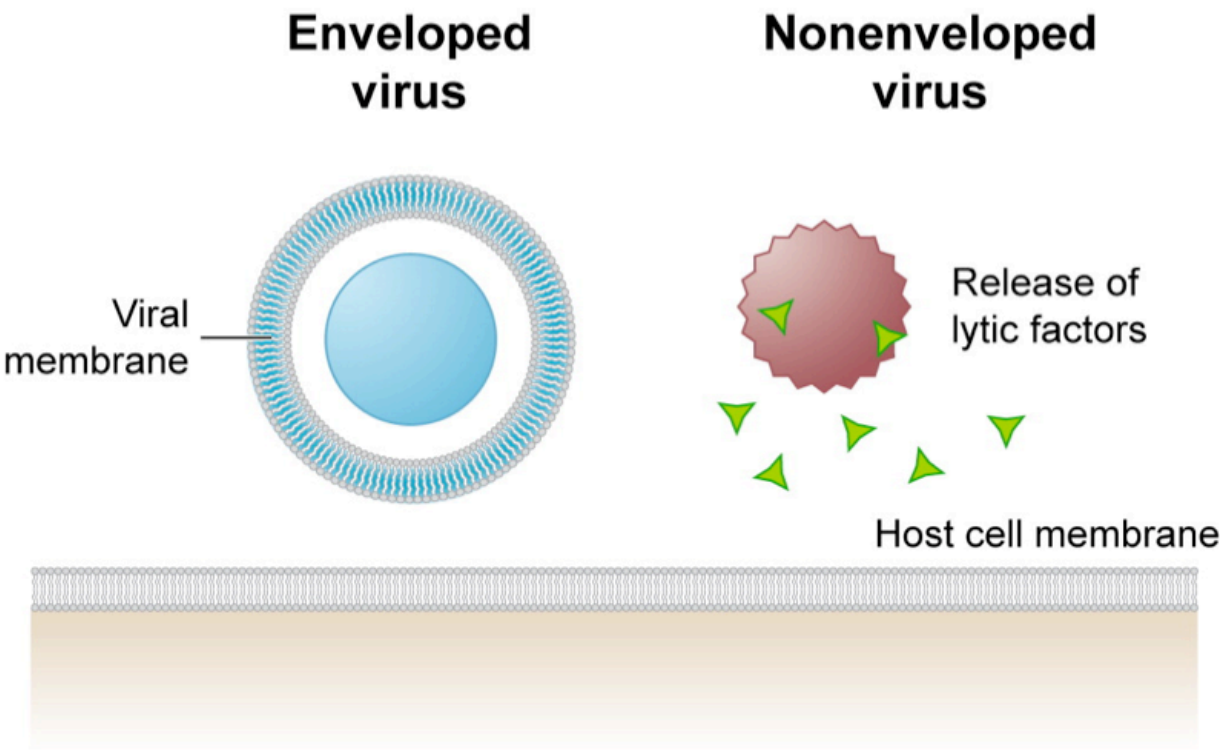

Figure 1. Enveloped and nonenveloped viruses employ different cell entry mechanisms. (Adapted from Billy Tsai[1])

Nudaurelia capensis ω Virus (NωV) is a relatively small and simple nonenveloped virus making it ripe for study. NωV infects butterflies and moths and is a member of the Tetraviridae family[5]. The virus consists of a capsid and a bipartite, linear ssRNA(+) genome contained within it. The capsid is made up of 240 chemically identical protein subunits (644 residues each) that assemble with T=4 icosahedral symmetry[6]. Though chemically identical, the coat protein subunits can be split into 4 quasi-equivalent subunits (A, B, C, D) that have functional differences due to their differing capsid positions. Initial assembly generates an immature particle called the procapsid, which is not infective. When the procapsid is exposed to a low pH (pH=5) environment it undergoes an irreversible maturation to develop infectivity[7] (**Fig. 2A**).

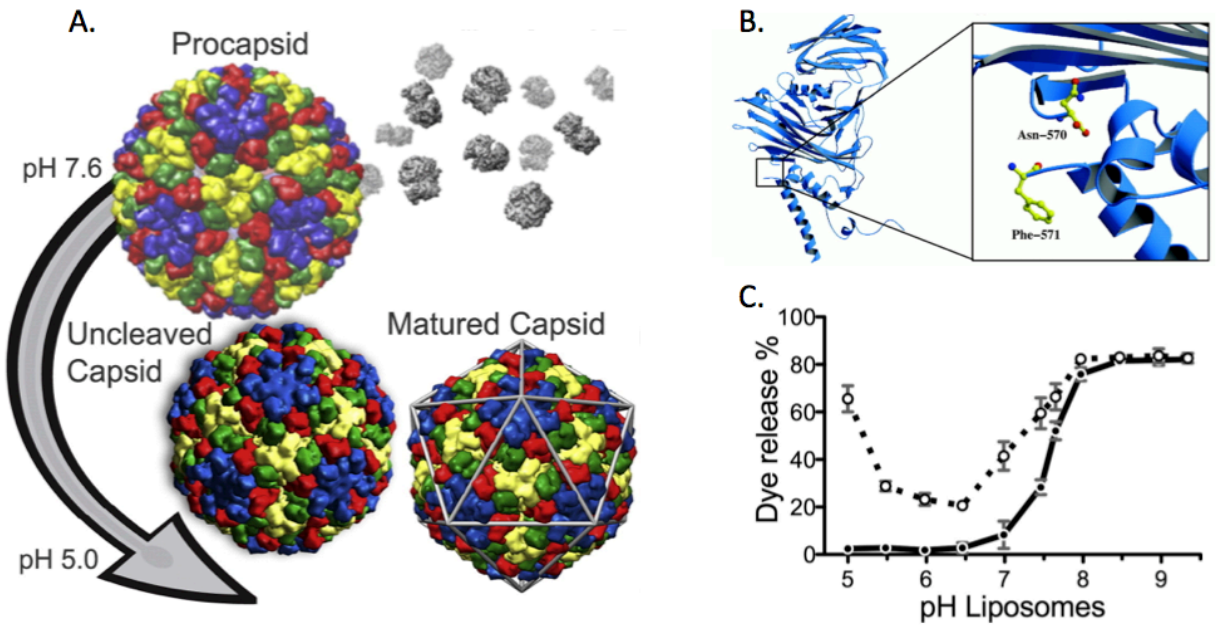

Figure 2. The life cycle of NωV utilizes a wide range of pH and involves an important autocatalytic cleavage of γ-peptide. A) Subunits A, B, C, D shown in blue, red, green, and yellow, respectively. B) Cleavage site of γ-peptide shown in the context of the rest of the A-subunit. C) Dye release from an artificial DOPC liposome caused by mature, full Virus-Like Particle (VLP) (solid line) vs. γ-peptide alone (dotted line). (Adapted from Domitrovic, et al.)[8,9,10]

Maturation is characterized by shrinking of the procapsid (490nm diameter) to a compact capsid (410nm diameter), which happens in conjunction with autocatalytic cleavage of the C-terminal 74 residues of every subunit (**Fig. 2B**). The cleaved peptide is referred to as γ-peptide. Cleavage is necessary to stabilize the mature form of the capsid[8]. To achieve this stability, the C-terminal helix of γ-peptide from C and D subunits acts as a molecular switch inserting itself into the contacts between capsid subunits. It is believed that after maturation, NωV is ejected into the alkaline hindgut of insects where it infects new host cells. While γ-peptides from the C and D subunits are involved in stabilizing the capsid, it has been shown that γ-peptide from the A subunits are released from the 5-fold axis of the capsid in alkaline conditions and are responsible

for lytic activity against the host cell membrane[9]. Interestingly, γ-peptide is strongly activated by alkaline pH and only minimally active in the pH range of the endosome. This is in contrast to most known viral, pH-sensitive lytic peptides that are activated by acidic pH and are thought to help the virus escape the endosome to access the interior of the host cell. Since externalization and activation of γ-peptide occurs only in alkaline conditions (**Fig. 2C**), it is thought that NωV directly penetrates the plasma membrane, rather than taking the endocytotic pathway, making it unique among known nonenveloped viruses.

There have been attempts to explain how the interaction between γ-peptide and host membrane facilitates infection, but as more details have been learned the mechanism has remained unclear[11,12]. Specifically, the first crystal structure revealed a 5-peptide helical bundle of the first (N-terminal) 29 residues (γN) of the lytic γ-peptides inside the capsid (**Fig. 3A**). Based on this, it was proposed that this bundle could be released from the capsid to form an RNA translocating pore in the host membrane (**Fig. 3B**). For this to be plausible, the helical bundle would need a hydrophobic exterior and polar interior to form favorable interactions in the membrane. However, inside the capsid the helical bundle features the opposite: a hydrophobic interior and a polar exterior. Additionally, a refined crystal structure published later reveals that the bundle is intercalated with 5 additional helices from neighboring subunits[12]. Release of this huge bundle at once is implausible and it is more likely that γ-peptides are released from the capsid individually. Given that the peptides do have an amphipathic character, pore formation is still plausible in a scenario where the peptides individually exit from the

capsid and then reorganize in the membrane to form a pore. Obviously, our understanding of this process is unclear. Thus, a study exploring the interactions between γN and a membrane could be valuable to gain new insights into the membrane poration process.

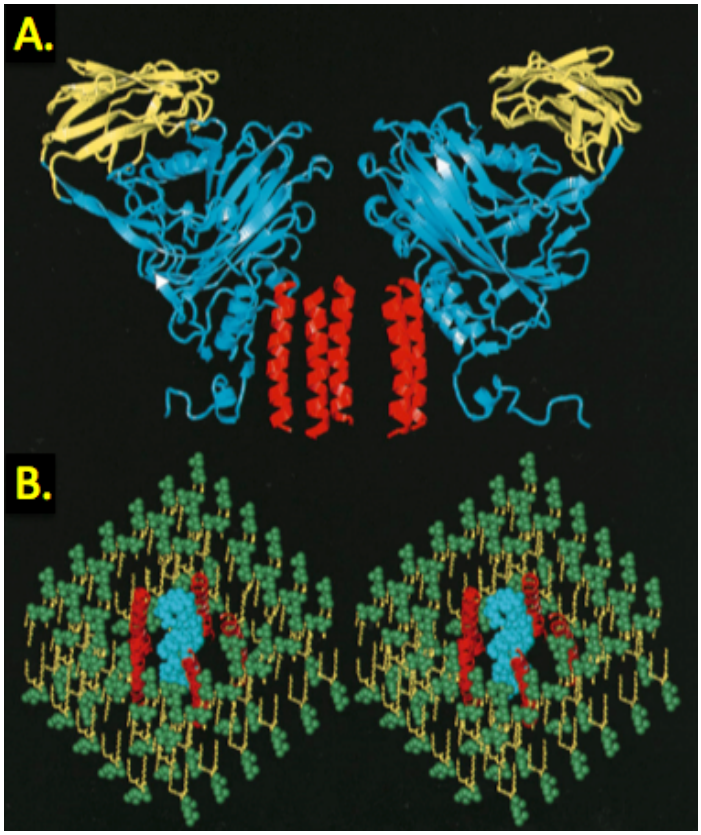

Figure 3. Depiction of a γ-peptide bundle forming a pore to translocate RNA across the host cell membrane. A) View of crystal structure showing γ-peptide bundle (red) located inside the capsid, contained by A subunits (blue and yellow). B) Artist depiction of a pore formed in the host cell membrane (green and yellow) allowing translocation of RNA (cyan). (Adapted from Johnson, et al.)[8]

The current study focuses on understanding how γN interacts with a membrane and how pH affects this interaction. Approaching these questions experimentally is challenging. Specifically, expressing lytic proteins is problematic since they are cytotoxic. Additionally, high-resolution structural techniques are likely to fail because of the membrane environment and inherently dynamic nature of membrane lysis. High-resolution structural techniques like X-ray crystallography require crystals, which can't be formed without adding a detergent to isolate the proteins from the membrane[13]. Additionally, peptides are known to be dynamic in stable pores[14]. The resolution and reliability of experimental structures decreases with increased dynamics. However,

these challenges do not apply to molecular dynamics simulations (MD). Using MD one can obtain atomistic resolution of peptide-membrane interactions under different pH conditions. Thus, MD simulations were performed to probe the pH-dependent mechanism of membrane lysis by γ-peptide.

METHODS:

We performed MD simulations of γN in a membrane environment using implicit and explicit representations of membrane and solvent. Coordinates for γN were taken from the A subunit of the crystal structure (PDB: 1OHF). The amino acid sequence is: FAAAVLAFAANMLTSVLKSEATTSVIKEL.

Implicit Membrane Simulation

To characterize the orientation and structural preference of γN in a membrane, we performed Temperature-based Replica Exchange Molecular Dynamics Simulations (REMD)[15] in conjunction with Constant pH Molecular Dynamics (CpHMD)[16] while utilizing the GBSW (generalized Born with a simple switching function) implicit membrane and solvent model[17,18]. Briefly, REMD is an enhanced sampling scheme, which simulates replicas (identical physical copies) of the system, but each replica is simulated at a different temperature (other variables can in principle be varied). At regular intervals (roughly ever 100 steps), neighboring windows attempt a swap, where the system coordinates would be exchanged. The swaps are accepted with a given probability based on the temperature and potential energy difference, based on the Metropolis criterion[19], which ensures a Boltzmann distribution is maintained in each

window. This scheme allows the system to cross enthalpic barriers more rapidly than an equilibrium simulation, yet preserves a proper Boltzmann distribution for each temperature. Our REMD simulations had 12 windows (or replicas) spanning temperatures from 300 K to 450 K and were run for 100 ns each for an aggregate sampling time of 1.2 μs. The starting conformation of γN was the same for each window: a tilted, transmembrane, helix (**Fig. 4**). The simulation was run at pH=8, the pH required for γ-peptide to have optimal activity[9]. pH was modeled using CpHMD where LYS (residues 18 and 27) and GLU (residues 20 and 28) were allowed to titrate based upon the λ-dynamics (extended Hamiltonian) approach. In this approach, the protonation dynamics are coupled to the conformational dynamics, which allows the probability of a given protonation state to be related to protein structure and local chemical environment. Lastly, the implicit solvent is modeled by having a dielectric constant of 80 where there would otherwise be water molecules modeled. The implicit membrane is a hydrophobic slab with a dielectric constant of 2 to model the hydrophobic core and a smoothing function implemented between the hydrophobic core and implicit solvent to model lipid headgroups. In our model the hydrophobic slab spans 30 Å (z=-15 Å to z=15 Å) and the smoothing function is 5 Å thick (z=12.5 Å to 17.5 Å and z=-12.5 Å to -17.5 Å). A concentration of 0.1 M NaCl was modeled in the implicit solvent through a Debye-Hückel approximation. A flat wall potential with a force constant of 100kcal/mol* $Å^2$ was used to prevent γ-peptide from drifting away from the membrane. All simulations were performed using the CHARMM27 force field[20] employing a 2 fs timestep.

Clustering was performed in MSMBuilder3[21] using the KCenters algorithm[22,23] to generate 8 clusters. pK_a calculations were done by solving the Henderson-Hasselbalch equation with relative concentrations of protonated to deprotonated residue being equivalent to relative number of frames in the simulation where a given titratable residue is protonated vs. deprotonated.

Explicit Membrane Simulations

To characterize differences in membrane binding between pH=6 and pH=8 we ran simulations with γN in solution, starting from less than 1 nm from the surface of a lipid bilayer membrane consisting of 75 1,2-Dioleoyl-sn-glycero-3-phosphocholine (DOPC) lipids on each leaflet. Solvent was modeled using the 3-point TIP3P water model and protein and lipids modeled using the CHARMM36 force field[24]. We modeled the different pH conditions by deprotonating Lys18 for pH=8 and leaving it protonated at pH=6. The simulations were setup using CHARMM-GUI[25] and were run using Gromacs 5[26]. Each simulation was pre-equilibrated following the standard CHARMM-GUI protocol, which includes 5000 energy minimization steps using the steepest descent algorithm followed by 275ps of restrained equilibrium simulations where the restraints were progressively lowered to relax the system. The simulations were run 250ns for each pH=6 and pH=8.

Umbrella Sampling

We combined Steered Molecular Dynamics (SMD)[27] with Umbrella Sampling (US)[28] to construct a free energy profile for insertion of γN into a membrane. SMD was performed by applying a harmonic biasing force on the center of mass of the first 3

residues pulling (along z-axis, bilayer normal) from a starting membrane-bound configuration to the lower leaflet of the bilayer. The force constant used was 30kJ/mol•Å^2 and the pulling rate was 1 Å/ns. Umbrella Sampling was performed by having 35 windows separated ~1 Å apart along the z-axis, perpendicular to the membrane. Each window included a harmonic restraint with a force constant of 30kJ/mol•Å^2 to hold the center of mass of the first 3 residues at its specified Z-coordinate value. In practice, this was all performed as one continuous simulation where the system alternated between SMD (pulling) for 1 ns followed by US for 10 ns. The total aggregate simulation time was ~500 ns, yielding equilibration times of 10-20 ns for each umbrella window. In order to construct a free energy profile, the reaction coordinate data was unbiased using the Weighted Histogram Analysis Method (WHAM)[29] software from Alan Grossfield (http://membrane.urmc.rochester.edu/content/wham) to obtain the potential of mean force (PMF). Both Lys residues were deprotonated to mimic the expected protonation state of Lys at pH=8 in a membrane[30]. The simulations were run with Gromacs 5 using the CHARMM36 force field.

RESULTS AND DISCUSSION:

Structure in an Implicit Membrane

Using a combination of REMD, CpHMD, and an implicit membrane we ran simulations starting from a transmembrane, helical conformation of γN to characterize its structural properties in a membrane environment. Clustering based on RMSD

revealed that the most highly populated conformation was a membrane-bound conformation with a kink at Lys18 (**Fig. 4**). 53.2% of conformations were within 3 Å RMSD of the center of this cluster, indicating that this conformation strongly dominates when γN is in a membrane environment at pH=8. This low energy conformation can be understood by looking at how polar and nonpolar residues are partitioned at the membrane-water interface. The N-terminal half is buried in the hydrophobic core of the membrane since it is dominated by nonpolar residues. The kink at Lys18 allows flexibility so that the polar residue dominated C-terminal region can avoid the hydrophobic core and interact more preferably with lipid head groups and solvent. Thus, it is likely that this kink is important in increasing membrane-binding affinity. Interestingly, the kink is at Lys18, which has its pK_a downshifted by ~1 pH unit (pK_a =9.65, calculation described in methods) relative to its solution pK_a of 10.54. It is known that γ-peptide goes from inactive to active as the pH changes from 6 to 8[9]. The obvious explanation for activation might be the change in protonation state of a histidine residue. However, γ-peptide does not contain any histidine residues. Given the pK_a shift and the kink at Lys18, we believe that deprotonation of Lys18 promotes the kink, which increases membrane binding affinity and thus, membrane disruption.

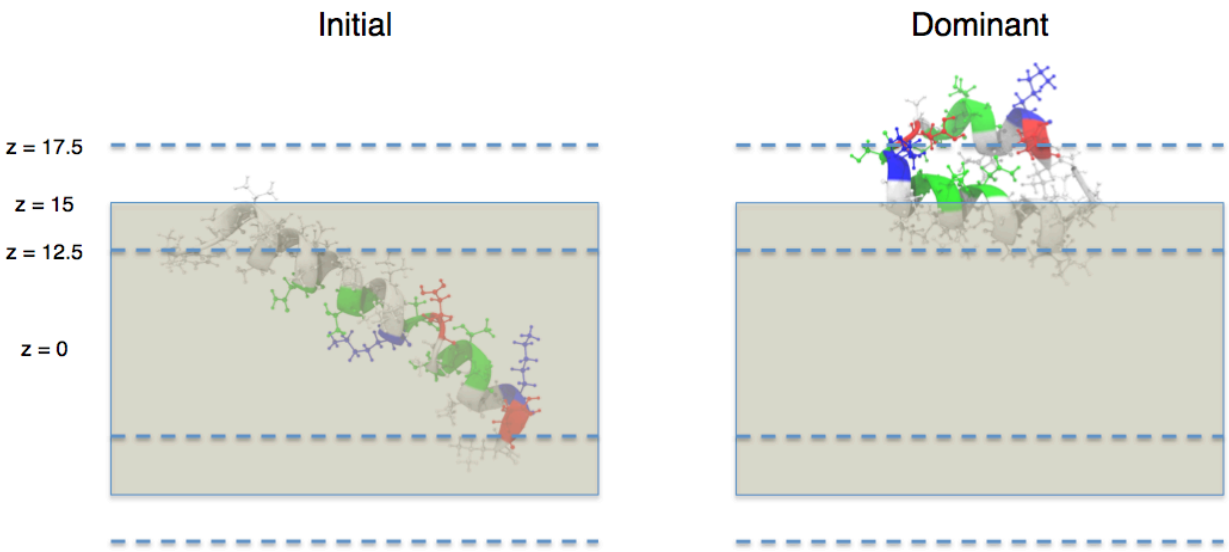

Figure 4. γN prefers a kinked membrane bound conformation. Simulation began from a helical transmembrane conformation (*left*). Structure taken from highest populated cluster (*right*). Membrane depth is in Å. Dotted blue lines indicate start and end of dielectric constant smoothing region. Tan indicates the low dielectric constant, membrane core. Polar, negatively charged, positively charged, and nonpolar are colored green, red, blue, and white, respectively.

Explicit Membrane Binding

We performed simulations to determine γN's ability to bind a DOPC membrane and determine how pH changes might alter the binding process. Here, we started simulations with γN, at pH=6 or pH=8, in solution above a membrane. Based on the pK_a data from the previous simulation, we modeled pH=6 and pH=8 by protonating and deprotonating Lys18, respectively. Under both pH conditions, γN bound to the membrane within 50 ns. Thus, the activation of γ-peptide cannot be explained by simply enhancing the membrane binding at pH=8. However, the final membrane bound state is starkly different when comparing pH=6 to pH=8. At pH=6, γN binds as a rigid helix only halfway inserted into the membrane (**Fig. 5**). Interestingly, at pH=8 (deprotonated Lys18), γN binds with a similar kink as identified in the previous implicit membrane

simulation run at pH=8 (titration performed on the fly) (**Fig. 5**). Thus, raising the pH from pH=6 to pH=8 generates a kink that is caused by deprotonating Lys18. With this kink, γN is more compact and all of its residues are interacting with the membrane hydrophobic core or the polar lipid headgroups. Thus, the kink allows higher affinity binding to the membrane. In the context of NωV infection, higher binding affinity means more γ-peptides will be membrane-bound than in solution, which would cause more membrane lysis.

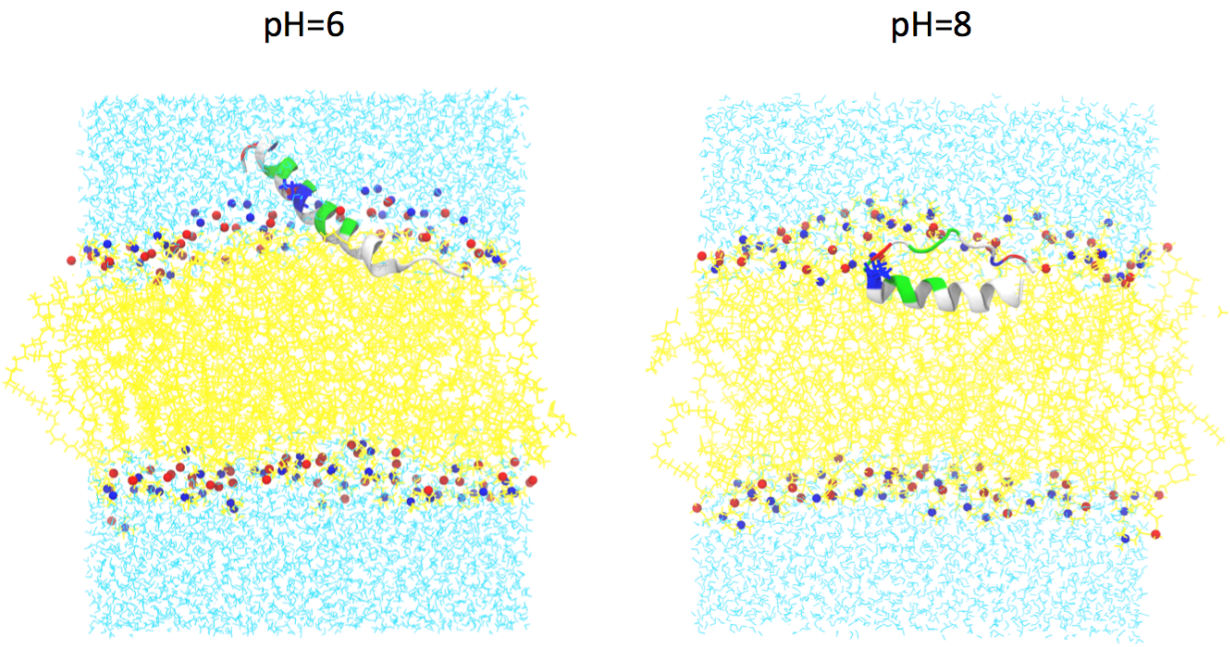

Figure 5. Binding differences between pH=6 and pH=8. Snapshots from the end of pH=6 (*left*) and pH=8 (*right*) simulations. Protein coloring scheme same as previous figure. Membrane core, phosphates, cholines, and water are colored yellow, red, blue, and cyan, respectively.

From a previous experiment performed by Domitrovic et al.,[9] we see that membrane lysis is at a minimum between pH=5.5 and pH=6.5 followed by a linear increase in lytic activity through pH=8 (**Fig. 2C**). Thus, it is plausible that as the pH is

raised from pH=6 to pH=8 Lys18 becomes deprotonated on more γ-peptides, generating the kink, which increases membrane binding affinity and consequently, membrane lysis.

Energetics of Membrane Insertion

To characterize the membrane disruption process, we ran Umbrella Sampling simulations to construct a free energy profile for insertion of γN at pH=8, starting from a membrane bound helical conformation. In order to insert and cross the membrane, γN forces a lipid flip-flop, dragging lipids with it from one leaflet to the other (**Fig. 6A,C**). This mechanism appears to be similar to Melittin pore formation as discussed in chapter 1. Throughout the insertion process, a lipid head group tracks closely with the positively charged N-terminus of γN. The negatively charged phosphate head group of this lipid shields the positively charged N-terminus from interacting only with the hydrophobic core of the membrane. Still, there is a cost to flip-flopping lipids and moving charges through a membrane. This process has a free energy cost of ~60kcal/mol (**Fig. 6B**), making it an extremely rare event that would not happen under any normal conditions. Thus, γN as a single peptide is unlikely to insert into host cell membranes. Membrane insertion and disruption may be achieved either by several γN peptides working in conjunction, or possibly the full-length γ-peptide can insert on its own. While these questions are interesting, they introduce many additional degrees of freedom that make calculations using all-atom MD intractable given reasonable computational resources.

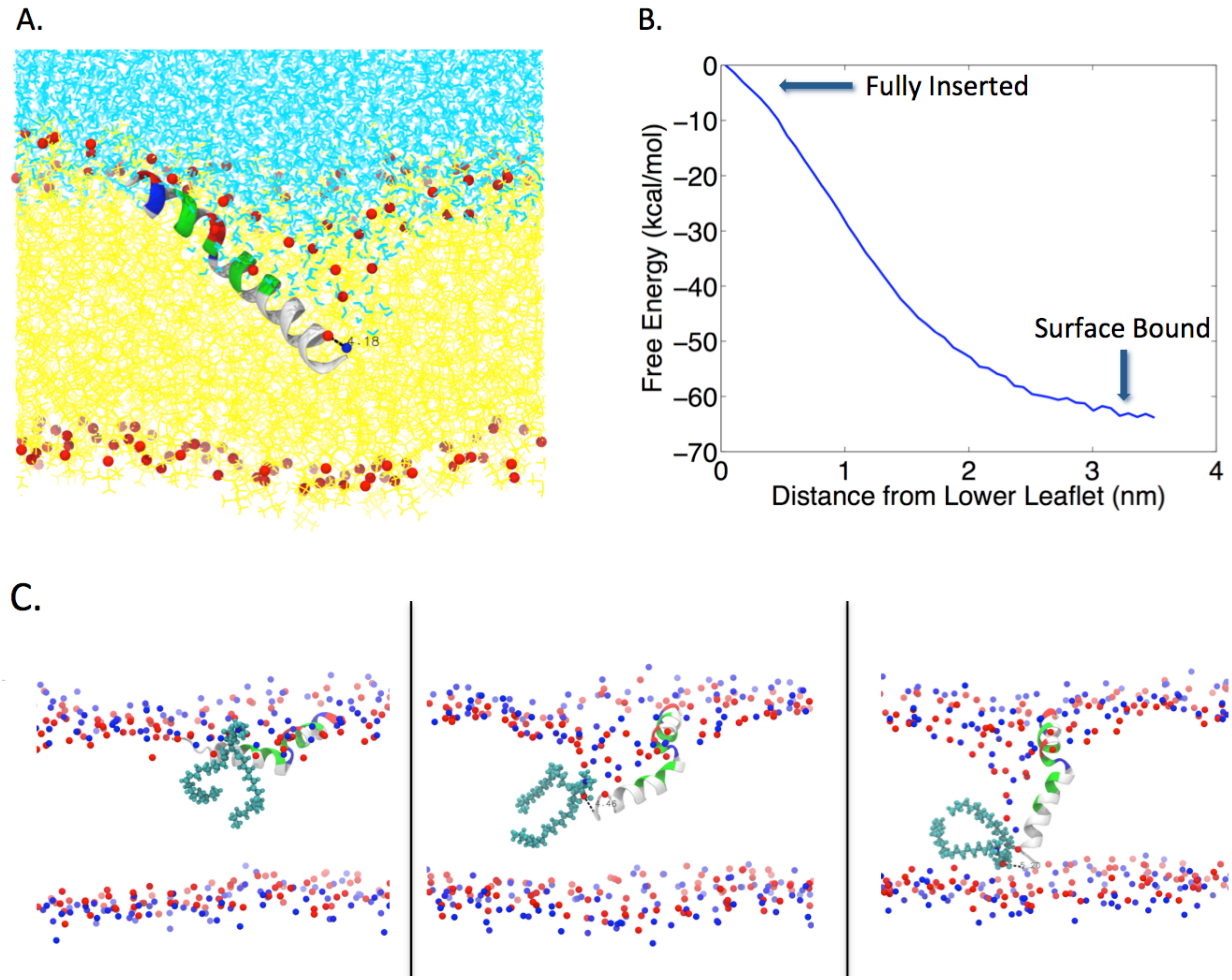

Figure 6. Membrane Insertion of γN. A) Screenshot halfway through the insertion process. Distance between (+)-charged N-terminus (blue ball) and a lipid head group is shown by the black dotted line. B) PMF taken from Umbrella Sampling. C) Snapshots along the insertion process taken to highlight lipid flip-flop. A single lipid (cyan) is shown to track with γN across the membrane.

CONCLUSIONS:

In this work, we used MD simulations to better understand how γ-peptide achieves pH-dependent membrane lysis. We modeled the α-helical N-terminal segment of γ-peptide, referred to as γN, as this segment had been previously proposed to form pores in host cell membranes[9]. Simulation data reveals that γN has a different

membrane binding mechanism at pH=6 and pH=8. At pH=6, γN binds as an α-helix halfway inserted into the membrane and halfway in solution. At pH=8, residue Lys18 is deprotonated, which destabilizes the α-helix, generating a kinked conformation. This kinked conformation allows tighter membrane binding than at pH=6, which may explain the higher membrane lytic activity at pH=8. How this lytic activity is achieved is still not well understood. Our insertion simulation suggests that γN can cause a lipid flip-flop while crossing the membrane. However, the free energy cost of this transition is unreasonably high (~60kcal/mol) and thus, unlikely to happen. Previous studies have shown that for pore-forming peptides, multiple peptides inserting together lowers the free energy cost of insertion[31]. Thus, understanding how γ-peptide disrupts the host cell membrane would likely require simulations of multiple peptides in a membrane environment. As stated above, studying this problem using all-atom models is difficult without extraordinary computational resources. Thus, there is incentive to use coarse-grained models, which allow simulation of greater system sizes at a reduced computational cost by sacrificing atomistic resolution, if they are at a resolution high enough to properly model the system of interest.

REFERENCES:

[1] B. Tsai. *Annu. Rev. Cell Dev. Biol., 2007,* Vol. 23:23-43.
[2] Banerjee M, Johnson JE. *Curr Protein Pept Sci.* 2008 Feb;9(1):16-27.
[3] Cosset FL, Lavillette D., *Advances in Genetics*. 2011;73:121–183.
[4] S.C. Harrison. *Nat. Struct. Mol. Biol.*, 2008, 15:690‑698.
[5] D.K. Agrawal, J.E. Johnson. *Virology.* 1992, 2:806-14.
[6] D.K. Agrawal, J.E. Johnson. *Virology.* 1995, 207(1):89-97.
[7] Canady MA, Tsuruta H, Johnson JE. *J Mol Biol.* 2001, 311(4):803-14
[8] Taylor DJ, Krishna NK, Canady MA, Schneemann A, Johnson JE. *J Virol.* 2002, 76(19):9972-80.
[9] Domitrovic T, Matsui T, Johnson JE. *J Virol*. 2012, 86(18):9976-82.
[10] T. Domitrovic et al., *J. Mol. Biol.,* 2013, 425 (9):1488-96.
[11] Munshi S, Liljas L, Cavarelli J, Bomu W, McKinney B, Reddy V, Johnson JE. *J Mol Biol*. 1996, 261(1):1-10.
[12] Lee KK, Tang J, Taylor D, Bothner B, Johnson JE. *J Virol*. 2004, 78(13):7208-16.
[13] I Moraes et al., *Biochim. Biophys. Acta.* 2014; 1838(1):78-87.
[14] J.M. Leveritt et al., *Biophys. J.*, 2015; 108(10):2424-6.
[15] Y. Sugita and Y. Okamoto*, Chem. Phys. Lett.* 1999 Nov 7;314:141-51.
[16] Khandogin, J. & Brooks, C. L., III. *Biophys J.* 2005; 89:141–157.
[17] W. Im, M.S. Lee, and C.L. Brooks III. *J. Comp. Chem.* 2003; 24:1691-1702.
[18] W. Im, M. Feig, and C.L. Brooks III. *Biophys J.* 2003; 85:2900-18.
[19] W.K. Hastings. *Biometrika*. 1970; 57(1):97-109.
[20] A. MacKerell et al., *J. Phys. Chem. B.* 1998; 102: 3586-3616.
[21] K.A. Beauchamp et al., *JCTC.* 2011; 7(10):3412-19.
[22] Gonzalez, Teofilo F. *Theor. Comput. Sci*. 38 (1985): 293-306.
[23] Beauchamp, Kyle A., et al. *J. Chem. Theory. Comput*. 7.10 (2011): 3412-3419.
[24] R.B. Best et al., *JCTC.* 2012; 8:3257-73.
[25] S. Jo et al., *J. Comp. Chem.*, 2008 Mar 19; 29 (11): 1859-65.
[26] MJ Abraham et al., *SoftwareX*, 2015 Sep; 1-2:19-25.
[27] S. Izrailev et al., *Biophys J.*, 1997; 72:1568-81.
[28] R.K. Schmidt et al., *J. Phys. Chem.,* 1995; 99(29):11339-43.
[29] B. Roux, *Comp Phys Comm* 1995, *91*, 275–282.
[30] A. Panahi and C.L. Brooks, *J. Phys. Chem.,* 2015, 119:4601-7.
[31] D. Sun et al., *Langmuir*, 2015; 31(34):9388-9401.

Chapter 3:

Evaluation of the Hybrid Resolution PACE Model for the Study of Folding, Insertion, and Pore Formation of Membrane Associated Peptides

ABSTRACT:

The PACE force field presents an attractive model for conducting molecular dynamics simulations of membrane-protein systems. PACE is a hybrid model, in which lipids and solvents are coarse-grained consistent with the MARTINI mapping, while proteins are described by a united-atom model. However, given PACE is linked to MARTINI, which is widely used to study membranes, the behavior of proteins interacting with membranes has only been limitedly examined in PACE. In this study PACE is employed to examine the behavior of several peptides in membrane environments, namely WALP peptides, melittin and influenza hemagglutinin fusion peptide (HAfp). Overall, we find PACE provides an improvement over MARTINI for modeling helical peptides, based upon the membrane insertion energetics for WALP16 and more realistic melittin pore dynamics. Our studies on HAfp, which forms a helical hairpin structure, do not show the hairpin structure to be stable, which may point toward a deficiency in the model.

INTRODUCTION:

Molecular dynamics (MD) simulations of membrane protein systems can be informative of many biological processes, including ion transport,[1] cell signaling,[2] and protein translocation,[3] among others.[4] Recent studies have utilized MD to explore and

validate complete models of enveloped virus particles including HIV-1[5] and Influenza A.[6]. While MD can be harnessed to perform rigorous kinetic,[7] thermodynamic[8] or mechanical[9] analyses on biological systems, the method can also be a great tool for understanding the qualitative features of a system at atomic resolution; a "computational microscope",[10] as it has been described.

However, a limitation of MD is the typical timescales of all-atom simulations are in the nanoseconds (ns) to microseconds (ms) range. These timescale limits are being pushed by advances in modern high performance architectures, GPU computing[11] and specialized hardware such as the ANTON/ANTON 2 machine.[12,13] All-atom simulations are often augmented by enhanced sampling methods to accelerate protein dynamics (recently reviewed in[14,15]) and improve convergence of computed properties. However, these methods may significantly increase the computational cost (e.g. replica-exchange, umbrella sampling), may obscure the gathering of kinetic/mechanistic information (e.g. replica-exchange) or may involve non-Boltzmann sampling that can complicate (or prevent) the calculation of thermodynamic quantities (e.g. metadynamics, accelerated MD, temperature-accelerated MD, adaptive-biasing force). While all-atom force-fields provide the most accurate classical description of biomolecules, coarse-grained (CG) models have been successfully developed to capture the essential physics, while omitting some atomic details that may not be critical to longer length scale phenomena. The utilization of CG models provides a means to probe longer timescale and larger lengthscale biomolecular phenomena at a reduced computational cost. Of course the

ability to access longer time scales is of little benefit if the models are not capable of providing a reasonable physical representation of the system.

Lipid systems has been an area where CG models have been particularly fruitful. Lipids are smaller than proteins and nucleic acids and do not fold into complex three-dimensional structures. The important features of lipids are their flexible hydrophobic tails attached to polar and charged head groups; properties which drive self-assembly into a variety of aggregated states, such as bilayers, micelles, and inverted hexagonal phases. These physical characteristics and ability to form aggregated states have been successfully modeled at descriptions less detailed than fully atomistic. The most detailed CG models are the united atom (UA) variety, in which only non-polar hydrogens are not explicitly represented. Popular UA models include those by Berger[16] and from the GROMOS force-field.[17] More aggressive CG models have been put forth by Klein,[18] Marrink (MARTINI)[19] and Voth[20] in which multiple heavy (non-hydrogen) atoms (~4) are mapped into a single CG interaction site. These models attempt to capture the net chemical properties of the atoms, which are represented by the CG beads. Whereas, even more aggressive CG models, referred to as mesoscopic models, focus on the bulk physical properties of the lipids. These models represent an individual lipid molecule by just a few CG particles, a notable model of this class is the three-bead model by Deserno.[21] In the mesoscopic models water is represented as a continuum solvent, which is also what is done in some CG models that use a lipid heavy atom 4-to-1 mapping.

Another approach to coarse-graining membranes is to completely remove the explicit representation of the bilayer and instead represent it implicitly as a continuum medium with a low dielectric constant. The implicit membrane representation is commonly predicated upon calculating the solvation free energy based upon a generalized Born formalism. The implicit membrane models may have a simple smoothing function[22] between the hydrophobic core and solvent dielectrics, or may have a more complex composition, with multiple slabs of different dielectric constants representing the different chemical environments of a bilayer, as function of bilayer depth.[2] An advantage of implicit membrane models is they are designed to be integrated with all-atom protein models. However, the lack of specific lipid-protein (and protein-solvent) interactions can be a downfall, as well as the inability to model heterogeneous membranes or membrane defects such as toroidal pores.

While these different CG models have their various strengths and weaknesses, the MARTINI model is one of the most widely used. The MARTINI model was initially developed for modeling lipids and surfactants,[24] and now offers hundreds of lipid types. The popularity of the model was established through studies which showed good agreement with experimental results in the study of lipid phase behavior[25,26] and membrane mechanical properties.[27] The force field has been extended to include proteins and DNA and has been utilized in the study of several membrane proteins including cytochrome proteins[28,29] and GPCRs.[30] However a limitation in the MARTINI protein model is that secondary structure remains fixed through a simulation, so folding studies are not directly accessible.

Given that the MARTINI model provides a fast and reasonably accurate model for a wide range of lipids, it is desirable to couple this model with a higher resolution protein model that is suitable to study folding and secondary structure changes. Hybrid resolution models have recently been developed to achieve this goal, pairing MARTINI with the GROMOS all-atom force field[31,32] or the PACE UA force field.[33,34]

PACE was originally parameterized to be compatible with MARTINI water,[35] but was then extended to include MARTINI phosphatidylcholine (PC) lipids.[33] In the study that extended PACE to included PC lipids the tilting of WALP peptides in a transmembrane (TM) state was shown to be in reasonable agreement with experimental measurements. Dimerization of glycophorin A was also examined in a membrane environment and the simulated dimer structure was in excellent agreement with the NMR dimer structure. Further parameterization of the model to improve backbone solvation properties and the inclusion of partial charges on acidic and basic residues, allowed for successful folding of several peptides up to 73 residues long in an aqueous environment.[33,34] Subsequent studies of protein folding in aqueous environments include detailed analysis of the TRP-cage and WW-domain, which revealed the model is capable of capturing complex folding pathways involving both on- and off-pathway intermediates.[36] Also, b-amyloid fibril elongation was investigated with the PACE model using replica-exchange and kinetic network analysis, revealing mechanistic insights. [37]

The inclusion of PACE in the CHARMM-GUI[38,39] website has made the simulation of protein-membrane systems widely accessible. Currently, protein-bilayer systems can be constructed with 15 different lipid types with 4 different head groups

(PC, PE, PS, PG).[39] Recent studies of membrane-protein systems using PACE include protonation-dependent conformational changes in lactose permease[40] and gating of the heat activated TRPV1 channel.[41]

While the availability and ease to setup membrane-protein systems through the CHARMM-GUI interface makes PACE an attractive model, there have been limited studies on the conformational dynamics of peptides in membrane environments with the latest version of PACE.[34] A recent study investigated the ability of PACE to capture the environmental sensitive folding behavior of the TMX3 peptide.[42] TMX3 is a 31 residue peptide which is largely disordered in water but becomes helical in membrane environments. In PACE, TMX3 showed similar degrees of folding in water and non-polar cyclohexane, which was explained by the inability of the CG MARTINI water to hydrogen bond with the protein backbone. They reasoned the lack of competition between backbone-backone and water-backbone hydrogen bonds leads to an overstabilization of folded structures in aqueous environments in PACE. In this work we will explore the folding characteristics of the helical WALP16, WALP19 and WALP23 peptides as well as the influenza hemagglutinin fusion peptide (HAfp), which forms a helical hairpin structure on micelles. In addition, we examine the dynamics of a tetramer of melittin peptides in a TM state for their ability to form transient water pores. We also performed simulations of the melittin and HAfp systems in MARTINI to provide a basis for evaluation, while the WALP simulations are largely compared against previous all-atom simulations. The goal of this study is to evaluate the suitability of PACE for studying relatively small, but dynamic peptides in membrane environments.

METHODS:

The PACE force field has been described in detail previously,[33-35] we will only briefly summarize the main features of the model. PACE models solvents and lipids consistent with the MARTINI force field[19], while proteins are represented by a united atom model, where heavy atoms and polar hydrogens are explicitly represented. Cross-resolution terms are optimized against thermodynamic data, and partial charges are included on acidic and basic residues, which interact through Coulomb potentials in the UA representation, but CG and UA interactions are handled by an effective Lennard-Jones (LJ) potential. The Coulomb energies are calculated with a relative dielectric e_r = 15, as is standard in MARTINI simulations with non-polarizable water.

A modified version of NAMD2.9[43] was used for all PACE simulations. Input files were generated through the CHARMM-GUI website,[38] which produces the standard six equilibration steps during which protein and lipid restraints are gradually released, as well as production run inputs. Parameters for the production runs include a 5-fs timestep with the neighbor list being updated every 10 steps. The electrostatic and van der Waals interactions were shifted to zero between 0 and 1.2 nm and 0.9 and 1.2 nm, respectively. The temperature was maintained at 303.15 K using a Langevin thermostat with a damping coefficient of 1/ps. Semi isotropic pressure coupling was applied using the Langevin piston method to maintain the pressure at 1 atm in the normal and lateral directions to the bilayer.

WALP Simulations

Five equilibrium simulations of 10 μs each were performed on WALP16, WALP19 and WALP23 peptides, which have sequences GWW$(LA)_5$WWA, GWW$(LA)_6$LWWA, and GWW$(LA)_8$LWWA, respectively. Each WALP peptide was simulated starting from an extended, TM configuration. Additionally, WALP16 and WALP19 were simulated from an extended structure starting ~2 nm above the membrane. The extended protein structures were generated using CHARMM.[44] The protein-membrane systems were generated with CHARMM-GUI to construct systems containing 250 total POPC lipids and enough water for a distance of at least 7 nm between the bilayer leaflets and their closest periodic image leaflet. Additionally, a 0.10 M NaCl concentration was used. TM structures were inserted into the membrane using the replacement method. Both termini were capped with acetyl (N-term) and N-methyl amide (C-term) groups. WALP16 and WALP19 were also simulated for 5 ms in a pure solvent system, starting from an extended conformation, to evaluate folding in the absence of a bilayer.

Steered Molecular Dynamics (SMD) was used to generate a series of snapshots along a de-insertion pathway, which would be subjected to umbrella sampling. This was done for WALP16, by pulling the center of mass of the peptide out of the membrane at a constant velocity of 1 Å/25 ns in the positive Z-direction. The SMD simulations applied a force constant of 5 kcal/mol/Å^2 which was applied to the backbone atoms. The pull was done at a sufficiently slow rate to prevent any disruptions to the membrane that might not be able to re-equilibrate during umbrella sampling. Additionally, a Z-position restraint with a 0.5 kcal/mol/Å^2 force constant was applied to all phosphate groups of the lower

leaflet of the membrane to prevent translation of the entire membrane in the positive Z-direction.

Umbrella sampling[45] was used to obtain a free energy profile describing the membrane insertion of WALP16. The distance between the center of mass (COM) of WALP16, based upon the a-carbons positions, and the center of mass of the bilayer, based upon all phosphate groups, was harmonically restrained in the Z-direction. 50 configurations were sampled, which spanned 0 to 50 Å COM separations, with approximately 1 Å spacing between windows. In each window the COM separation was maintained by a 5 kcal/mol/Å^2 force constant. Each window was simulated for 1 μs, and the reaction coordinate data was unbiased using the Weighted Histogram Analysis Method (WHAM)[46] software from Alan Grossfield (http://membrane.urmc.rochester.edu/content/wham) to obtain the potential of mean force (PMF).

Melittin Simulations

We simulated a tetramer of the antimicrobial peptide melittin in a tetrameric pore structure to evaluate whether PACE could accurately model the dynamics of a transient pore. The initial tetramer structure was obtained by backmapping[47] a system which was constructed in MARTINI (see below). The system consisted of 136 total DPPC lipids and a 0.10 M NaCl salt concentration. The PACE simulation was run for 5 μs.

We also ran a CG MARTINI[48,49] simulation with the GROMACS 4.6.5 package[50] of a system containing pre-arranged trans-membrane tetramer of melittin embedded in DPPC membrane. The system was generated using the insane.py script.[51] The melittin

tetramer structure was generated by manipulating the crystal structure (PDB ID: 2MLT).[52] Specifically, four melittin peptide molecules were placed in a symmetric arrangement where all four peptides had their respective C-termini embedded in the upper bilayer leaflet. The tetramer was constructed by rotating and translating the melittin molecules such that the hydrophilic residues of each peptides faced each other while the hydrophobic residues faced outward toward the hydrophobic core of the DPPC bilayer, analogous to the protocol of Leveritt *at al.*[53] The bilayer contained 136 lipid molecules, and the system was solvated with 2928 CG MARTINI polarizable water particles and NaCl was added to a concentration of 0.1 M.

After the initial setup, the system was minimized using the steepest descent algorithm, followed by NVT equilibration simulation for 100 ns with position restraints on the peptides, while allowing the lipids, water and ions to relax. Next, all the restraints were removed and NPT simulations were performed for 5 μs using the Berendsen coupling scheme with the temperature maintained at 323 K and pressure kept at 1.0 bar with semi-isotropic coupling. The time constants for the pressure and temperature couplings were 3.0 and 1.0 ps, respectively, and the compressibility was set at 3×10^{-5} bar^{-1}. Electrostatic interactions were calculated using the Particle Mesh Ewald (PME) relative dielectric of $e_r = 2.5$, in accordance with the protocols for MARTINI simulations using polarizable water. The simulation was performed utilizing periodic boundary conditions in X,Y, and Z-directions, with a time step of 0.020 ps (except for NVT simulation where the time step was 0.010 ps). The nonbonded LJ interactions were smoothly shifted to zero between 0.9 and 1.3 nm.

HAfp Simulations

We studied the 23-residue variant of the HAfp N-terminal region, which has sequence GLFGAIAGFIEGGWTGMIDGWYG. The structure of the 23-residue HAfp has been determined by NMR in the presence of DPC micelles, and the structure displays a kinked topology, often referred to as a helical hairpin.[54] We evaluated the ability of PACE to properly fold HAfp in a membrane environment. We performed two 5 μs simulations starting from random extended conformations, one membrane bound and the other in solution ~2 nm above the membrane. The membrane was composed of 250 DLPC molecules in an attempt to match the chain length of lipids in a DPC micelle. Again, these systems were generated in CHARMM-GUI at a NaCl concentration of 0.1 M. Similar simulations were performed, but starting from the folded NMR structure (PDB ID: 2KXA). Six additional simulations starting from the folded, membrane bound state at different insertion depths were performed for 3-5μs each. The termini were charged in all simulations.

We also performed simulations of HAfp in the MARTINI force field using simulation parameters consistent with the MARTINI simulations of melittin. The system was generated from the HAfp NMR structure (PDB ID:2KXA), with charged termini. A bilayer of 250 DLPC lipids was constructed with the insane.py script and the system was solvated within polarizable water and ionized with NaCl at a 0.1 M concentration. Simulations were performed from starting configurations in which the peptide was placed ~2 nm above the membrane or in a membrane bound configuration. To construct the system topology a DSSP[55] structure file was supplied. The DSSP

algorithm classified residues 2-21 as being helical, even though a kink occurs in the central region of the peptide. Simulations run using this native DSSP file caused the kinked helix to form a straight helix within the first ns of the simulations. Therefore, we created a modified DSSP file in which residues 10-12 were changed from type helix (H) to type loop (blank). Using this modified topology the system did not have a strong propensity for forming a straight helix and these are the simulations we analyzed for this study.

Trajectory Analysis

Trajectory analyses were performed using a combination of CHARMM, MDTraj[56] and GROMACS tools. Helicity measurements were taken every 1 ns using the COOR SECS function in CHARMM. Distance calculations between membrane and protein were done using MDTraj. Contacts between residues and the membrane were analyzed in CHARMM, using a cutoff of 5 Å. Tilt angles were calculated from the normalized dot product between the membrane normal and vector along the helical axis, described by the vector connecting the geometric center of the first four and last four residues. Representative snapshots were selected by using a script derived from the MDTraj example for finding a centroid. The algorithm computes all pairwise RMSDs between conformations and calculates a similarity score for each conformation. The similarity score exponentially decreases with increasing distance to favor conformations that are very close to a significant number of other conformations. Melittin simulations in both PACE and MARTINI, were analyzed with the g_density tool from GROMACS, after re-aligning the membrane to the center for each snapshot of the trajectory. The density of

water and at the center of the bilayer was computed using 11 slabs, and the sixth slab was treated as the center. The full density profiles used 100 slabs. HAfp kink angle calculations were performed with the QUICK function in CHARMM. The kink angle was computed between the three centers of mass defined by the Ca atoms in residues 3-6, 12-13 and 18-21. The choice to not use terminal residues was to avoid artifacts due to helix fraying.

RESULTS AND DISCUSSION:

WALP Peptides

WALP peptides are designed peptides,[57] which form a TM α-helix. Several all-atom simulations have been performed on WALP peptides interacting with membranes[58-61] and provide ample data to compare our PACE simulations against.

Folding & Insertion from Solution Phase

We performed equilibrium molecular dynamics simulations on three different length WALP peptides, WALP16, WALP19 and WALP23 to assess the ability of the PACE force field to properly fold the peptides into an α-helix structure and whether the environment (aqueous or membrane embedded) affected the folding. Furthermore, as these peptides are known to insert into membranes to form a TM configuration, we initiated simulations with an unfolded peptide in the solution phase to see if the peptides would spontaneously insert into the membrane phase and to see to what degree the folding and insertion process are coupled.

Simulations of WALP16 and WALP19 were initiated in an unfolded configuration in the solvent phase, approximately 2 nm away from the bilayer. The initial configuration

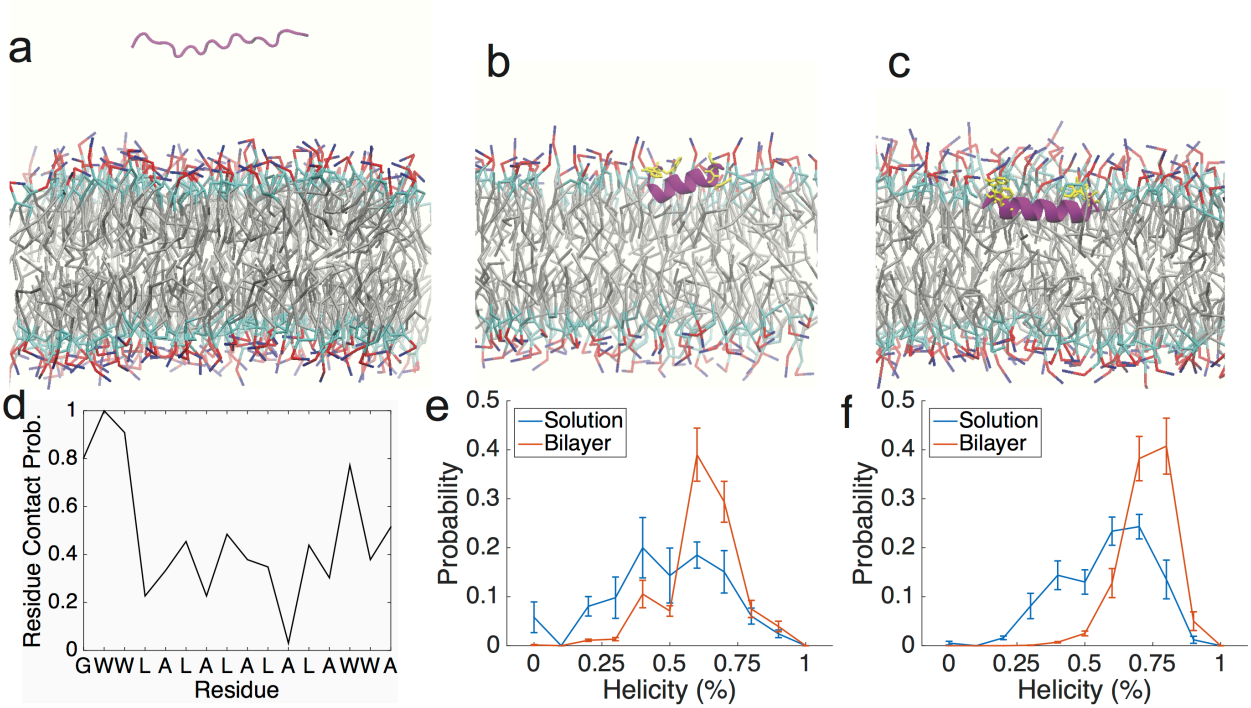

Figure 1. Folding of WALP16 and WALP19 from solution. a) Initial structure for the WALP16 folding simulation. b) WALP16 and WALP19 (c) membrane bound centroid structures. d) Normalized contact probability for WALP16 during the initial membrane binding (first 1.3 ms) phase of simulation. e-f) Helical probabilities for WALP16 (e) and WALP19 (f) for when the peptide in solution and membrane bound. Helical probabilities were computed over 5 ms of data. The simulations were divided into 1 ms blocks from which the mean probabilities and standard errors (represented by the error bars), were computed. TRP residues are shown in yellow in (b) and (c). Membrane coloring: tails=silver, glycerol=cyan, phosphate=red, choline=blue.

for WALP16 is shown in **Fig. 1a.** Both WALP16 and WALP19 fold into a helical structure and spontaneously insert into the membrane (**Fig. 1b-c**). The peptides become buried under the phosphate head groups allowing the hydrophobic residues to partition from the polar solvent/headgroup region to the nonpolar lipid core. The WALP peptides remain inserted for the duration of the 10 μs simulations, but do not transition to a TM configuration.

PACE captures important aspects of the transition from solution to a membrane bound state. Previous studies done on WALP and other membrane inserting peptides

have shown TRP-membrane interactions to be important for binding and insertion into the membrane.[62] Contact analysis shown in **Fig. 1d** and **Fig. S1** reveals that TRP

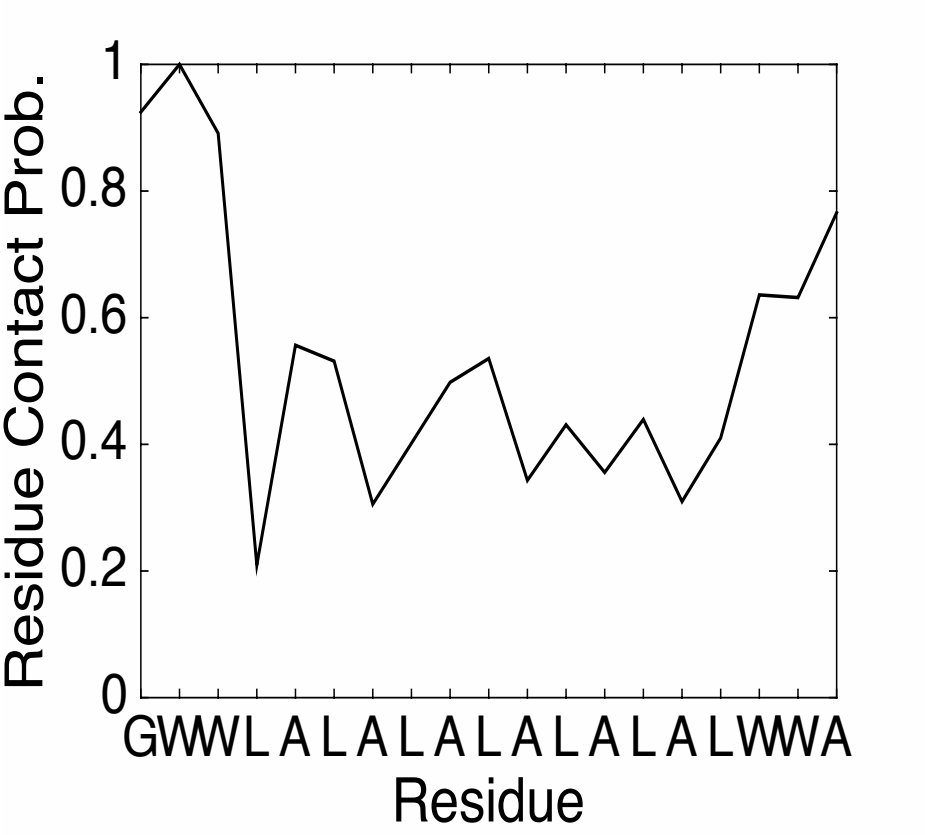

Figure S1. Normalized contact probability for WALP19 during the initial membrane binding (first 3.5 μs) phase of simulation.

residues make the most contacts with the membrane leading up to insertion. This suggests that when the peptide associates with the membrane it is mainly caused by favorable TRP-membrane interactions. Most of the contacts are transient, until two TRP residues become anchored into the membrane. At this point the insertion of the peptide is very rapid (~1 ns), as evidenced by the steep drop in center of mass distance between peptide and phosphate head groups (**Fig. S2**).

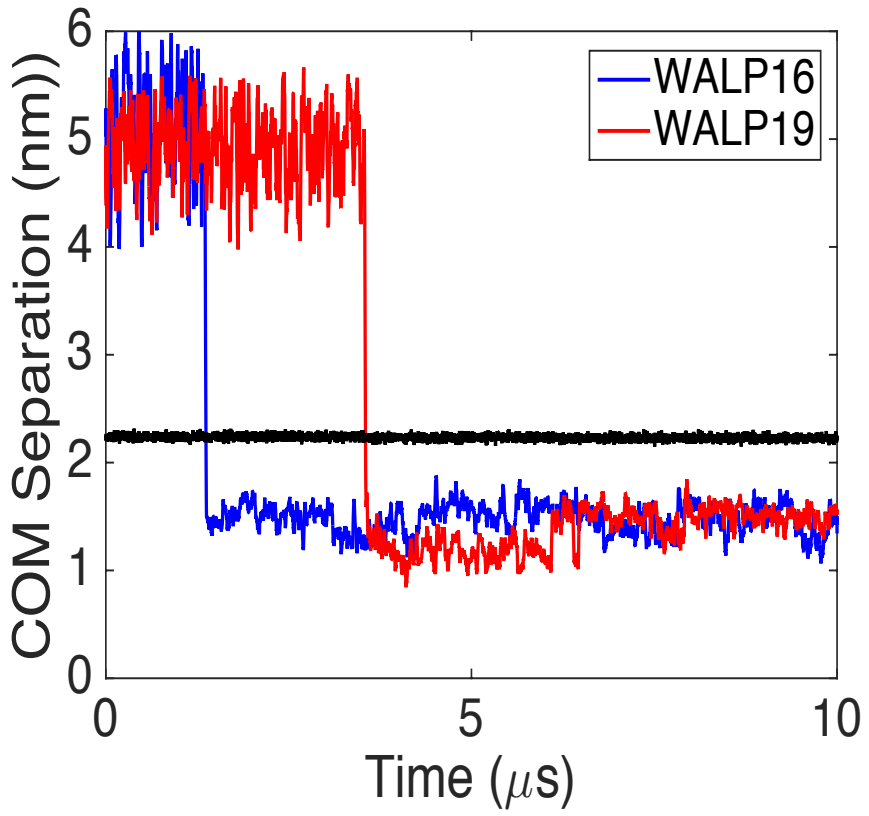

Figure S2. Center of mass separation between WALP peptides and bilayer. The mean phosphate position of the upper leaflet is shown in black.

An important aspect of evaluating PACE is whether the CG resolution of the solvent and membrane is sufficient to influence the protein structural properties. We do see that while both WALP16 and WALP19 sample folded configurations in solvent, binding to the membrane shifts the populations toward more folded conformations for both WALP16 and WALP19 (**Fig. 1e-f**). This analysis was performed by comparing WALP16/19 in a pure solvent simulation against the membrane bound configuration.

The structure of WALP peptides in aqueous solution is elusive to determine experimentally due to the hydrophobic aggregation properties. Atomistic simulations are suited to examine this problem, but differing results have been reported. Unbiased simulations have shown unfolding to occur in solution,[60] while metadynamics simulations have shown both folded and unfolded conformations are thermally accessible and the folded state is the free energy minimum.[63] There is a possibility that PACE is overstabilizing the helical content in the aqueous phase, but it does show that the membrane environment can influence the peptide structure.

Folding from TM State

In the above simulations initiated in the solvent we did not observe a transition to a TM helical conformation. This may be due to inadequate sampling (kinetic barriers) or due to the model favoring the interfacial configuration over a TM state (thermodynamic bias). To further evaluate this behavior, we initiated simulations with the peptides embedded in the membrane in an unfolded TM configuration (**Fig. 2a**). In addition to WALP16 and WALP19 we also simulated WALP23. The WALP16 peptide rapidly formed a TM helix (**Fig. 2b**), but it did not remain stable throughout the simulation. After approximately 4 ms WALP16 transitioned to an interfacial state, initially in a linear helical conformation, which then transitioned to a helical hairpin structure (**Fig. 2c**). The hairpin structure we observe is similar to hairpin structures observed in previous all-atom simulations of WALP peptides.[59,60]

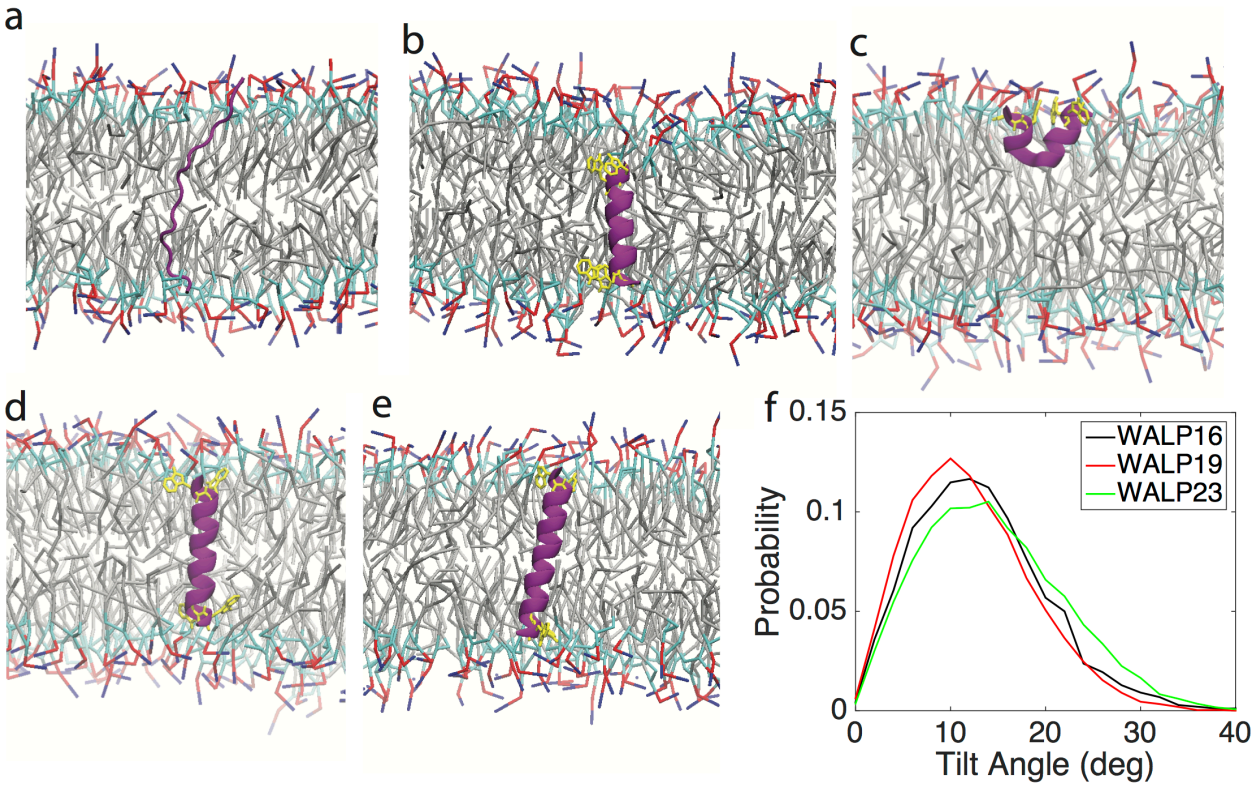

Figure 2. WALP folding from a TM state. a) Initial configuration for WALP16 TM folding simulation. WALP16 initially folds to a TM helix (b), but transitions to a surface bound helical hairpin (c) after ~4 ms. WALP19 (d) and WALP23 (e) remain in a TM helical state for the full 10 ms simulations. f) Tilt angle distributions for WALP peptides. Distributions were calculated between 500 ns and 10 ms for WALP19 and WALP23 and between 500 ns and 4 ms for WALP16.

Unlike WALP16, both WALP19 and WALP23 remain in a folded TM state through 10 ms simulations (**Fig. 2d-e**). An important structural feature of single pass TM helices is their orientation relative to the membrane normal. Different length helices tilt to different degrees to minimize hydrophobic mismatch. Tilt angles of WALP using PACE have been measured previously,[33] but not in a POPC bilayer, or since the recent force field optimization.[34] The tilt angle probabilities are shown in **Fig. 2f**, where it can be seen that WALP23 samples the highest tilt angles. For WALP19 and WALP23, the mean tilt angles are 12.1° and 14.6° respectively, which are in excellent agreement with the values of 12.5° and 14.9°, determined from all-atom enhanced sampling calculations

in POPC membranes.[61]. However, we calculate WALP16 to have an average tilt angle of 13.2°, which does not agree with the aforementioned study that determined a 6.4° tilt angle for WALP16. Given that WALP16 undergoes a large transition from a TM helix to an interfacially bound state, we interpret the larger tilting to be due the peptide being frustrated in the TM state.

Free Energy of WALP16 Membrane Insertion

WALP16 is expected to remain stable in a TM configuration in POPC, and therefore we further investigated the energetics of WALP16 membrane insertion. We performed umbrella sampling calculations to determine the PMF to transport WALP16 from solvent to a TM configuration. Computing a converged PMF of partitioning a peptide from solvent to membrane is challenging in atomistic simulations, and can take an exorbitant amount computational resources.[64] Using PACE we are able to perform 1 ms sampling in each umbrella window, which may approximate up to 10 ms of sampling in an atomistic simulation, due to the inherent acceleration of protein dynamics in PACE.[34] We are able compute a converged PMF (**Fig. 3a, S3**), which shows there is a free energy minimum at both the TM state and also at the interfacial state, when the center of mass separation between the peptide and membrane is approximately 1.5 nm. The PMF is consistent with our unbiased simulations, which show WALP16 does not transition from the surface to a TM state, which can be attributed to both a barrier (~3 kcal/mol) and the TM state not having a lower free energy.

Atomistic simulations using the GROMOS force-field have been used to compute the PMF of WALP16 insertion into a POPC bilayer.[63] In that calculation, umbrella

sampling was employed with 200 ns/window sampling along the insertion pathway. The PMF showed an overall insertion free energy change of approximately -35 kcal/mol, and the TM state was favored over the interfacial state by about 6 kcal/mol, though no significant barrier between the TM and interfacial state was observed. A subsequent study performed a multiscale sampling approach to compute the PMF of WALP16 insertion into a POPC bilayer.[65] A similar profile was computed using the multiscale sampling method the previous study, but there were some differences. The overall insertion free energy was reduced to about -30 kcal/mol and a small barrier (~2 kcal/mol) to transition from the interfacial to TM was observed. The DG between interfacial and TM states was about 6 kcal/mol, which was consistent in both PMFs.

In comparison to the AA PMFs of

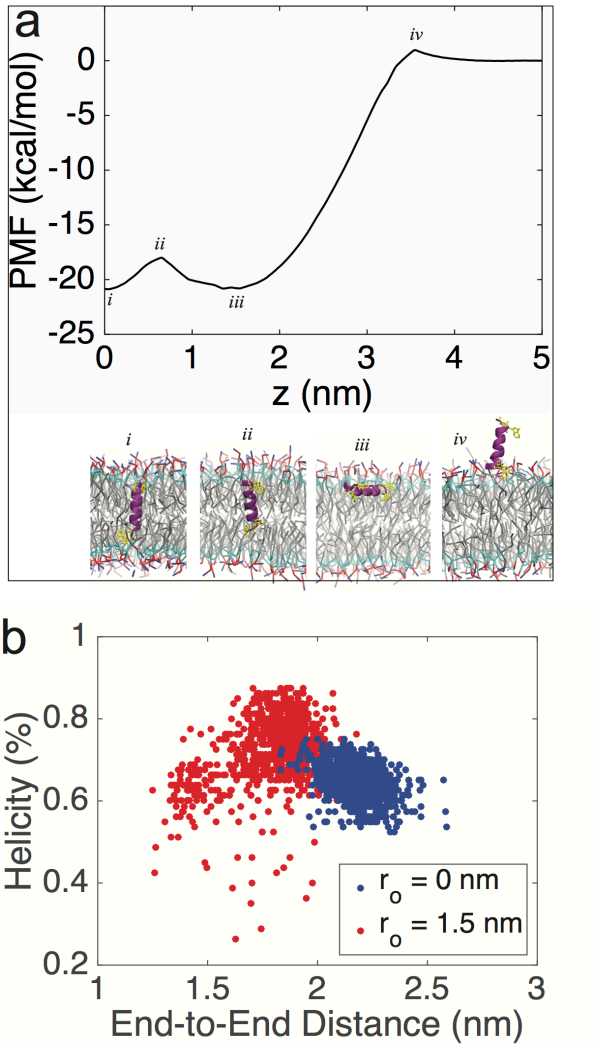

Figure 3. WALP16 insertion energetics. a) The PMF of WALP16 membrane insertion shows equal free energy minima at the TM and interface states. Centroid structures from the umbrella sampling windows corresponding to minima and maxima are indicated by roman numerals and shown below. b) The peptide end-to-end distance and helicity for the umbrella sampling window when the center of mass separation was restrained to 0 nm (TM) and 1.5 nm (interfacial). Data was analyzed between 200-1000 ns, the helicity values are smoothed over 5 ns, to remove the discreteness of the helicity values.

WALP16 insertion, it appears PACE is overstabilizing the interfacial state. This same problem was observed in pure MARTINI simulations, though there the interfacial state of WALP16 was globally stable and the barrier separating TM and interfacial states was considerably larger than what we compute for PACE.[63] In the same paper by Bereau *et al.,* they compute an atomistic insertion PMF, which shows a similar barrier between interfacial and TM configurations as we compute in PACE, but the TM state free energy is lower than the interfacial state by ~7 kcal/mol. The CG-PLUM model is also presented by Bereau *et al.,* which shows a downhill PMF of insertion, similar to the atomistic PMF, though that model does not produce a barrier between interfacial and TM states, and it produces a smaller DG of insertion by about 13 kcal/mol (DG = -22 PLUM; DG = -35 GROMOS), similar to PACE.

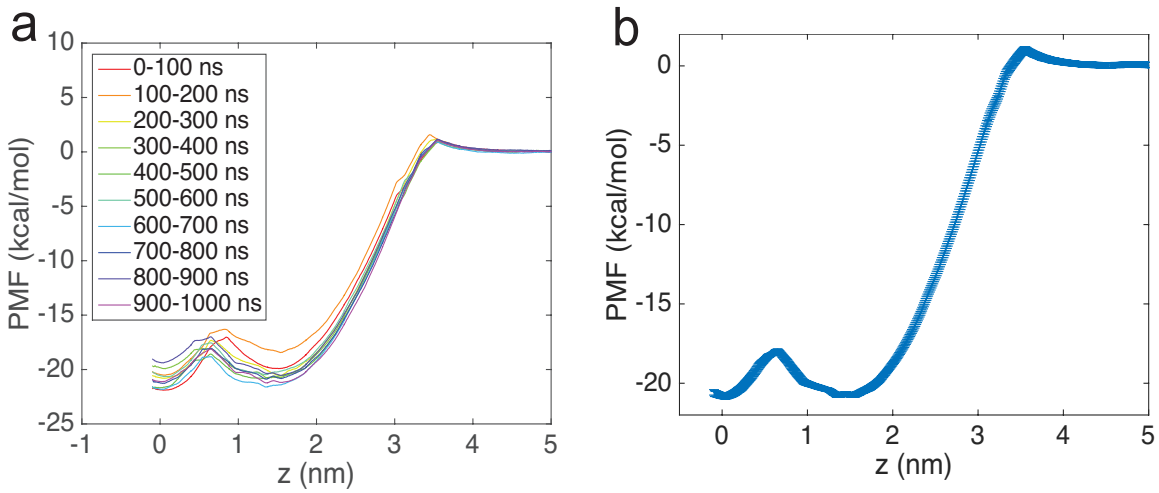

Figure S3. **a)** Membrane insertion PMF for WALP16 computed for 100 ns segments. The curves only fluctuate, but do not drift after the first 200 ns. The final PMF presented in the main text was computed using data from 200-1000 ns. b) PMF with errorbars. Errorbars are computed as the standard error, based upon the PMFs computed in 100 ns blocks from 200 ns to 1000 ns.

While PACE shows a significant improvement over MARTINI, the inability to predict the TM state as globally stable is troubling. A potential source of error in the PACE calculation may be due to the MARTINI representation of POPC lipids. We have used the historical five-bead representation of the oleoyl tail, though a four bead model has been recently introduced.[51] In the study by Bereau *et al.,* they compared the five-bead and four-bead models, and found that switching to the four-bead POPC model stabilized the TM state of WALP16 by nearly 10 kcal/mol. It is likely the inability of POPC to thin sufficiently to match the hydrophobic thickness of WALP16 in an ideal a-helical conformation is resulting in raising the energy of the TM state. From our umbrella sampling data, we examined the helicity and peptide end-to-end distance in the two minima and find that at the TM state the peptide is extended and does not sample conformations with as high helical content as at the interface (**Fig. 3b.**). With the thinner four-bead model WALP16 would likely not have to extend and break its helix to span the membrane. Thus, if we observed a similar degree of stabilization by switching to the four-bead model, we would expect our PACE PMF to be in excellent agreement with the atomistic PMF from Bereau *et al.*

Melittin

Antimicrobial peptides (AMP) are of great interest for their potential antibiotic properties, and hence understanding their mechanism of action is of great importance. Melittin is a well studied AMP, which has been shown to form pores within membranes that are transient at low peptide:lipid ratios, but can be stabilized at higher peptide concentrations.[66,67] A transient melittin pore has been observed in long-time scale

atomistic MD simulations, starting from a tetrameric TM orientation.[53] Simulations using MARTINI have also shown formation of transient melittin pores, though high peptide:lipid (1:21) ratios were required and only a single peptide was observed in a TM configuration.[68] We have utilized PACE to simulate a tetramer of melittin in a TM state embedded in a DPPC bilayer, to evaluate if water could permeate into the membrane interior and to what extent the protein dynamics were qualitatively similar to multi-ms all-atom simulations. We also performed simulations using MARTINI from the same starting tetramer configuration to evaluate what benefits PACE may have over a pure MARTINI representation.

In comparing our PACE and MARTINI simulations, we find that in PACE the peptides can adopt more varied orientations, whereas the peptides in MARTINI remain quite stable in the TM state. Furthermore, the PACE simulations show the ability to form transient water solvated pores, while the MARTINI simulation displays only minimal amounts of water permeation into the middle of the membrane. The water and phosphate densities for both PACE and MARTINI are shown in **Fig. 4a-b**, respectively, where the densities are calculated for each ms of the simulations. To further understand the pore dynamics we calculated the density at the center of the membrane every ns (**Fig. 4c**). Snapshots from the PACE simulations are shown which depict a configuration when a peptide has moved into a more lateral orientation (**Fig. 4d**), when the pore is well solvated and lipid phosphate groups have moved toward the bilayer center (**Fig. 4e**), and when the pore center is desolvated (**Fig. 4f**). Overall the PACE peptide and pore dynamics are qualitatively similar to those observed in all-atom

simulations,[69] whereas the melittin peptides in the MARTINI simulations appear to be overly stable. It may be possible to optimize the MARTINI model by removing or reducing secondary structure restraints to increase peptide flexibility, but the lack of water permeation may indicate the peptide interactions are too attractive, which would require a more extensive re-parameterization. The PACE model shows promising characteristics in modeling oligomeric TM helical pore structures and warrants further investigation.

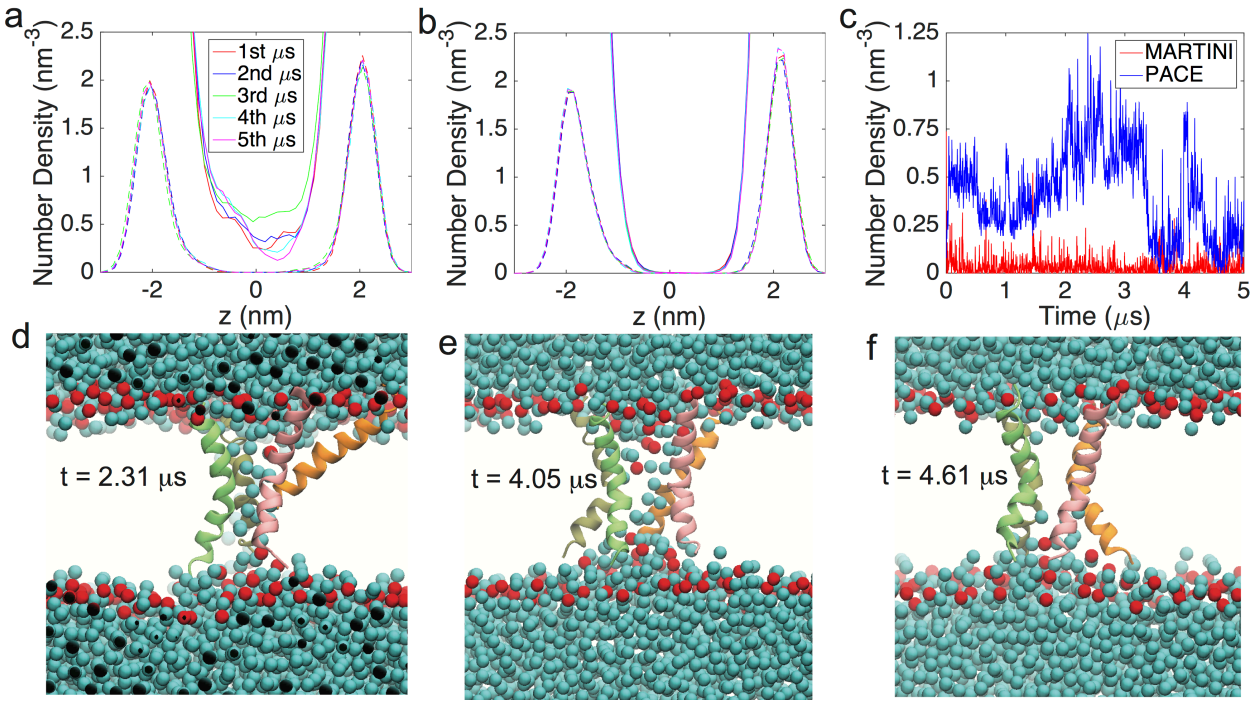

Figure 4. TM melittin tetramer in PACE and MARTINI. Water (solid lines) and phosphate (dashed lines) number density in PACE (a) and MARTINI (b) for each ms of the simulations. The water density is multiplied by four to account of the CG water mapping. Densities were computed using 100 slabs. c) Density of water at the center of the bilayer using 11 slabs, again multiplied by 4. d-f) Snapshots from the PACE simulation at various time points, CG waters are represented as cyan spheres and lipid phosphate groups as red spheres.

HAfp

While PACE performs well in modeling the simple a-helical WALP peptides and the slightly kinked melittin, we wanted to explore the applicability of PACE to a more

complex protein structure in a membrane bound state. The influenza hemagglutinin fusion peptide (HAfp) provides a more complex protein structure to analyze. Based upon an NMR structure, the 23-residue HAfp peptide adopts a helical hairpin (kinked) conformation, when bound to a micelle.[54] We have conducted simulations of the 23-residue HAfp in solution and in the presence of a DLPC bilayer to determine the ability of PACE to fold into and maintain a sharply kinked conformation consistent with the NMR structure. Our simulations entailed starting from folded and unfolded states, in solution and embedded into the bilayer, as well as starting from the folded configuration embedded in the bilayer at a range of insertion depths between 8 to 20 Å from the bilayer midplane. The HAfp peptide did not maintain a structure highly similar to the NMR structure in any of our simulations. Overall our findings show that HAfp prefers a straight helical conformation on the membrane (**Fig. 5a-b**), and while it samples a wide range of kink angles in solution (**Fig. 5b**) the angle of kinking is not as severe as the NMR hairpin structure, which is around 40^{o}. The simulation that started with HAfp folded on the membrane does maintain a kinked conformation during the first ms, during which it samples a much narrower range of kink angles than during the solution simulation. However, even during the first ms of the membrane bound simulation the kink angle is around 60^{o}, which is considerably larger than the NMR structure kink angle.

It should be noted that the configuration of HAfp is rather complex, as variety of factors including peptide sequence, peptide length, pH, and detergent/lipid composition can effect the peptide structure.[70] For the 23-residue HAfp, subpopulations of more open configurations have been detected by NMR. On a micelle at pH 4 an open

structure was shown to be in equilibrium with the hairpin structure, with the open state constituting ~20% of the population.[71] Whereas on a bilayer at neutral pH the hairpin and a semi-closed conformations were shown to be in equilibrium, with the semi-closed state constituting ~30% of the population.[72] The kink angle in the semi-closed conformation is ~20^{o} wider than the hairpin kink angle, which may correlate with our observation of structures having kink angles around 60^{o}. Nonetheless the tight hairpin structure is the dominant species in these studies and inability of PACE to show the hairpin as a stable structure is in disagreement with experiments.

It has been acknowledged that PACE may overstabilize protein hydrogen bonds[37,42] due to the inability for the water or lipids to form hydrogen bonds with the protein. This results in random coil configurations being disfavored and may also be driving HAfp toward a straight helix. Another concern is that even when HAfp is forming a kinked structure it is not consistent with the dominant NMR observed structure. To try to understand why the hairpin is so rapidly destabilized in the PACE simulations we closely examined the initial 10 ns of the membrane bound simulation. When comparing the NMR structure **(Fig. 5c**) to the structure at 10 ns (**Fig. 5d**), there is a switch in the hydrogen bonds in the kinking region (residue 11-13). In the NMR structure GLY12 forms *i*-*i*+4 H-bonds with both GLY8 and GLY16, whereas in the 10 ns structure the bonding has switched to GLU11-GLY16 and GLY12-THR15, causing a change to a more open kink angle. A contributing factor to this switch may be the interaction between the only two charged residues, GLU11 and ASP19. The change in distance between ASP19 and GLU11 is highly correlated with the changes in the hydrogen

bonding distances around the kink, as is shown in **Fig. 5e**. In the NMR structure the minimal distance between any of the GLU/ASP carboxylic acid oxygens is 8.85 Å, which is at a maximum in the PMF of ASP/GLU sidechain separation distance in the latest PACE parameterization, which includes partial charges.[34] In the 10 ns structure the distance has decreased to 6.8 Å, which is actually a minimum in the PACE ASP-GLU sidechain separation PMF. This is a somewhat non-intuitive finding given that the expectation for bringing two negatively charged side chains into closer proximity would be a destabilizing effect, or at the least a flat interaction until they are close enough to repulse, which is what the all-atom OPLS PMF shows. The maximum appearing in the PACE PMF which does not appear in the OPLS PMF, may be inconsequential to many systems, but it may be having an undesired effect on the HAfp structure and causing it to deviate from the native hairpin configuration.

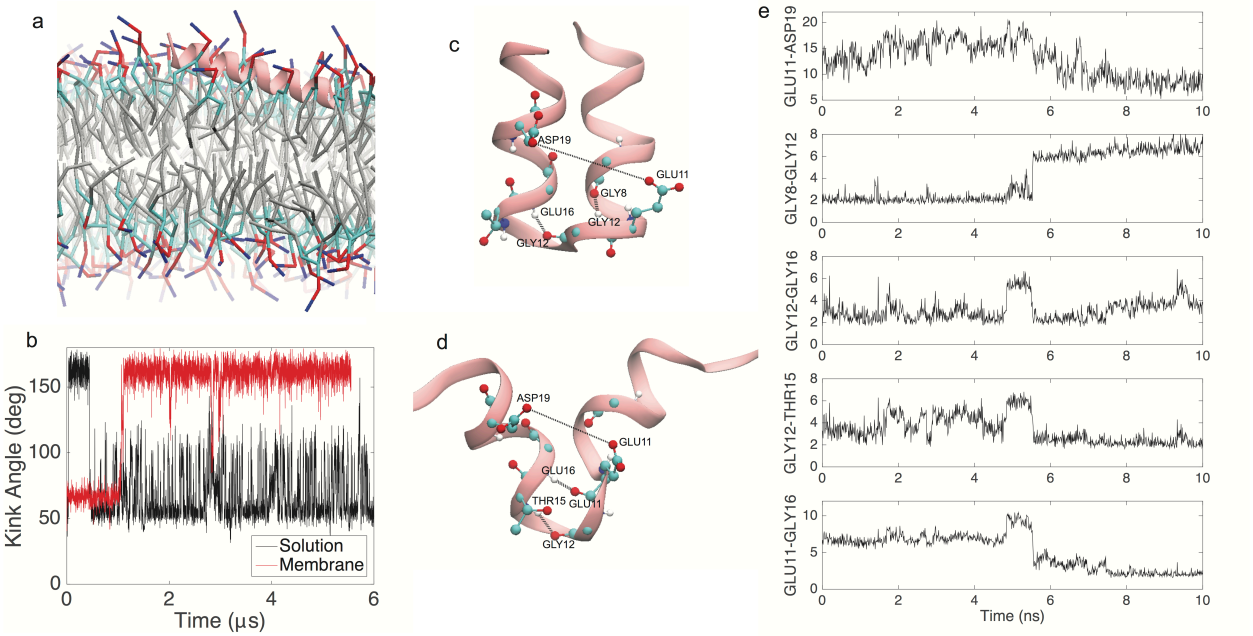

Figure 5. HAFP Analysis. a) Centroid structure from simulation initiated in hairpin conformation on the membrane. b). Kink angles for simulations started from hairpin structure in solution and on membrane. c) NMR determined structure, the H-bonds involving GLY12 and the distance between ASP19 and GLU11 are shown. d) Structure at 10 ns, for simulation starting in hairpin on the membrane. Distances between ASP19 and GLU11 side chain oxygens as well as H-bonds involving GLU11 and GLY12 are shown. e) Distance of first 10ns of membrane bound simulations of the distance pairs denoted in subfigures (c) and (d).

While PACE does not maintain a tight hairpin structure for HAfp we investigated whether MARTINI could be suitable for modeling the HAfp-membrane system. We ran analogous simulations to the PACE simulations where the kinked HAfp was started on a membrane and also ~2 nm above a DLPC membrane. The simulations started in solution rapidly associate (within 20 ns) with the membrane, while the simulations started on the membrane remain in a membrane-associated state throughout the simulation. The HAfp kink angle probability distribution for MARTINI is compared with the PACE simulations in Fig. S4. The MARTINI simulation does not maintain a tight helical hairpin, but produces a broad angular distribution with a maximum around 100^{o}. The peaks in the MARTINI angle distribution coincide with peaks in the PACE

simulations and therefore similar substates may be sampled by the different force fields. The broadness in the MARTINI distribution indicates rapid sampling of a wide range of conformations, which does not appear to be well supported by the majority of studies on the 23-residues HAfp by NMR[54,71] or atomistic MD.[73,74]

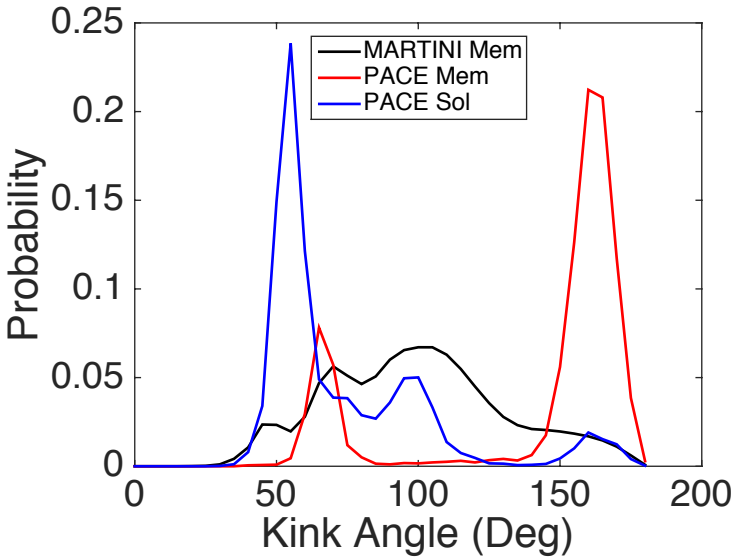

Figure S4. HAfp kink angle probability distribution comparison between MARTINI and PACE. For all simulations the initial conformation was consistent with the NMR determined structure (PDB ID:2KXA), and the simulations were run for 5 ms. The MARTINI simulation started in solution produced a kink angle probability distribution highly similar to the MARTINI membrane simulation and is therefore not shown.

CONCLUSIONS:

In this work we have examined the ability of PACE to model membrane peptide interactions for several well studied peptides. Our investigation of the designed WALP peptides show that an interfacially bound configuration is overstabilized in PACE, though we believe the stability of this state would be diminished by switching to the four-bead POPC MARTINI model. The tilt angles of WALP19 and WALP23 are in excellent agreement with all-atom simulations. Our studies of the antimicrobial melittin peptide have shown that a TM tetrameric pore modeled in PACE shows behaviors qualitatively consistent with all-atom simulations and offers significantly improved protein dynamics

compared with MARTINI. Our studies on the influenza hemagglutinin fusion peptide showed PACE did not favor the helical hairpin structure determined by NMR. A possible source of inaccuracy is the introduction of partial charges on only the charged residues. For negatively charged side chains there is a maximum in the separation distance PMF, which may drive conformational changes away from native structures. Overall PACE is a promising model for studying membrane-peptide dynamics, though further refinement of the charged interactions and compensating for the inability of CG particles to hydrogen-bond with the protein are avenues which could potentially improve the accuracy of this force field.

REFERENCES:

[1] J. Ostmeyer, S. Chakrapani, A. C. Pan, E. Perozo, B. Roux, *Nature* 2013, *501*, 121–124.

[2] R. O. Dror, D. H. Arlow, P. Maragakis, T. J. Mildorf, A. C. Pan, H. Xu, D. W. Borhani, D. E. Shaw, *Proc. Natl. Acad. Sci. U.S.A.* 2011, *108*, 18684–18689.

[3] J. Gumbart, C. Chipot, K. Schulten, *Proc. Natl. Acad. Sci. U.S.A.* 2011, *108*, 3596–3601.

[4] M. Chavent, A. L. Duncan, M. S. Sansom, *Curr Opin Struct Biol* 2016, *40*, 8–16.

[5] G. Zhao, J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken, et al., *Nature* 2014, *497*, 643–646.

[6] T. Reddy, D. Shorthouse, D. L. Parton, E. Jefferys, P. W. Fowler, M. Chavent, M. Baaden, M. S. P. Sansom, *Structure* 2015, *23*, 584–597.

[7] J. K. Weber, R. L. Jack, V. S. Pande, *J. Am. Chem. Soc.* 2013, *135*, 5501–5504.

[8] E. R. May, K. Arora, C. L. Brooks III, *J. Am. Chem. Soc.* 2014, *136*, 3097–3107.

[9] E. R. May, C. L. Brooks III, *Phys Rev Lett* 2011, *106*, 188101.

[10] E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas, K. Schulten, *Structure* 2009, *17*, 1295–1306.

[11] M. S. Friedrichs, P. Eastman, V. Vaidyanathan, M. Houston, S. Legrand, A. L. Beberg, D. L. Ensign, C. M. Bruns, V. S. Pande, *J Comput Chem* 2009, *30*, 864–872.

[12] D. E. Shaw, J. P. Grossman, J. A. Bank, B. Batson, J. A. Butts, J. C. Chao, M. M. Deneroff, R. O. Dror, A. Even, C. H. Fenton, et al., in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, Ieee, 2014, pp. 41–53.

[13] D. E. Shaw, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Lerardi, I. Kolossváry, J. L. Klepeis, T. Layman, et al., *Commun. ACM* 2008, *51*, 91–7.

[14] T. Mori, N. Miyashita, W. Im, M. Feig, Y. Sugita, *BBA - Biomembranes* 2016, 1–54.

[15] C. Abrams, G. Bussi, *Entropy* 2014, *16*, 163–199.

[16] O. Berger, O. Edholm, F. Jähnig, *Biophys J* 1997, *72*, 2002–2013.

[17] S.-W. Chiu, S. A. Pandit, H. L. Scott, E. Jakobsson, *J Phys Chem B* 2009, *113*, 2748–2763.

[18] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, P. B. Moore, M. L. Klein, *J Phys Chem B* 2001, *105*, 9785–9792.

[19] S.-J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, *J Phys Chem B* 2007, *111*, 7812–7824.

[20] S. Izvekov, G. A. Voth, *J Phys Chem B* 2005, *109*, 2469–2473.

[21] I. R. Cooke, K. Kremer, M. Deserno, *Phys Rev E Stat Nonlin Soft Matter Phys* 2005, *72*, 011506–4.

[22] W. Im, M. Feig, C. L. Brooks III, *Biophys J* 2003, *85*, 2900–2918.

[23] S. Tanizaki, M. Feig, *J. Chem. Phys.* 2005, *122*, 124706.

[24] S.-J. Marrink, A. H. de Vries, A. E. Mark, *J Phys Chem B* 2004, *108*, 750–760.

[25] S.-J. Marrink, J. Risselada, A. E. Mark, *Chem Phys Lipids* 2005, *135*, 223–244.
[26] E. R. May, D. I. Kopelevich, A. Narang, *Biophys J* 2008, *94*, 878–890.
[27] E. May, A. Narang, D. Kopelevich, *Phys Rev E* 2007, *76*, 021913.
[28] C. Arnarez, J.-P. Mazat, J. Elezgaray, S.-J. Marrink, X. Periole, *J. Am. Chem. Soc.* 2013, *135*, 3112–3120.
[29] C. Arnarez, S. J. Marrink, X. Periole, *Scientific Reports* 2013, *3*, 1263.
[30] J. M. Johnston, H. Wang, D. Provasi, M. Filizola, *PLoS Comput Biol* 2012, *8*, e1002649.
[31] A. J. Rzepiela, M. Louhivuori, C. Peter, S.-J. Marrink, *Phys. Chem. Chem. Phys.* 2011, *13*, 10437–12.
[32] T. A. Wassenaar, H. I. Ingólfsson, M. Prieß, S.-J. Marrink, L. V. Schäfer, *J Phys Chem B* 2013, *117*, 3516–3530.
[33] C.-K. Wan, W. Han, Y.-D. Wu, *J. Chem. Theory Comput.* 2012, *8*, 300–313.
[34] W. Han, K. Schulten, *J. Chem. Theory Comput.* 2012, *8*, 4413–4424.
[35] W. Han, C.-K. Wan, F. Jiang, Y.-D. Wu, *J. Chem. Theory Comput.* 2010, *6*, 3373–3389.
[36] W. Han, K. Schulten, *J Phys Chem B* 2013, *117*, 13367–13377.
[37] W. Han, K. Schulten, *J. Am. Chem. Soc.* 2014, *136*, 12450–12460.
[38] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, et al., *J. Chem. Theory Comput.* 2015, *12*, 405–413.
[39] Y. Qi, X. Cheng, W. Han, S. Jo, K. Schulten, W. Im, *J. Chem. Inf. Model.* 2014, *54*, 1003–1009.
[40] Y. Jewel, P. Dutta, J. Liu, *Proteins* 2016, 1–8.
[41] W. Zheng, F. Qin, *J Gen Physiol* 2015, *145*, 443–456.
[42] Z. Jia, J. Chen, *J Comput Chem* 2016, *37*, 1725–1733.
[43] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J Comput Chem* 2005, *26*, 1781–1802.
[44] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al., *J Comput Chem* 2009, *30*, 1545–1614.
[45] G. M. Torrie, J. P. Valleau, *J Comput Phys* 1977, *23*, 187–199.
[46] B. Roux, *Comp Phys Comm* 1995, *91*, 275–282.
[47] T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S.-J. Marrink, D. P. Tieleman, *J. Chem. Theory Comput.* 2014, *10*, 676–690.
[48] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, S.-J. Marrink, *J. Chem. Theory Comput.* 2008, *4*, 819–834.
[49] D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman, S.-J. Marrink, *J. Chem. Theory Comput.* 2013, *9*, 687–697.
[50] B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, *J. Chem. Theory Comput.* 2008, *4*, 435–447.
[51] T. A. Wassenaar, H. I. Ingólfsson, R. A. Böckmann, D. P. Tieleman, S.-J. Marrink, *J. Chem. Theory Comput.* 2015, *11*, 2144–2155.

[52] T. C. Terwilliger, D. Eisenberg, *J. Biol. Chem.* 1982, *257*, 6010–6015.
[53] J. M. Leveritt, A. Pino-Angeles, T. Lazaridis, *Biophys J* 2015, *108*, 2424–2426.
[54] J. L. Lorieau, J. M. Louis, A. Bax, *Proc Natl Acad Sci U S A* 2010, *107*, 11341–11346.
[55] W. Kabsch, C. Sander, *Biopolymers* 1983, *22*, 2577–2637.
[56] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, V. S. Pande, *Biophys J* 2015, *109*, 1528–1532.
[57] J. A. Killian, *FEBS Letters* 2003, *555*, 134–138.
[58] A. E. García, J. N. Onuchic, *Proc. Natl. Acad. Sci. U.S.A.* 2003, *100*, 13898–13903.
[59] W. Im, C. L. Brooks III, *Proc. Natl. Acad. Sci. U.S.A.* 2005, *102*, 6771–6776.
[60] M. B. Ulmschneider, J. P. F. Doux, J. A. Killian, J. C. Smith, J. P. Ulmschneider, *J. Am. Chem. Soc.* 2010, *132*, 3452–3460.
[61] T. Kim, W. Im, *Biophys J* 2010, *99*, 175–183.
[62] H. Nymeyer, T. B. Woolf, A. E. García, *Proteins* 2005, *59*, 783–790.
[63] T. Bereau, W. F. D. Bennett, J. Pfaendtner, M. Deserno, M. Karttunen, *J. Chem. Phys.* 2015, *143*, 243127–12.
[64] C. Neale, J. C. Y. Hsu, C. M. Yip, R. Pomès, *Biophys J* 2014, *106*, L29–31.
[65] T. Bereau, K. Kremer, *J Phys Chem B* 2016, *120*, 6391–6400.
[66] M.-T. Lee, W.-C. Hung, F.-Y. Chen, H. W. Huang, *P Natl Acad Sci Usa* 2008, *105*, 5087–5092.
[67] M. T. Lee, T. L. Sun, W. C. Hung, H. W. Huang, *Proc Natl Acad Sci U S A* 2013, *110*, 14243–14248.
[68] K. P. Santo, S. J. Irudayam, M. L. Berkowitz, *J Phys Chem B* 2013, *117*, 5031–5042.
[69] J. M. Leveritt III, A. Pino-Angeles, T. Lazaridis, *Biophys J* 2015, *108*, 2424–2426.
[70] R. Worch, *Acta Biochimica Polonica* 2014, *61*, 421–426.
[71] J. L. Lorieau, J. M. Louis, *Proc Natl Acad Sci U S A* 2012, *109*, 19994–19999.
[72] U. Ghosh, L. Xie, L. Jia, S. Liang, D. P. Weliky, *J. Am. Chem. Soc.* 2015, *137*, 7548–7551.
[73] A. R. Brice, T. Lazaridis, *J Phys Chem B* 2014, *118*, 4461–4470.
[74] J. L. Baylon, E. Tajkhorshid, *J Phys Chem B* 2015, *119*, 7882–7893.

Chapter 4:

Closing Remarks and Future Directions

The current study has focused on using MD simulations to better understand membrane active peptides. The initial study focused on how pH activates the membrane lytic activity of γ-peptide from NωV. Simulations identified a different mode of binding at pH 6 (inactive) and pH 8 (active). Specifically, the increase in pH promotes deprotonation of a critical Lys residue, which causes the helical γ-peptide to kink and become more compact. This kink allows γ-peptide to become more buried in the membrane. Kink formation is known to aid in activity of other membrane active peptides discussed in Chapter 1. Thus, the pH dependent structural changes are potentially important in regulating the lytic activity of γ-peptide. The resources available for this study only permit studying a single peptide's interaction with the membrane with all-atom resolution. No membrane disruption was observed by a single peptide, even in the active pH 8 environment. A study of many peptides in a membrane environment would help better understand membrane disruption by γ-peptide.

Simulating membrane disruption requires long timescales and large system sizes since membrane reorganization is slow and multiple peptides are often required to disrupt the membrane. Previously, there was no proven model to effectively study pore formation by membrane active peptides. All-atom simulations are accurate, but can't reach timescales to sufficiently sample membrane disruption. Existing coarse-grained models like MARTINI can access long timescales, but do not have high enough

resolution to model protein dynamics well enough to properly model structural changes during the transitions from solution phase to membrane pores. Our study evaluating the recently developed PACE force field has shown that PACE is an appropriate model choice when modeling membrane pores formed by membrane active peptides. For this application PACE is a promising compromise in being able to properly model pore dynamics, while accessing long timescales. However, PACE shows mixed results for properly folding and inserting membrane peptides. While PACE accurately folded and inserted the simple transmembrane WALP helices, PACE failed to accurately model the kinked membrane bound conformation and dynamics of HAfp. Thus, one should use PACE to understand protein dynamics on a membrane only in conjunction with structural data from experiment.

Going forward, PACE may be useful for studying pore formation and stabilization by lytic peptides from nonenveloped viruses. γ-peptide from NωV is not an ideal candidate since its activity is pH-dependent and PACE has not been optimized for any environment other than neutral pH. However, Flockhouse virus (FHV) is a related virus that also employs a short, pore-forming lytic peptide (also called γ-peptide) that is active at neutral pH and has been structurally characterized on a membrane. PACE could be used to gain insight on the structure and dynamics of the pores formed by FHV γ-peptide. Understanding this specific case is useful, but its serves a more general purpose. The mechanistic details of pore formation are not well understood and PACE can help to further advance our understanding of this critical biological process.