2022

Enhanced Sampling Techniques Reveal Key Information about Membrane Active Peptides

Nicolas Charles Frazee
West Virginia University, ncf0003@mix.wvu.edu

Enhanced Sampling Techniques Reveal Key Information about Membrane Active Peptides

Nicolas Frazee

Dissertation submitted to the Eberly College of Arts and Science at West Virginia University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Chemistry

Blake Mertz, Ph.D., Committee Chairperson
Justin Legleiter, Ph.D.
Aldo Romero, Ph.D.
Brian Popp, Ph.D.
Peng Li,Ph.D.

C. Eugene Bennet Department of Chemistry

Morgantown, West Virginia
2022

Keywords: pHLIP, lariat peptides, cyclic peptides, Huntington's Disease, amyloid formation, lipid tail unsaturation, WESTPA, enhanced sampling, GaMD, constant pH, transmembrane α-helix, aphipathic α-helix

ABSTRACT

Enhanced Sampling Techniques Reveal Key Information about Membrane Active Peptides

Nicolas Frazee

Membrane active peptides (MAPs) show promise in terms of future drug development. Whether it be adapting macrocycles for use in targeting protein-protein interactions or adopting cell-penetrating peptides (CPPs) to cause lysis in target cells, the field is burgeoning with possibilities for designer drugs. The following work encompasses the exploration of three such peptides. The pH-Low Insertion Peptide (pHLIP) is a membrane-active peptide that spontaneously folds into a transmembrane α-helix upon acidification. This activity enables pHLIP to potentially act as a vector for drugs related to diseases characterized by acidosis such as cancer or heart ischemia. First, we explored the conformational space sampled by pHLIP while in bulk solution via constant pH molecular dynamics (MD) simulations. It was determined that pHLIP's acidic residues are similar to single-residue-in-solution values and the P20G maintains a higher helicity in solution than wt-pHLIP. The next study was on the 17 N-terminal residues of the huntingtin (htt) protein (Nt17). Nt17 is essential for the patheoogenesis of Huntington's Disease through its role in both htt aggregation and membrane association. We investigated Nt17 and its association with three model membranes with a consistent headgroup and tails with varying degrees of unsaturation and length. We found no correlation between the effect of lipid vesicles on aggregation and the degree of htt-lipid complexes formed, supporting that the properties of the membrane have direct influence on the aggregation mechanism. We also determined that Nt17-membrane association is regulated by complimentarily-sized hydrophobic residues in Nt17 and defects in the lipid bilayer. Finally, we developed a high-throughput assay for determining the

permeability cyclic peptides. Targeting protein-protein interactions with traditional small-molecule drugs can be challenging when the binding pocket is too large. However, cyclic peptides are the key to targeting these interactions: passive membrane permeability and more consistent structure to prevent off targeting. Using a library of peptides with known permeability, we performed Gaussian accelerated molecular dynamics (GaMD) simulations on roughly 200 peptides of the library in octanol and water to estimate their permeability. Initially, we did not directly replicate the permeability from the experimental results, however, we did replicate the trend that correlates N-methylation and permeability. After using PCA to determine the peptides with biggest difference in populations between octanol and water, we more successfully reproduced the permeability data from experiment for those peptides.

ACKNOWLEDGEMENTS

I am overwhelmingly appreciative for all the support and guidance I received from many individuals in my life to be able to accomplish this task.

I'd like to thank Dr. Blake Mertz for acting as my advisor and mentor for my years at WVU. I feel like one of the main reasons I picked the Mertz Lab as my research group when I first joined was that I felt like we would be able to get along as friends, and I wasn't wrong. Thank you for all the guidance you've given, for always having my back, and for teaching me what a research professor can be.

I'd like to thank my parents for never expecting but always encouraging my success. This philosophy sometimes made motivating myself challenging but always meant that I was doing what I was doing for myself and no one else. To the rest of my family, thank you for always filling my life with happiness and laughter; when I'm around you lot of noisy, obnoxious, hilarious people I feel most at home.

I'd also like to give thanks for everyone who has mentored me along my academic career: Dr. Baldauff for encouraging me to switch to chemistry and go to grad school, as well as her insane, fun antics; Dr. Williams for reminding me that school is for learning and not necessarily for grades; and Dr. LaCount for mentoring me during my first research project. I'd also like to thank my PhD committee members Dr. Legleiter, Dr. Popp, Dr. Li, and Dr. Romero for their feedback over the course of my graduate career. I'd especially like to thank Dr. Legleiter for his guidance during our collaboration with him; I love his inquisitiveness and excitement for digging deeper and tying disparate facts together into a cogent story.

This also wouldn't have been possible without the friends I've made throughout my life, new and old, who have always been there for me when I needed them. I'd like to specifically thank Maryssa Beasley, Sharon Groover, and Blaine McClay for being my friends in the program keeping me sane; Kyle Billings, Violeta Burns, and Hannah Scott for always making the lab a joy. To everyone else, you know who you are and I love you for it.

Contents

List of Figures

List of Tables

1. INTRODUCTION

1.1 THE PERMEATION AND ASSOCIATION OF MEMBRANE ACTIVE PEPTIDES

Membrane active peptides (MAPs) are a broad class of proteins responsive to the presence of a phospholipid bilayer, either transitioning cargo through it or disrupting it to cause cell lysis. This makes MAPs an attractive alternative to current pharmaceuticals to attack invasive membrane-bound species. In particular, antimicrobial peptides (AMPs), integral components of the host defense systems with variations unique to organisms, are being adapted for use to target a variety of invasive species[1]. AMPs are usually under 30 residues long and have a higher than average concentration of hydrophobic residues to allow themselves to associate more easily with the acyl chains of the phospholipid bilayer. Most AMPs operate by disrupting the membrane, ultimately causing lysis. These peptides are often randomly coiled in solution and transition to a more rigid α-helix or β-sheet structure in the presence of a cell membrane. This forces the backbone into a more rigid position, given the hydrogen bonds, which may be responsible for modulating the potency with which the peptide disrupts the bilayer[2, 3]. However these peptides do not uniformly target membranes often preferring either charged or neutral membranes. Though this preference may purely be in association and not because the AMP cannot disrupt all membrane compositions [4, 5, 6]. Given this, there is potential that AMPs can be tuned to address the growing ranks of bacteria that have developed resistance against the traditional antibacterial drugs. Although AMPs generally cause lysis in cells, there are peptides that at lower concentrations will simply localize to the the target cell without harm; these peptides in particular show promise for diagnostic imaging.

Additionally, cell penetrating peptides (CPPs) are being studied as vectors for drug delivery; these short (under 30 residues) peptides are capable of transferring cargo several

times larger than its own molecular weight[7]. There are many different processes by which CPPs cross the cell membrane; some are more benign (reliance on transient membrane disruption or endocytosis) while some are aggressive enough to cause cell lysis. The most interesting in terms of drug delivery employ predominantly more harmless methods to cross the membrane. At higher concentrations, very cationic CPPs (i.e., Arg9, *tat*, penetratin) can effectively cross anionic synthetic membranes, however this most likely occurs via disruption of the bilayer structure[8]. Instead, with moderate to lower peptide concentrations or in bilayers with too few anionic lipids, many cationic CPPs will not cross the bilayer; therefore, they seem almost entirely dependant on anionic lipids to cross the membrane. Some more anionic CPPs (i.e., pHLIP, ATRAM) will localize to tissues that are acidotic allowing potential to target CPP drug delivery to acidotic cells such as those related to cancer and heart ischemia.

Membrane-permeable cyclic peptides are attractive to many scientists and pharmaceutical companies as potential drug candidates[9]. The selection of possible peptide sizes allows selectivity for the target receptor and modulates the protein-protein interaction. While linear peptides are flexible, and therefore lack affinity, selectivity, and bioavailability, cyclic peptides are more rigid and adopt structures with affinity for protein surfaces with high specificity. Paramount to the identification of drug-viable cyclic peptides is the determining the potential aqueous structures of the peptide and the membrane permeability of the peptide.

1.2 MEMBRANE-ACTIVE PEPTIDES IN THIS WORK

1.2.1 pH low insertion peptide (pHLIP)

The first MAP explored in this work, the pH Low Insertion Peptide (pHLIP)[10], is 36 residues and anionic that is randomly coiled both in solution (state I) and when bound to a membrane at neutral pH (state II) (Fig. 1.1). Upon acidification, pHLIP will insert unidirectionally, forming a transmembrane α-helix (state III). This transition occurs with sufficient free energy that cargo of various sizes and polarities can be attached to the C-terminus of pHLIP and shuttled across the membrane[11]. Wild type (wt) pHLIP will insert in the pH range of 5.2-6.5[12].

Through a combination of circular dichroism (CD) and measurement of pHLIP's inherent tryptophan fluorescence over a pH-gradient, the basic mechanism through which

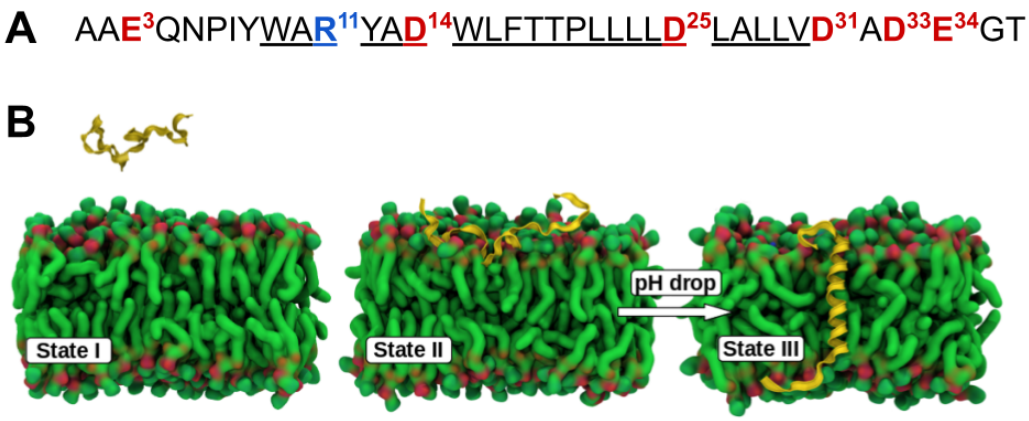

Figure 1.1: Description of pHLIP. A) The primary sequence of the pH Low Insertion Peptide. *Red:* acidic residues; *Blue:* basic residue; *Underline:* transmembrane region. B) pHLIP insertion into a bilayer. *Red/green:* phospholipids; *yellow ribbon:* pHLIP's protein backbone. pHLIP adopts a randomly coiled conformation in solution(state I). Once bound to a membrane (state II), upon acidification, pHLIP will spontaneously and unidirectionally insert into the membrane, forming a transmembrane α-helix (state III).

pHLIP is acidified and inserts has been determined. Based on existing spectroscopic data[12], pHLIP interacts with the membrane in the following order: 1) partition the sinker stretch (residues 21-30) into the head group region, 2) the transmembrane region will form helix and move into the head group region, and 3) the C-terminus will partition and eventually transition entirely through the bilayer with the N-terminus remaining in the extracellular medium. Due to the spontaneity of pHLIP's insertion into the bilayer, transitioning pHLIP from state II to state III would incur a great energy penalty and research is underway to determine this penalty. However, fine details relating pHLIP's sequence to structure and function are yet to be understood. To fully model pHLIP's transition from state I-III, we need to acquire atomistic knowledge of the system in an environment similar to a cancer cell target.

1.2.2 The 17 N-terminal residues of the huntingtin protein (Nt17)

Huntington's Disease (HD) is a fatal neurodegenerative disorder caused by an expanded glutamine repeat region (polyQ) within the huntingtin protein (htt). The progression of HD involves the formation of amyloid inclusions which is driven by aggregation of the huntingtin protein in people with the disease. The 17 N-terminal residues of the huntingtin protein

(Nt17) is fundamentally important for the pathogenesis of HD due to its role in the formation of these aggregates[13, 14, 15]. Nt17 is intrinsically disordered in bulk solution but shifts conformations in the presence of a binding partner to an amphipathic α-helix[16, 17], an α-helix with distinct hydrophobic and hydrophilic sides. Nt17 α-helices can interact with each other (Fig. 1.2) to form several multimeric species that based on the region of Nt17 exposed to the solvent, can cause Nt17 to drive the htt oligomerization[18]. These oligomers are a potent toxic aggregate species, and their formation is predicated on the presence of Nt17, as when it is removed from the htt, the protein no longer forms nonamyloid oligomers[15, 19].

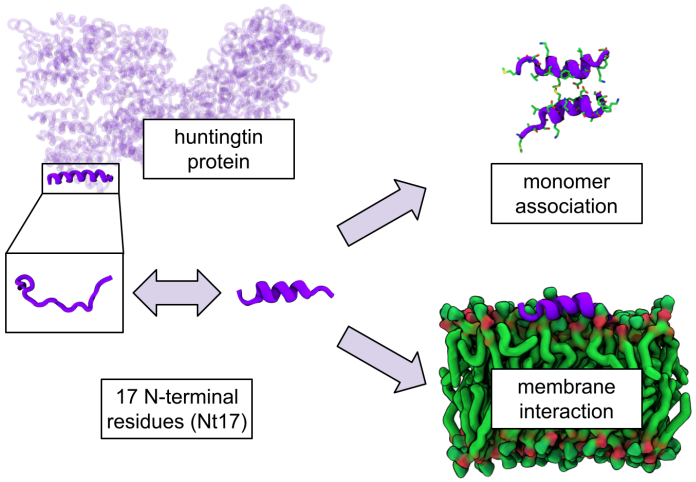

Figure 1.2: Function of Nt17. The 17 N-terminal residues of the huntingtin protein (Nt17) is unstructured in bulk solution. However, it can undergo a conformational shift to an amphipathic α-helix when in proximity to a binding partner. Nt17 is essential for both the formation of huntingtin aggregates and interaction with the cell membrane.

Nt17 is not only responsible for htt aggregation but also acts as a lipid binding domain (Fig. 1.2) for htt exhibiting many of the same properties associated with the amphipathic α-helical motif, namely being associated with lipid binding[20][21] and a preference for curvature induced defects[22]. When a polyQ peptide is flanked with Nt17, it effectively disrupts lipid vesicles, whereas the polyQ peptide alone does not interact significantly[23]. Therefore, Nt17 is absolutely necessary for htt to interact with the membrane.

1.2.3 Cyclic lariat peptides

A particular class of cyclic peptides known as lariat depsipeptides make up roughly 30% of the naturally occurring cyclic peptides according to a survey[24] of the Natural Products Atlas[25]. The majority of these peptides are cyclized by a connection between the C-terminal carboxylic acid and a side-chain hydroxyl group to form an ester linkage. Of particular interest is the library of peptides studied by C. Kelly, et al.[24] (Figure 1), 4096 nine-residue peptides with an ester linkage between the C-terminus and Thr3 with the members of the library being various possible N-methylation sites as well as each residue being able to have either a D or L chirality. The study from C. Kelly, et al. determined there was a strong positive relationship between the permeability of the peptides and both the number of heterochiral residues (D chirality) and the degree of N-methylation.

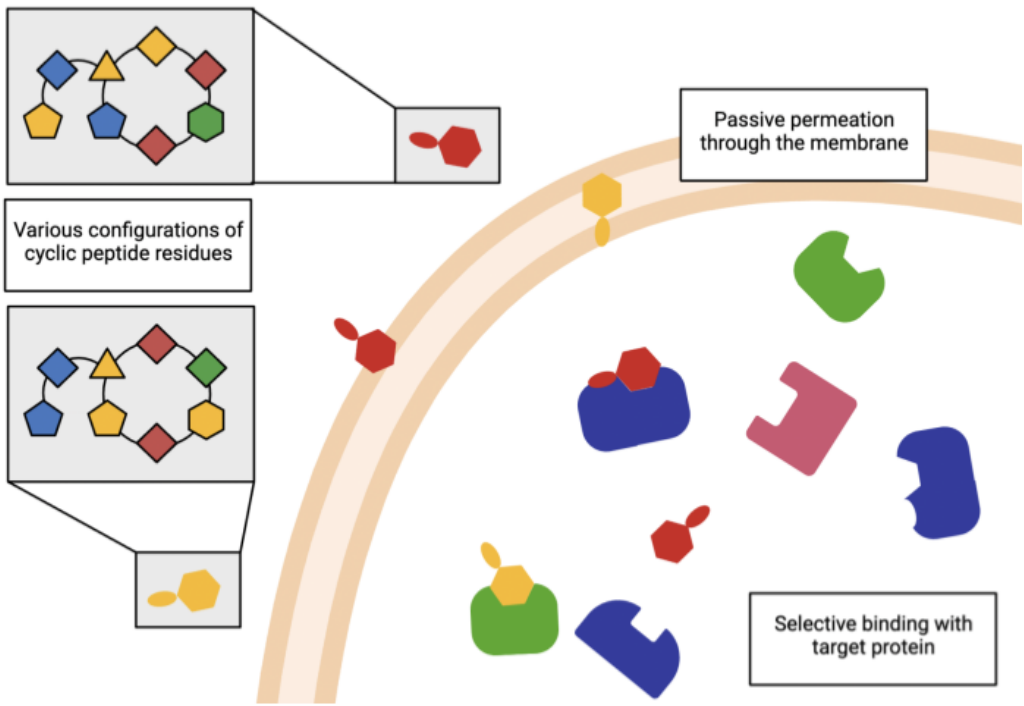

Figure 1.3: Description of lariat peptides. Cyclic lariat peptides are highly customizable analogues for small molecule drugs. Unlike linear small peptides, the ring structure of the peptides reduces flexibility and therefore increases target specificity.

1.3 MODELING MEMBRANE-ACTIVE PEPTIDES VIA MOLECULAR DYNAMICS

Molecular dynamics (MD) is an effective tool to evaluate the properties of MAPs even when simply using equilibrium MD; the high resolution in both time and model allow high precision when investigating these short peptides. Recently, Chen, et al. utilized long-scale, unbiased atomistic MD simulations to reveal that a single point mutation strongly affected the development of functional dynamic and heterogeneous aggregates. This ultimately led to a framework that allowed prediction of aggregation and assembly based solely on peptide sequence[26]. Another study[27] determined via circular dichroism that WALP peptides remained inserted in the bilayer and fully helical at temperatures as high as 90 °C. These higher temperatures were used to accelerate equilibrium MD simulations increasing their rate of sampling by $\tilde{5}$0-500 times. This allowed the observation of a folded insertion pathway and the direct calculation of insertion kinetics. While equilibrium MD is a very effective tool for MAP investigation, there are several enhanced sampling techniques that can potentially answer the pertinent questions while using fewer computing hours.

Enhanced sampling techniques are methods that allow the user to answer some specific questions with their MD simulations. For example, Neale, et al. made use of umbrella sampling (US), an enhanced sampling method that requires breaking a reaction coordinate into windows and simulating the system between each of the windows, to compute the standard binding free energy of an AMP called indolicidin as it inserts into a phospholipid bilayer[28]. Several other enhanced sampling methods allow for the calculation of a binding free energy (i.e, metadynamics[29, 30]), or the measurement of the permeability of a peptide[31]. Replica-exchange with solvent tempering (REST) is an effective means of accelerating simulations towards convergence by ultimately increasing the conformational sampling of the system; the implementation by Appadurai, et al. is especially tailored for MAPs as it was specifically designed to increase sampling accurately across a wide variety of proteins[32]. Finally, of particular interest to highly anionic or cationic MAPs, constant pH is a method that allows protonation states to fluctuate as the simulation progresses. This has been used to great effect on pHLIP, as it depends on protonation for insertion, by Vila-Viçosa, et al. to determine membrane-induced pK_a shifts in pHLIP and the L16H variant[33]. In concert, these methods allow exploration of several different important aspects of MAPs.

1.4 THIS WORK

The subsequent chapters of this document contain applications of enhanced sampling MD methods on several MAPs. Chapter 2 is a description of the fundamentals of molecular dynamics and some of the approximations taken to accelerate the processes. Next, we explore the application of the constant pH method on pHLIP and the P20G variant in solution with different salt concentrations. Chapter 4 follows with the investigation of Nt17 and its association with model membranes with varying chain lengths in the tails via a weighted-ensemble approach. Finally, Chapter 5 is a description of the development of a method to efficiently test the permeability of potential cyclic peptide drugs using gaussian-accelerated MD.

1.5 BIBLIOGRAPHY

[1] F. G. Avci, B. Sariyar Akbulut, and E. Ozkirimli. “Membrane Active Peptides and Their Biophysical Characterization”. In: *Biomolecules* 8.3 (Aug. 22, 2018). ISSN: 2218-273X. DOI: 10.3390/biom8030077. pmid: 30135402. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6164437/ (visited on 12/07/2018).

[2] K. Luna-Ramírez, M.-A. Sani, J. Silva-Sanchez, J. M. Jiménez-Vargas, F. Reyna-Flores, K. D. Winkel, C. E. Wright, L. D. Possani, and F. Separovic. “Membrane interactions and biological activity of antimicrobial peptides from Australian scorpion”. eng. In: *Biochim biophys acta* 1838.9 (Sept. 2014), pp. 2140–2148. ISSN: 0006-3002. DOI: 10.1016/j.bbamem.2013.10.022.

[3] D. I. Fernandez, M.-A. Sani, A. J. Miles, B. A. Wallace, and F. Separovic. “Membrane defects enhance the interaction of antimicrobial peptides, aurein 1.2 versus caerin 1.1”. eng. In: *Biochim biophys acta* 1828.8 (Aug. 2013), pp. 1863–1872. ISSN: 0006-3002. DOI: 10.1016/j.bbamem.2013.03.010.

[4] S.-C. Park, J.-Y. Kim, S.-O. Shin, C.-Y. Jeong, M.-H. Kim, S. Y. Shin, G.-W. Cheong, Y. Park, and K.-S. Hahm. “Investigation of toroidal pore and oligomerization by melittin using transmission electron microscopy”. eng. In: *Biochem biophys res commun* 343.1 (Apr. 2006), pp. 222–228. ISSN: 0006-291X. DOI: 10.1016/j.bbrc.2006.02.090.

[5] L. Yang, T. A. Harroun, T. M. Weiss, L. Ding, and H. W. Huang. “Barrel-stave model or toroidal model? A case study on melittin pores.” In: *Biophys j* 81.3 (Sept. 2001), pp. 1475–1485. ISSN: 0006-3495. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1301626/ (visited on 09/10/2022).

[6] F. Yoneyama, Y. Imura, K. Ohno, T. Zendo, J. Nakayama, K. Matsuzaki, and K. Sonomoto. “Peptide-lipid huge toroidal pore, a new antimicrobial mechanism mediated by a lactococcal bacteriocin, lacticin Q”. eng. In: *Antimicrob agents chemother* 53.8 (Aug. 2009), pp. 3211–3217. ISSN: 1098-6596. DOI: 10.1128/AAC.00209-09.

[7] M. Lindgren, M. Hällbrink, A. Prochiantz, and Ü. Langel. “Cell-penetrating peptides”. In: (), p. 5.

[8] H. Herce, A. Garcia, J. Litt, R. Kane, P. Martin, N. Enrique, A. Rebolledo, and V. Milesi. “Arginine-Rich Peptides Destabilize the Plasma Membrane, Consistent with a Pore Formation Translocation Mechanism of Cell-Penetrating Peptides”. In: *Biophys j* 97.7 (Oct. 2009), pp. 1917–1925. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2009.05.066. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2756373/ (visited on 09/11/2022).

[9] C. Morrison. “Constrained peptides’ time to shine?” en. In: *Nat rev drug discov* 17.8 (Aug. 2018), pp. 531–533. ISSN: 1474-1776, 1474-1784. DOI: 10.1038/nrd.2018.125. URL: http://www.nature.com/articles/nrd.2018.125 (visited on 04/01/2022).

[10] J. F. Hunt, P. Rath, K. J. Rothschild, and D. M. Engelman. “Spontaneous, pH-Dependent Membrane Insertion of a Transbilayer α-Helix†”. In: *Biochemistry* 36.49 (Dec. 1997), pp. 15177–15192. ISSN: 0006-2960. DOI: 10.1021/bi970147b.

[11] D. Thévenin, M. An, and D. M. Engelman. “pHLIP-mediated Translocation of Membrane Impermeable Molecules into Cells”. In: *Chemistry & biology* 16.7 (July 31, 2009), pp. 754–762. ISSN: 1074-5521. DOI: 10.1016/j.chembiol.2009.06.006. pmid: 19635412. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2741147/ (visited on 10/04/2018).

[12] H. L. Scott, J. M. Westerfield, and F. N. Barrera. “Determination of the Membrane Translocation p K of the pH-Low Insertion Peptide”. In: *Biophysical journal* 113.4 (Aug. 2017), pp. 869–879. ISSN: 00063495. DOI: 10.1016/j.bpj.2017.06.065. URL:

http://linkinghub.elsevier.com/retrieve/pii/S0006349517307531 (visited on 08/14/2018).

[13] T. E. Williamson, A. Vitalis, S. L. Crick, and R. V. Pappu. “Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin”. eng. In: *J mol biol* 396.5 (Mar. 2010), pp. 1295–1309. ISSN: 1089-8638. DOI: 10.1016/j.jmb.2009.12.017.

[14] M. Jayaraman, R. Kodali, B. Sahoo, A. K. Thakur, A. Mayasundari, R. Mishra, C. B. Peterson, and R. Wetzel. “Slow amyloid nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments”. eng. In: *J mol biol* 415.5 (Feb. 2012), pp. 881–899. ISSN: 1089-8638. DOI: 10.1016/j.jmb.2011.12.010.

[15] M. Jayaraman, R. Mishra, R. Kodali, A. K. Thakur, L. M. I. Koharudin, A. M. Gronenborn, and R. Wetzel. “Kinetically competing huntingtin aggregation pathways control amyloid polymorphism and properties”. eng. In: *Biochemistry* 51.13 (Apr. 2012), pp. 2706–2716. ISSN: 1520-4995. DOI: 10.1021/bi3000929.

[16] M. Michalek, E. S. Salnikov, and B. Bechinger. “Structure and Topology of the Huntingtin 1–17 Membrane Anchor by a Combined Solution and Solid-State NMR Approach”. In: *Biophys j* 105.3 (Aug. 2013), pp. 699–710. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2013.06.030. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3736738/ (visited on 09/09/2022).

[17] M. Michalek, E. S. Salnikov, S. Werten, and B. Bechinger. “Membrane interactions of the amphipathic amino terminus of huntingtin”. eng. In: *Biochemistry* 52.5 (Feb. 2013), pp. 847–858. ISSN: 1520-4995. DOI: 10.1021/bi301325q.

[18] J. R. Arndt, S. G. Kondalaji, M. M. Maurer, A. Parker, J. Legleiter, and S. J. Valentine. “Huntingtin N-terminal monomeric and multimeric structures destabilized by covalent modification of heteroatomic residues”. In: *Biochemistry* 54.28 (July 2015), pp. 4285–4296. ISSN: 0006-2960. DOI: 10.1021/acs.biochem.5b00478. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4617767/ (visited on 09/09/2022).

[19] R. Mishra, M. Jayaraman, B. P. Roland, E. Landrum, T. Fullam, R. Kodali, A. K. Thakur, I. Arduini, and R. Wetzel. “Inhibiting the nucleation of amyloid structure in a huntingtin fragment by targeting α-helix-rich oligomeric intermediates”. eng. In: *J mol biol* 415.5 (Feb. 2012), pp. 900–917. ISSN: 1089-8638. DOI: 10.1016/j.jmb.2011.12.011.

[20] V. Brass, E. Bieck, R. Montserret, B. Wölk, J. A. Hellings, H. E. Blum, F. Penin, and D. Moradpour. “An amino-terminal amphipathic alpha-helix mediates membrane association of the hepatitis C virus nonstructural protein 5A”. eng. In: *J biol chem* 277.10 (Mar. 2002), pp. 8130–8139. ISSN: 0021-9258. DOI: 10.1074/jbc.M111289200.

[21] E. R. Georgieva, S. Xiao, P. P. Borbat, J. H. Freed, and D. Eliezer. “Tau binds to lipid membrane surfaces via short amphipathic helices located in its microtubule-binding repeats”. eng. In: *Biophys j* 107.6 (Sept. 2014), pp. 1441–1452. ISSN: 1542-0086. DOI: 10.1016/j.bpj.2014.07.046.

[22] H. Cui, E. Lyman, and G. A. Voth. “Mechanism of Membrane Curvature Sensing by Amphipathic Helix Containing Proteins”. In: *Biophys j* 100.5 (Mar. 2011), pp. 1271–1279. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2011.01.036. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3043213/ (visited on 09/09/2022).

[23] K. A. Burke, K. J. Kauffman, C. S. Umbaugh, S. L. Frey, and J. Legleiter. “The interaction of polyglutamine peptides with lipid membranes is regulated by flanking sequences associated with huntingtin”. eng. In: *J biol chem* 288.21 (May 2013), pp. 14993–15005. ISSN: 1083-351X. DOI: 10.1074/jbc.M112.446237.

[24] C. N. Kelly, C. E. Townsend, A. N. Jain, M. R. Naylor, C. R. Pye, J. Schwochert, and R. S. Lokey. “Geometrically Diverse Lariat Peptide Scaffolds Reveal an Untapped Chemical Space of High Membrane Permeability”. en. In: *J. am. chem. soc.* 143.2 (Jan. 2021), pp. 705–714. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.0c06115. URL: https://pubs.acs.org/doi/10.1021/jacs.0c06115 (visited on 03/09/2022).

[25] J. A. van Santen, G. Jacob, A. L. Singh, V. Aniebok, M. J. Balunas, D. Bunsko, F. C. Neto, L. Castaño-Espriu, C. Chang, T. N. Clark, J. L. Cleary Little, D. A. Delgadillo, P. C. Dorrestein, K. R. Duncan, J. M. Egan, M. M. Galey, F. J. Haeckl, A. Hua, A. H. Hughes, D. Iskakova, A. Khadilkar, J.-H. Lee, S. Lee, N. LeGrow, D. Y. Liu, J. M. Macho, C. S. McCaughey, M. H. Medema, R. P. Neupane, T. J. O’Donnell, J. S. Paula, L. M. Sanchez, A. F. Shaikh, S. Soldatou, B. R. Terlouw, T. A. Tran, M. Valentine, J. J. J. van der Hooft, D. A. Vo, M. Wang, D. Wilson, K. E. Zink, and R. G. Linington. “The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery”. en. In: *Acs cent. sci.* 5.11 (Nov. 2019), pp. 1824–1833. ISSN: 2374-7943, 2374-7951. DOI: 10.1021/acscentsci.9b00806.

URL: https://pubs.acs.org/doi/10.1021/acscentsci.9b00806 (visited on 04/01/2022).

[26] C. H. Chen, C. G. Starr, S. Guha, W. C. Wimley, M. B. Ulmschneider, and J. P. Ulmschneider. “Tuning of a membrane-perforating antimicrobial peptide to selectively target membranes of different lipid composition”. In: *The journal of membrane biology* 254.1 (Feb. 2021), pp. 75–96. ISSN: 1432-1424. DOI: 10.1007/s00232-021-00174-1.

[27] M. B. Ulmschneider, J. P. F. Doux, J. A. Killian, J. C. Smith, and J. P. Ulmschneider. “Mechanism and kinetics of peptide partitioning into membranes from all-atom simulations of thermostable peptides”. In: *Journal of the american chemical society* 132.10 (Mar. 17, 2010). Publisher: American Chemical Society, pp. 3452–3460. ISSN: 0002-7863. DOI: 10.1021/ja909347x. URL: https://doi.org/10.1021/ja909347x (visited on 09/16/2022).

[28] C. Neale, J. C. Y. Hsu, C. M. Yip, and R. Pomès. “Indolicidin binding induces thinning of a lipid bilayer”. In: *Biophysical journal* 106.8 (Apr. 15, 2014), pp. L29–31. ISSN: 1542-0086. DOI: 10.1016/j.bpj.2014.02.031.

[29] G. H. Zerze, C. M. Miller, D. Granata, and J. Mittal. “Free energy surface of an intrinsically disordered protein: comparison between temperature replica exchange molecular dynamics and bias-exchange metadynamics”. In: *Journal of chemical theory and computation* 11.6 (June 9, 2015). Publisher: American Chemical Society, pp. 2776–2782. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.5b00047. URL: https://doi.org/10.1021/acs.jctc.5b00047 (visited on 09/16/2022).

[30] S. Lee, R. Liang, G. A. Voth, and J. M. J. Swanson. “Computationally efficient multiscale reactive molecular dynamics to describe amino acid deprotonation in proteins”. In: *Journal of chemical theory and computation* 12.2 (Feb. 9, 2016). Publisher: American Chemical Society, pp. 879–891. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.5b01109. URL: https://doi.org/10.1021/acs.jctc.5b01109 (visited on 09/16/2022).

[31] C. T. Lee, J. Comer, C. Herndon, N. Leung, A. Pavlova, R. V. Swift, C. Tung, C. N. Rowley, R. E. Amaro, C. Chipot, Y. Wang, and J. C. Gumbart. “Simulation-Based Approaches for Determining Membrane Permeability of Small Compounds”. en. In: *J. chem. inf. model.* 56.4 (Apr. 2016), pp. 721–733. ISSN: 1549-9596, 1549-960X. DOI:

10.1021/acs.jcim.6b00022. URL: https://pubs.acs.org/doi/10.1021/acs.jcim.6b00022 (visited on 03/25/2022).

[32] R. Appadurai, J. Nagesh, and A. Srivastava. “High resolution ensemble description of metamorphic and intrinsically disordered proteins using an efficient hybrid parallel tempering scheme”. In: *Nature communications* 12.1 (Feb. 11, 2021). Number: 1 Publisher: Nature Publishing Group, p. 958. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21105-7. URL: https://www.nature.com/articles/s41467-021-21105-7 (visited on 09/16/2022).

[33] D. Vila-Viçosa, T. F. D. Silva, G. Slaybaugh, Y. K. Reshetnyak, O. A. Andreev, and M. Machuqueiro. “Membrane-induced p ka shifts in wt-pHLIP and its I16h variant”. In: *Journal of chemical theory and computation* 14.6 (June 12, 2018), pp. 3289–3297. ISSN: 1549-9626. DOI: 10.1021/acs.jctc.8b00102.

[34] H. Cui, E. Lyman, and G. A. Voth. “Mechanism of Membrane Curvature Sensing by Amphipathic Helix Containing Proteins”. In: *Biophys j* 100.5 (Mar. 2011), pp. 1271–1279. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2011.01.036. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3043213/ (visited on 09/09/2022).

[35] M. Chaibva, K. A. Burke, and J. Legleiter. “Curvature enhances binding and aggregation of huntingtin at lipid membranes”. eng. In: *Biochemistry* 53.14 (Apr. 2014), pp. 2355–2365. ISSN: 1520-4995. DOI: 10.1021/bi401619q.

2. MOLECULAR DYNAMICS (MD) SIMULATIONS

2.1 EQUILIBRIUM MOLECULAR DYNAMICS

Molecular Dynamics (MD) simulation is a method for determining the state of a system at time, t, given an initial position, $t = 0$. A given state is defined by the three-dimensional position and momentum vector for each particle in the system; therefore, for a system of size N, the evolution of an MD simulation is the repeated solving of an equation involving $6N$ parameters, given a number of constraints. Restraints used here are based on the topology of the molecules (bonds, angles, and dihedral angles) as well as other non-bonded interactions involving the molecules. For this work, atomistic MD simulations (every atom is represented by a hard-sphere particle) were performed. These atoms follow the laws of classical mechanics:

$$v_i(t + \delta t) = v_i(t) + f_i(t)\delta t/m_i \tag{2.1a}$$

$$r_i(t + \delta t) = r_i(t) + v_i(t + \delta t)\delta t \tag{2.1b}$$

where v is velocity, r is position, f is the force, and δt is the simulation time step. i represents the ith particle. Usually initial structures for simulation are taken from known X-ray or NMR structure, and initial momenta follow a Maxwell-Boltzmann distribution based on temperature. The total energy from interactions (V_{total} are modeled as:

$$V_{total} = V_{bond} + V_{angle} + V_{dihedral} + V_{improper} + V_{LJ} + V_{coulomb} \tag{2.2}$$

where V_{bond} is the energy of oscillation of two atoms about equilibrium bond length, V_{angle} is the energy of oscillation of three atoms about equilibrium bond angle, $V_{dihedral}$ is the energy for torsional rotation of four atoms, $V_{improper}$ is the energy for rotation about an improper dihedral angle, the energy contribution from van der Waals interactions based on the Lennard-Jones potential, and energy from coulombic interactions.

Bonded interactions (**Eqs. 2.3a-d**) apply to specific groups of atoms determined before the simulation begins. Non-bonded interactions (**Eqs. 2.3e and 2.3f**) are pairwise interactions defined for every pair of atoms in the system. Therefore, an N-particle system would have N(N-1) pairs for these interactions. Since these systems often have 50k - 100k atoms, this is a steep computational cost.

$$V_{bond}(r) = \frac{1}{2}k_b(r - r_0)^2 \tag{2.3a}$$

$$V_{angle}(\theta) = \frac{1}{2}k_\theta(\theta - \theta_0)^2 \tag{2.3b}$$

$$V_{dihedral}(\phi) = \frac{1}{2}k_\phi(\phi - \phi_0)^2 \tag{2.3c}$$

$$V_{improper}(\psi) = \frac{1}{2}k_\psi(\psi - \psi_0)^2 \tag{2.3d}$$

$$V_{LJ}(r) = \epsilon[(\frac{R}{r})^{12} - 2(\frac{R}{r})^6] \tag{2.3e}$$

$$V_{coulomb}(r) = \frac{q_1 q_2}{4\pi\epsilon_0 r} \tag{2.3f}$$

The primary method to reduce this cost is to cut off the range at which van der Waals (**Eq. 2.3e**) interactions are calculated; they decay to zero more rapidly than the coulombic interactions (**Fig. 2.1**). Longer range interactions are handled using the particle mesh Ewald (PME) method: interactions are sorted into short and long range terms, and the latter is Fourier transformed to accelerate convergence.

The parameters in the force field allow the system to evolve in time. As it does so, given sufficient time, the system will explore all possible conformations. This means that at long timescales the ensemble average will be equivalent to the time average. This represents the ergodic hypothesis which is fundamental to the molecular dynamics simulation method. Therefore it is absolutely necessary that the system be able to evolve for a sufficiently long time. Given that sufficiently timescale has been reached the ensemble average of A can be calculated as:

$$< A >= \frac{\sum_i A_i e^{-\beta E_i}}{\sum_i e^{-\beta E_i}} \tag{2.4}$$

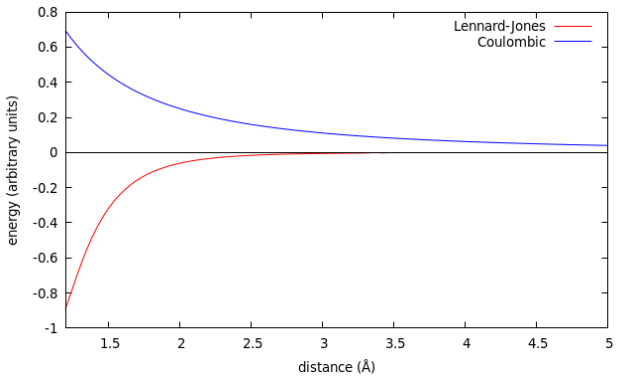

Figure 2.1: Comparison of the convergence of Coulombic and Lennard-Jones interactions. The energy contribution from Lennard-Jones decays to zero much more quickly than the energy associated with Coulombic interactions.

where β is the Boltzmann constant, V is the potential energy, and the subscript i represents individual conformations from the ensemble.

However, solving many biological problems require simulations of events that occur on timescales unattainable by most computers in a reasonable time. With some of the processes taking milliseconds or seconds, reaching a sufficient quantity of simulation time to represent these is unrealistic in MD simulation. The primary restriction on accessing these timescales is the timestep for calculating energies in the simulation. The proper timestep is set by the rate at which the bond between a hydrogen and a heavy atom fluctuate, and although this bond can be restrained to allow for a larger timestep, it remains on the order of femtoseconds which means millions of steps are necessary to attain even nanosecond timescales.

This ”brute force” method of MD simulations can be effective to solve some problems. However, To combat this, we utilize enhanced sampling methods to explore the phase space of the biological components in question. Chapter 3 of this work utilizes a method called constant pH to more efficiently sample protonation states by allowing them to vary during the simulation. In chapter 4, we employ a weighted ensemble method to encourage the exploration of the important phase space of a reaction coordinate. Finally, in chapter 5, gaussian-accelerated MD was used to more efficiently evolve possible conformations for the peptides.

2.2 CONSTANT pH MD

The constant pH method is a solution to effectively simulate pH in MD. Under normal equilibrium MD, bonds cannot be formed nor broken, therefore protonation states must be simulated individually. For example, this means that to fully simulate all possible combinations of states in the peptide in Fig. 2.2 64 individual simulations are necessary. Instead, under constant pH, the algorithm uses a pH value to determine energetically which sites are more likely to be protonated and is able to switch them on-the-fly in a single simulation.

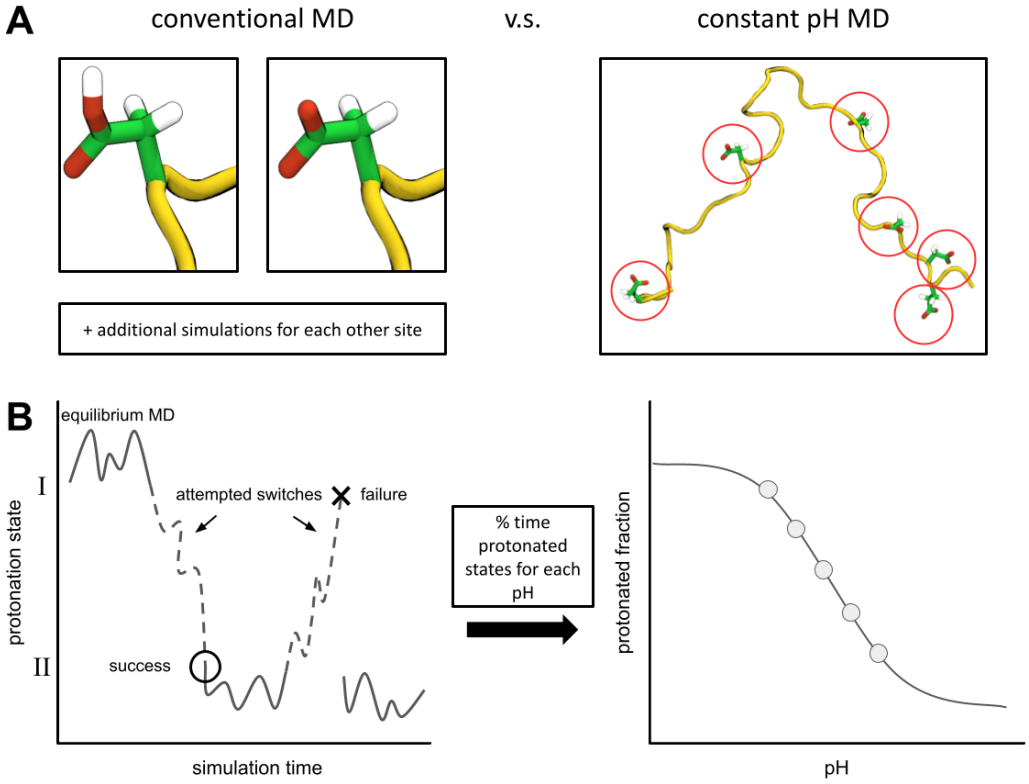

Figure 2.2: Description of the constant pH method. A) Simulating different protonation states with conventional MD requires a different simulation per protonation state while constant pH can simulate the system at a pH value in a single simulation. B) In the NAMD implementation of constant pH, the system goes through cycles of equilibrium MD and nonequilibrium/Monte Carlo during which switches in protonation are attempted. If the switch is accepted, the system will proceed with the new protonation state, otherwise it will remain in the current state. The %time protonated for each pH can be plotted to form a titration curve.

The constant pH implementation in NAMD[1] (the one used in this work) operates via cycles of equilibrium MD and nonequilibrium/Monte Carlo (noneq/MC). During equilibrium

MD, the system is able to adopt conformations based on the protonation state. As the noneq/MC portion of the cycle begins, the system attempts to switch the protonation state of each possible site. If the new state is energetically favorable, the simulation proceeds with the new protonation state. However, if the new state is not preferred then the system will continue with the original protonation state. As the constant pH cycles accrue, the protonation state of each site is tracked for that pH. This allows simulations from multiple pHs to plot a titration curve for each protonatable site based on the % time the site was protonated in each simulation.

2.3 THE WEIGHTED ENSEMBLE APPROACH

Utilizing a weighted ensemble (WE) in MD is most effective for fully exploring the phase space of a reaction coordinate. To do this, the phase space of the reaction coordinate must be broken into bin and a number of simulations or walkers to occupy those bins must be selected. Unlike many enhanced sampling methods that more efficiently sample a phase space, WE does so without energetically perturbing the system; instead, it allocates simulation time to evenly explore all bins. In the example in Fig. 2.3, the reaction coordinate is the distance between bilayer center and the peptide in efforts to simulate unbiased binding; two walkers per bin are represented. As the simulations evolve, one of them happens to have had the peptide move closer to the bilayer and into the second bin; here that means that WE will duplicate both walkers and iteration 3 begins with four walkers. In iteration 4, one of the walkers from the bin 2 happens to enter bin 1 over the simulation time, and therefore, since there are now three walkers in bin 1, WE trims one of the walkers to maintain two walkers per bin. This process continues until the desired number of iterations have been completed, and through this the phase space explored. This method can be effectively implemented via the WESTPA[2] as it is MD engine independent and comes with a host of helpful tools for viewing the progression through bins and analyzing the flux between bins.

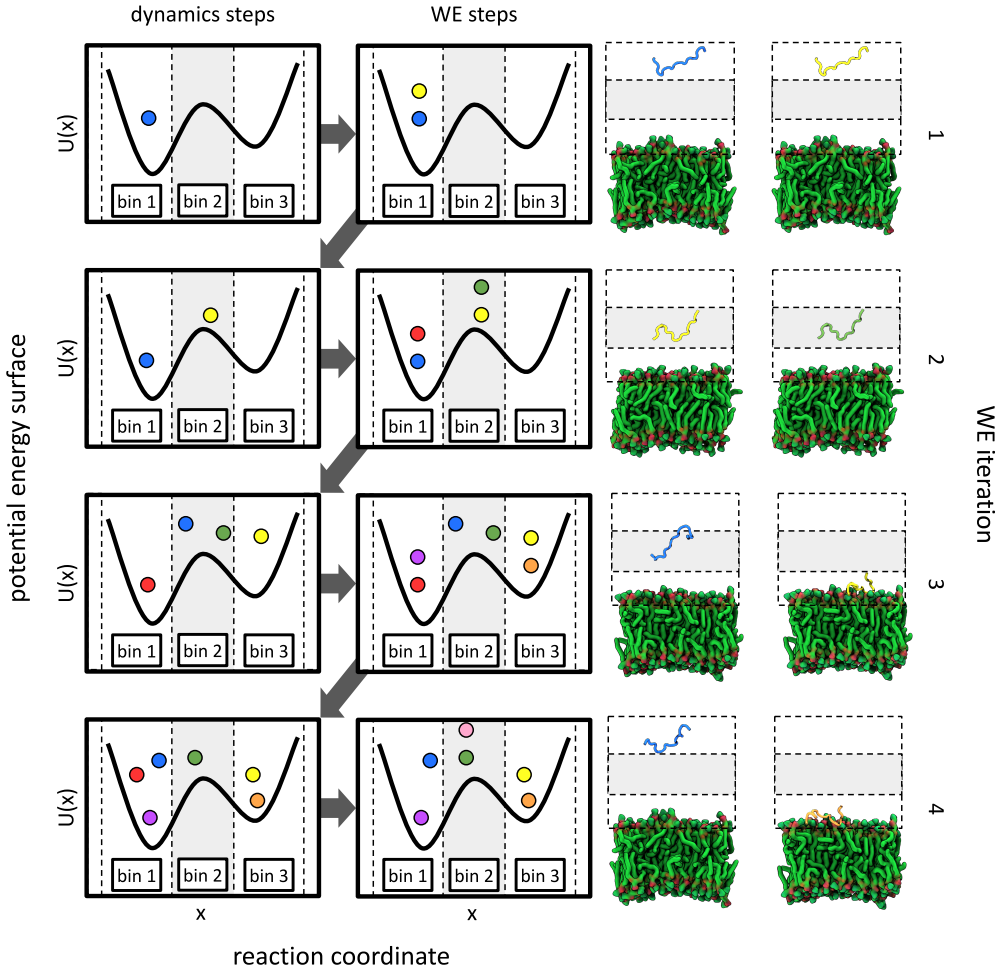

Figure 2.3: The process of the weighted ensemble approach. Walk-through of four iterations of a theoretical WE simulation. The images to the right of the plots represent the systems in the plot; color of the point in the plot correlates with color of peptide. Applying a weighted ensemble (WE) approach requires separating the reaction coordinate space into bins. During each WE step, simulations will be duplicated/trimmed to maintain the desired number of walkers per bin. In each MD step, the system is allowed to evolve unperturbed in equilibrium. This regime supports the full exploration of the reaction coordinate without energetically perturbing the system.

2.4 GAUSSIAN ACCELERATED MD

Gaussian accelerated MD[3, 4, 5, 6] (GaMD) is a method that adjusts the energy barriers between states so that higher energy states can be more easily accessed. This effectively enhances the conformational sampling of the subject. This is done by adding a harmonic boost potential (**Fig. 2.4**) to "fill in" the deepest wells of the potential. Consider a system

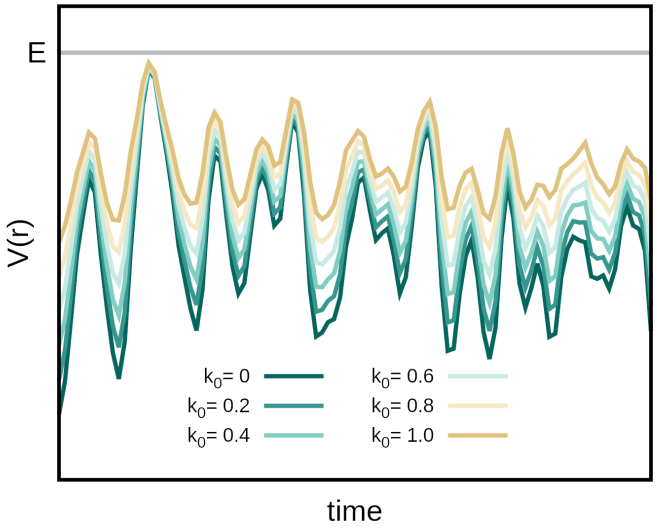

Figure 2.4: Gaussian accelerated MD boost potential.

with N atoms with positions $r \equiv \{\vec{r}_1, \cdots, \vec{r}_N\}$. When the potential of the system, $V(r)$, is less than a reference value, E, a modified potential, $V^*(r)$, is calculated as such:

$$V^*(r) = V(r) + \Delta V(r), \tag{2.5}$$

$$\Delta V(r) = \begin{cases} \frac{1}{2}k(E - V(r))^2, & \text{if } V(r) < E \\ 0, & \text{if } V(r) \geq E \end{cases} \tag{2.6}$$

where k is the harmonic force constant. E and k are both set automatically by the three rules endemic to the method. First, E must be set in the following range:

$$V_{max} \leq E \leq V_{min} + \frac{1}{k}, \tag{2.7}$$

where V_{max} and V_{min} are the maximum and minimum potential energies of the system respectively. A corollary to **Eq. 2.7** is that $k \leq \frac{1}{V_{max} - V_{min}}$. So if we define $k \equiv k_0 \cdot \frac{1}{V_{max} - V_{min}}$,

then $0 < k_0 \leq 1$. Finally, for proper energetic reweighting, the standard deviation of ΔV must have a sufficiently narrow distribution:

$$\sigma_{\Delta V} = k(E - V_{avg})\sigma_V \leq \sigma_0 \tag{2.8}$$

where V_{avg} and σ_0 are the average and standard deviation of the system potential energies, $\sigma_{\Delta V}$ is the standard deviation of ΔV with σ_0 as an upper limit specified by the user for proper reweighting. When E is set to the lower bound, $E = V_{max}$, k_0 is calculated as:

$$k_0 = min(1.0, k_0') = min(1.0, \frac{\sigma_0}{\sigma_V} \cdot \frac{V_{max} - V_{min}}{V_{max} - V_{avg}}). \tag{2.9}$$

However, when E is instead set to the upper bound, $E = V_{min} + \frac{1}{k}$, k_0 is calcualted as:

$$k_0 = k"_0 \equiv (1 - \frac{\sigma_0}{\sigma_V}) \cdot \frac{V_{max} - V_{min}}{V_{avg} - V_{min}} \tag{2.10}$$

if $k"_0$ is found between 0 and 1. Otherwise, k_0 is calculated using **Eq. 2.9**.

As can be seen, unlike other enhanced sampling methods that require the management of reaction coordinates, GaMD is very accessible in that there is little trial and error work to get the method running effectively. Also, because the boost potential is gaussian in nature, the potentials can be energetically reweighted leading to an accurate free energy landscape. GaMD is natively installed with both NAMD and AMBER.

2.5 BIBLIOGRAPHY

[1] B. K. Radak, C. Chipot, D. Suh, S. Jo, W. Jiang, J. C. Phillips, K. Schulten, and B. Roux. "Constant-pH Molecular Dynamics Simulations for Large Biomolecular Systems". In: *Journal of chemical theory and computation* 13.12 (2017-11-22), pp. 5933–5944. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.7b00875.

[2] M. C. Zwier, J. L. Adelman, J. W. Kaus, A. J. Pratt, K. F. Wong, N. B. Rego, E. Suárez, S. Lettieri, D. W. Wang, M. Grabe, D. M. Zuckerman, and L. T. Chong. "WESTPA: An Interoperable, Highly Scalable Software Package for Weighted Ensemble Simulation and Analysis". In: *J. chem. theory comput.* 11.2 (Feb. 2015). Publisher: American Chemical Society, pp. 800–809. ISSN: 1549-9618. DOI: 10.1021/ct5010615. URL: https://doi.org/10.1021/ct5010615 (visited on 06/23/2022).

[3] Y. T. Pang, Y. Miao, Y. Wang, and J. A. McCammon. “Gaussian accelerated molecular dynamics in NAMD”. In: *Journal of chemical theory and computation* 13.1 (Jan. 10, 2017), pp. 9–19. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.6b00931. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.6b00931 (visited on 09/14/2022).

[4] J. Wang, P. R. Arantes, A. Bhattarai, R. V. Hsu, S. Pawnikar, Y.-m. M. Huang, G. Palermo, and Y. Miao. “Gaussian accelerated molecular dynamics: principles and applications”. In: *WIREs computational molecular science* 11.5 (Sept. 2021). ISSN: 1759-0876, 1759-0884. DOI: 10.1002/wcms.1521. URL: https://onlinelibrary.wiley.com/doi/10.1002/wcms.1521 (visited on 09/14/2022).

[5] Y. Miao, V. A. Feher, and J. A. McCammon. “Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation”. In: *Journal of chemical theory and computation* 11.8 (Aug. 11, 2015), pp. 3584–3595. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.5b00436. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.5b00436 (visited on 09/14/2022).

[6] Y. Miao and J. A. McCammon. “Gaussian accelerated molecular dynamics: theory, implementation, and applications”. In: *Annual reports in computational chemistry* 13 (2017), pp. 231–278. ISSN: 1574-1400. DOI: 10.1016/bs.arcc.2017.06.005.

3. INTRAMOLECULAR INTERACTIONS PLAY A KEY ROLE IN STABILIZATION OF pHLIP AT ACIDIC CONDITIONS[1]

Membrane-active peptides (MAPs) are short-length peptides used for potential biomedical applications in diagnostic imaging of tissues, targeted drug delivery, gene delivery, and antimicrobials and antibiotics. The broad appeal of MAPs is that they are infinitely variable, relatively low cost, and biocompatible. However, experimentally characterizing the specific properties of a MAP or its many variants is a low-resolution and potentially time-consuming endeavor; molecular dynamics (MD) simulations have emerged as an invaluable tool in identifying the biophysical interactions that are fundamental to the function of MAPs. In this chapter, a step-by-step approach to discreetly model the binding, folding, and insertion of a membrane-active peptide to a model lipid bilayer using MD simulations is described. Detailed discussion is devoted to the critical aspects of running these types of simulations: prior knowledge of the system, understanding the strengths and weaknesses of molecular mechanics force fields, proper construction and equilibration of the system, realistically estimating both experimental and computational timescales, and leveraging analysis to make direct comparisons to experimental results as often as possible.

3.1 INTRODUCTION

Intrinsically-disordered proteins (IDPs) are ubiquitous in nature, most notably hypothesized to play a role in the onset of various neurodegenerative disorders such as Alzheimer's and

[1] This article/chapter was published in *Journal of Computational Chemistry*, **42**, Frazee, N., Mertz, B., Intramolecular interactions play key role in stabilization of pHLIP at acidic conditions, Copyright 2021 Wiley Periodicals LLC.

Parkinson's associated with formation of amyloid plaques [1] or to facilitate liquid-liquid phase separations involved with membraneless organelles [2]. The hallmark of IDPs is their ability to transition from an unstructured conformation as monomers to fibrillar complexes with well-defined secondary structure [3]. Aggregation of IDPs is partially understood; theoretical approaches have recently been developed to predict conformational samplinng of IDPs [4] as well as sequence-dependent propensity to drive liquid-liquid phase separations [5, 6]. However, comprehensive fundamental understanding of these phenomena still remains elusive. Membrane-active peptides (MAPs) are a subset of IDPs that are characterized by their ability to undergo partitioning-folding coupling at the membrane interface [7], leading to a broad range of potential applications (e.g., anitbiotics [8], targeted drug delivery [9], and gene therapy [10]).

The pH-Low Insertion Peptide (pHLIP) is a MAP that has shown much promise in biomedical applications such as diagnostic imaging of tumors [11] and targeted drug delivery of chemotherapeutics.[9] The key to the function of pHLIP is its acid-sensitivity. Under normal pH, pHLIP remains in a coiled conformation, diffusing through solution and freely associating/dissociating to and from the plasma membrane of cells. Upon diffusion into tissues with an acidic pH (e.g., the tumor microenvironment[12]), the acidic residues in pHLIP become protonated, triggering folding and unidirectional insertion into the plasma membrane. [13, 14] However, a key overlooked aspect of the biophysics of pHLIP is its behavior in solution. Both *in vitro* and *in vivo* studies of pHLIP typically use a concentration of 2 μM to prevent aggregation of the peptide in solution, a low concentration for effective use. Understanding of the behavior of pHLIP in solution could play a key role in enhancing its effectiveness in clinical applications.

What we do know of pHLIP's behavior in solution is that it forms a tetramer at concentrations higher than 8 μM, but is predominantly monomeric below that threshold.[14] Aggregation is driven by the large number of hydrophobic residues in the C-terminal half of pHLIP. Circular dichroism (CD) spectroscopy reveals little secondary structure formation in the peptide. However, it is possible for pHLIP to sample secondary structural conformations, as shown in a recent molecular dynamics (MD) study. [15] Even though aggregation is a critical step in the mechanism of many cell-penetrating peptides,[16] this mainly occurs at the membrane surface. In fact, higher concentrations of pHLIP tend to decrease its effectiveness at the membrane surface (the "parking problem" effect).[17] This points to a

need for a more detailed characterization of the behavior of pHLIP in solution.

With that in mind, we utilized constant pH molecular dynamics (CPHMD) simulations to probe the effect of changes in pH of bulk solvent on pHLIP. Previous experimental studies clearly demonstrated that the interstitial region of solid tumors is highly acidic (pH $<$ 5.5).[18] Based on our previous work, even a slight change in the protonation of pHLIP can shift the conformational distribution of the peptide;[15] a more accurate approach like CPHMD could provide invaluable insights extending our fundamental understanding of pHLIP. Constant pH approaches have been successfully applied to other complicated short peptides to identify differences in conformation based as a function of pH.[19, 20, 21, 22] In this work, we found that pHLIP is generally insensitive to changes in pH. In addition, intramolecular interactions and mutations can lead to subtle shifts in conformational sampling of the peptide. Finally, physiological salt concentrations play a role in desensitizing pHLIP to changes in pH.

3.2 METHODOLOGY

3.2.1 System setup

The structure for pHLIP was obtained from helix C of bacteriorhodopsin and mutated in VMD with psfgen[23] to correspond with pHLIP-4 used in Karabadzhak et al (AEQNPIYWARYAD-WLFTTPLLLLDLALLVDADEGT);[24] a second structure was mutated for the P20G systems. 20 trajectories each for pHLIP and the P20G variant were spawned and heated at 700 K for 5 ns. K-means clustering in LOOS[25] was performed on the trajectories; the top fifteen representative structures for each system were used as the starting positions after verifying they had proper ω dihedral angles. Each structure was solvated in VMD with a water box with 15 Å of padding on all sides. Ions were then added with the *ionize* plugin in VMD at two concentrations: 1) with just sodium counter ions present (0 mM) and 2) 150 mM NaCl. System sizes ranged from 35,000 to 40,000 atoms.

3.2.2 Simulations

All constant pH simulations were performed from pH 2.5 to 8.0 at 0.5 pH increments with the constant pH implementation[19] in NAMD 2.12[26] using the CHARMM36 forcefield.[27] Each of the four conditions (pHLIP, P20G:0 mM NaCl, 150 mM NaCl) and starting structure (15 for each condition) was run for 4000 non-equilibrium Monte Carlo/equilibrium MD cycles

with a τ_{switch} of 15 ps and an equilibrium time of 10 ps. There are 18 μs of simulation time for each of the four conditions. Additionally, for confirmation of the measured pK_a values, constant pH simulations using the same parameters and the same 15 starting structures for pHLIP with 0 mM and 150 mM NaCl were run for 500 non-equilibrium Monte Carlo/equilibrium MD cycles with an increased initial pK_a of 6 for each acidic residue (**Fig. 3.1**).

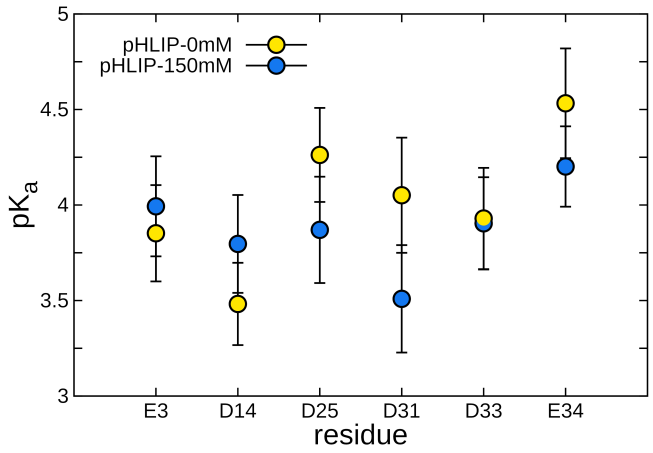

Figure 3.1: Effect of higher default pKa on determination of pKa's of pHLIP at 0 and 150 mM NaCl.

3.2.3 Analysis

Of the 4000 frames from each condition, pH, and starting structure, the last 3700 were used for analysis, as the decorrelation time was around 300 cycles for all conditions (**Fig. 3.2**). Plots were generated with gnuplot.[28] *Acidic residue titration.* The cphanalyze tool in pyNAMD[29] was used on all of the .cphlog files from the condition to compute the titration curves and pK_a. *Radius of gyration and end-to-end distance.*The *rad-gyr* tool in LOOS was employed to measure the radius of gyration for the peptide and the *interdist* tool in LOOS was used to measure the end-to-end distance. The two were correlated and binned for each condition. The representative structures were determined by finding each frame that occupied the most populated bin, determining the average structure of the frames, and finally finding the frame from the trajectories closest to the average structure. *Helicity.* The *stride* algorithm in VMD was used to determine the secondary structure of each residue in

each trajectory. *Salt bridges.* The *saltbr* plugin in VMD was used to measure salt bridges interactions with the default cutoff distance of 3.2 Å. *R11-D14 distance.* The *interdist* tool in loos was used to measure the distance between the terminal N atoms in R11 and terminal O atoms in D14 and separated based on the secondary structure measurements with *stride*. *Ion contacts.* The *contact-time* tool from LOOS was used to measure contacts between the the sodium ions and the terminal oxygen atoms of the acidic residues with a cutoff of 5 Å. *Solvent accessible surface area.* The *measure sasa* tool in VMD was used to measure the solvent accessible surface area of the selected residues. *Ramachandran plots.* The *ramachandran* tool in LOOS was used to measure the φ/ψangles of the peptide. *Native contacts.* The *native_contacts* tool in LOOS was used to measure the fraction of native contacts in the helical region (9-30) with respect to the crystal structure. *Hydrogen Bonds.* The *hbonds* plugin in VMD was used to measure hydrogen bonds with only the polar atoms and a cutoff of 5 Å.

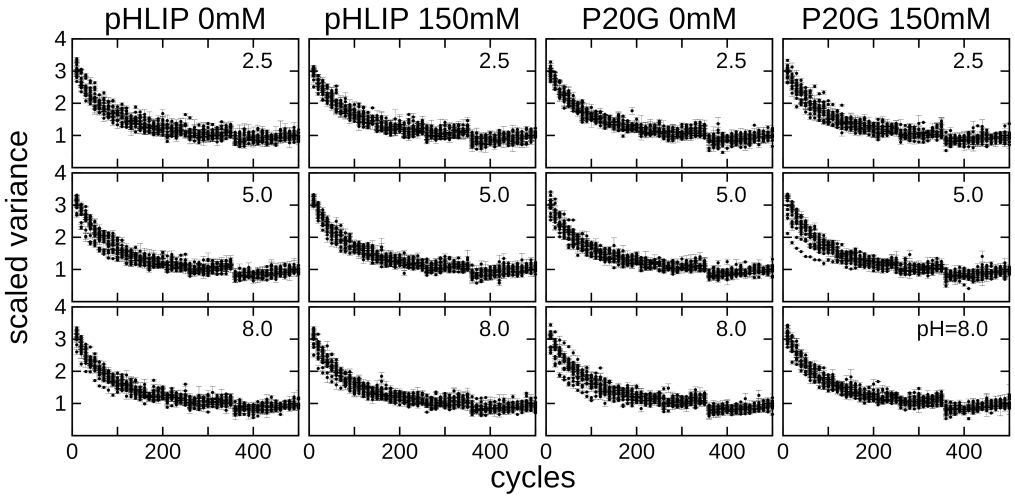

Figure 3.2: Decorrelation times as a function of number of MC/MD cycles in CPHMD simulations.

3.3 RESULTS

3.3.1 Constant pH MD effectively samples conformational space of pHLIP

Membrane-active peptides, like most IDPs, sample many conformations in solution. This is mainly due to partitioning-folding coupling [7], as pHLIP lacks a membrane to help stabilize

its folded state [30]. In the absence of a membrane, any formation of secondary structure is attributable to a balance between Coulomb repulsion of charged sidechains, hydrophobicity of the non-polar residues in pHLIP, and intramolecular hydrogen bond formation [31]. Initial analysis of our simulations show that pHLIP and P20G sample the majority of conformational space, regardless of condition (**Fig. 3.3**). If we compare our simulations to the fully-folded structure of pHLIP (based on the X-ray crystal structure of helix C from bacteriorhodopsin), we see that both pHLIP and P20G partially sample the native state (10-15 % of the peptide, **Fig. 3.4**). Looking at per-residue interactions, in general protonation of acidic residues (i.e., lower pH) tends to drive closer interactions in the interior of pHLIP and P20G (**Fig. 3.5**).

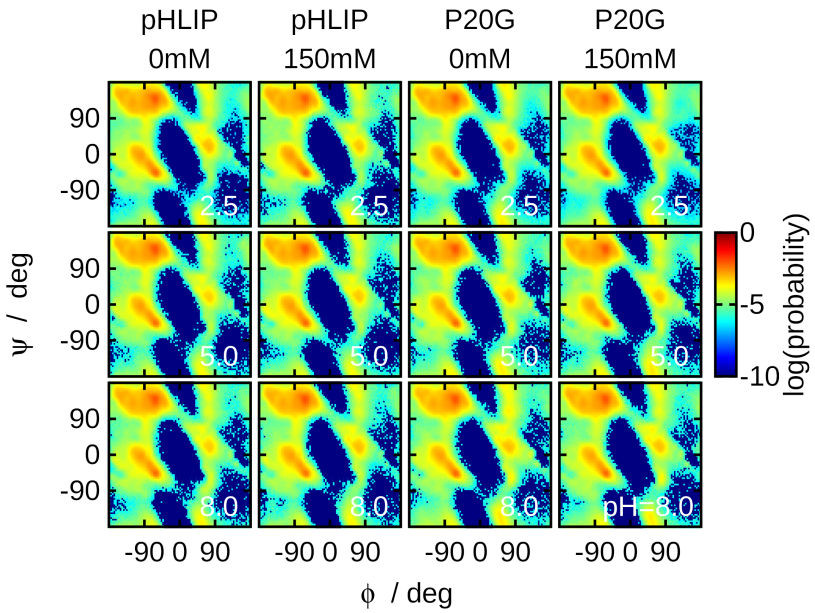

Figure 3.3: Distribution of backbone dihedrals for pHLIP and P20G as a function of pH and salt concentration.

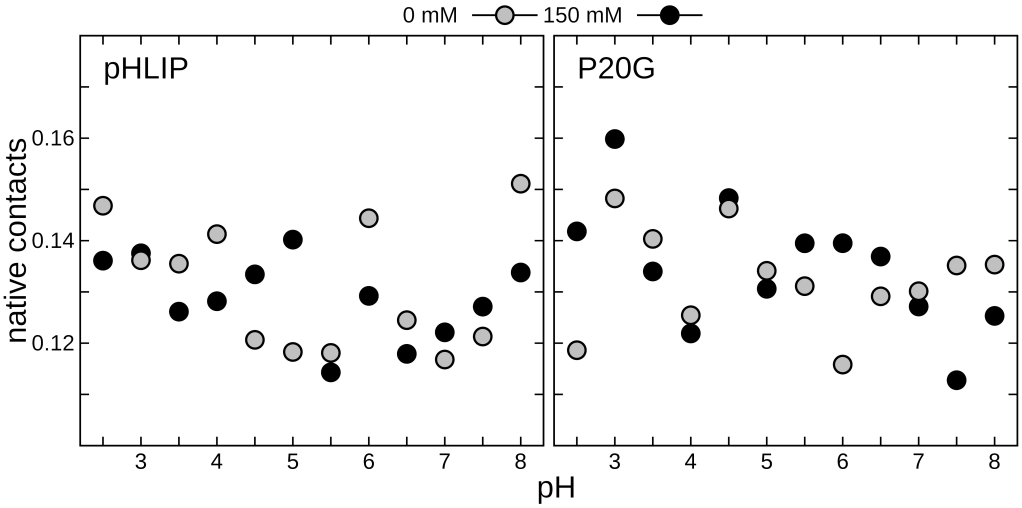

Figure 3.4: Sampling of native contacts of pHLIP and P20G as a function of pH and salt concentration.

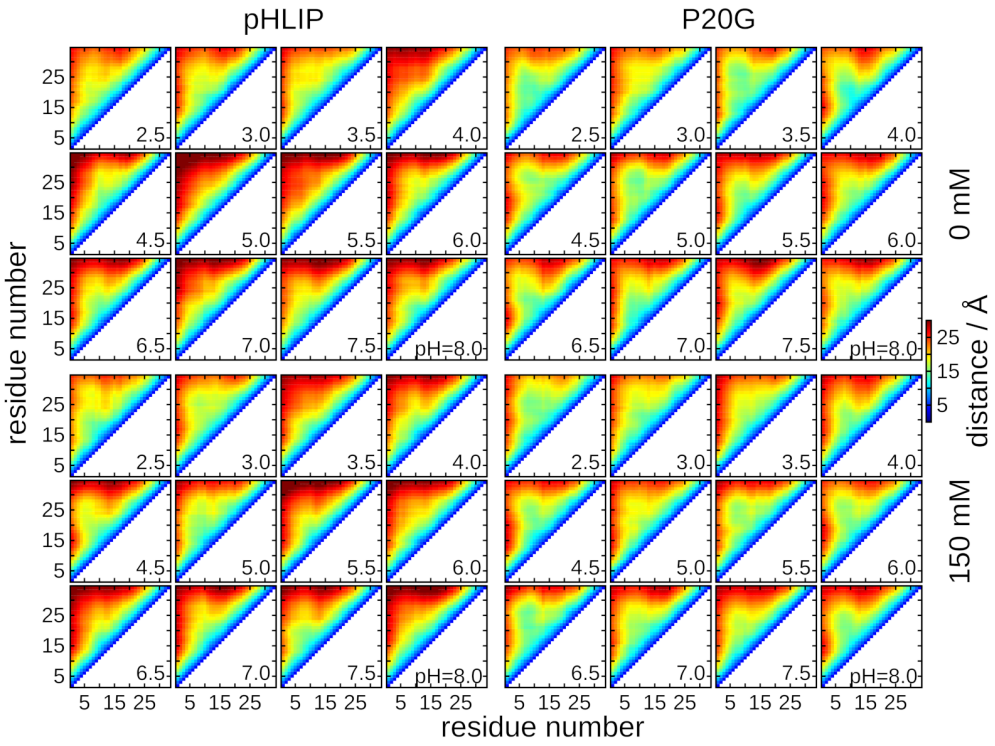

Figure 3.5: Acidic pH decreases intramolecular interactions in pHLIP. *Upper left:* residue-residue distances for pHLIP at 0 mM NaCl from pH 2.5 to 8.0. *Lower left:* residue-residue distances for pHLIP at 150 mM NaCl from pH 2.5 to 8.0. *Upper right:* residue-residue distances for P20G pHLIP at 0 mM NaCl from pH 2.5 to 8.0. *Lower right:* residue-residue distances for P20G pHLIP at 150 mM NaCl from pH 2.5 to 8.0.

3.3.2 Salt and sequence effects on the pKa of acidic residues

Our first goal was to determine the individual pK_as for acidic residues in pHLIP. Previously, we showed that each residue undergoes independent conformational sampling,[15] which should lead to distinct pK_as. In our CPHMD simulations, a clear difference exists between the aspartic acid (D14, D25, D31, and D33) and the glutamic acid (E3 and E34) residues, with pK_as comparable to a single amino acid in solution (D = 3.86, E = 4.25, **Figs. 3.6-3.9**). In addition, E3 and D14 have noticeably lower pK_as than the other glutamic and aspartic acids. This behavior is most likely due to their proximity to positively-charged residues (the N-terminus for E3 and R11 for D14), a phenomena that occurs in numerous protein systems such as the aspartic acid that acts as a proton acceptor in the microbial proton pump, bacteriorhodopsin [32].

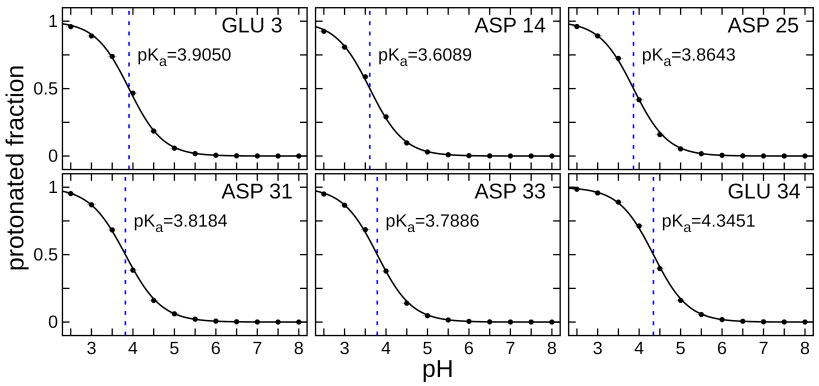

Figure 3.6: pKa for acidic residues in pHLIP at 0 mM NaCl.

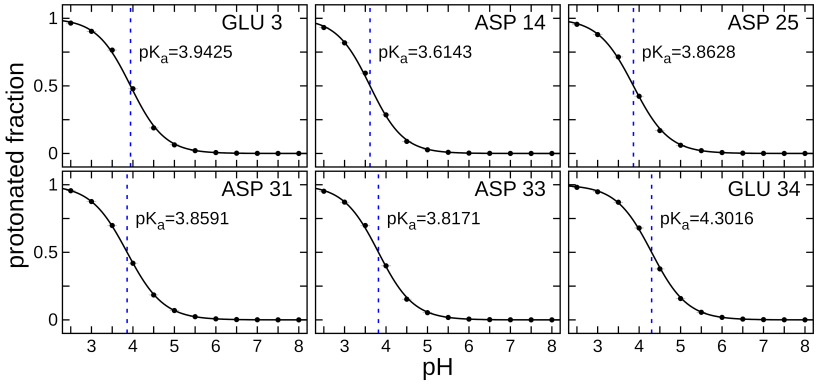

Figure 3.7: pKa for acidic residues in P20G at 0 mM NaCl.

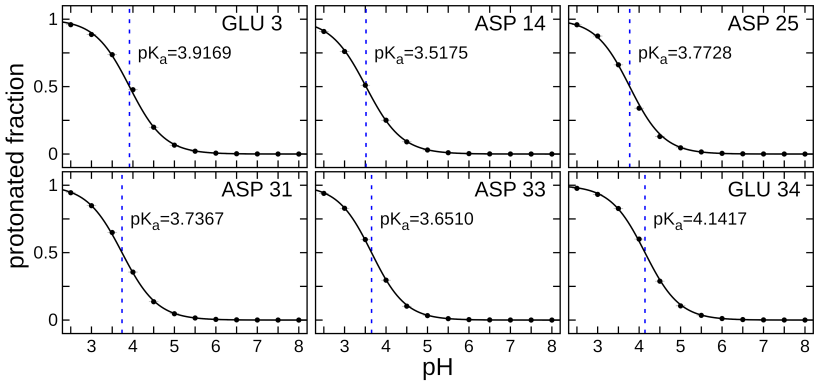

Figure 3.8: pKa for acidic residues in pHLIP at 150 mM NaCl.

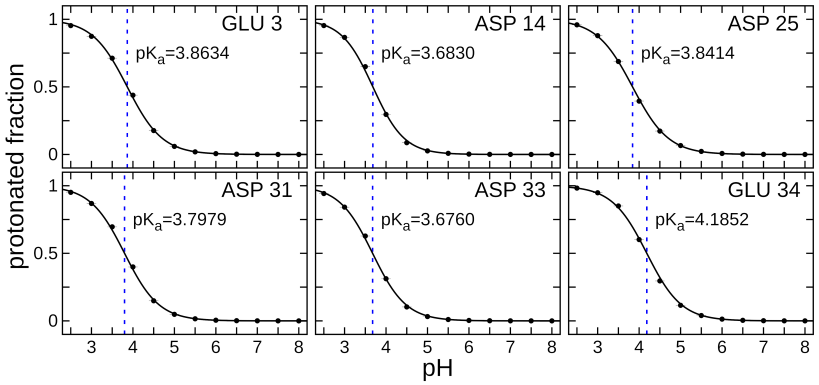

Figure 3.9: pKa for acidic residues in P20G at 150 mM NaCl.

Upon addition of 150 mM NaCl, several pK_as change for both pHLIP and P20G, often a decrease of 0.1-0.2 pH units (**Fig. 3.10**). This indicates that 1) higher salt concentrations have a weak screening of protons in solution and 2) salts could also contribute to greater solubility of pHLIP (i.e., less aggregation). (Physiological salt concentrations have been shown to decrease the pK_a of acidic residues in staphylococcal nuclease (SNase), a model protein that has pH-dependent behavior [33].) As mentioned above, D14 consistently has the lowest pK_a, with the most pronounced decrease for pHLIP at 150 mM NaCl. In addition to the proximity of R11, D14 is embedded among several aromatic residues (Y8, W9, Y12, W15, F17), which can provide local stabilization via π-π stacking and cation-π interactions (**Fig. 3.11**). This trend is consistent with solid-state NMR experiments on pHLIP in state II,[34] and could suggest that conformational sampling of pHLIP subtly influences the accessibility of D14 for titration both in solution and at the membrane interface.

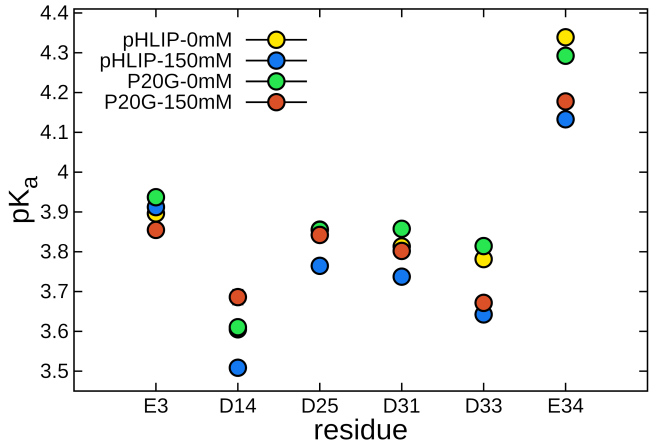

Figure 3.10: Determination of per-residue pK_as for acidic residues in pHLIP. Estimated pK_as via the cphanalyze tool.

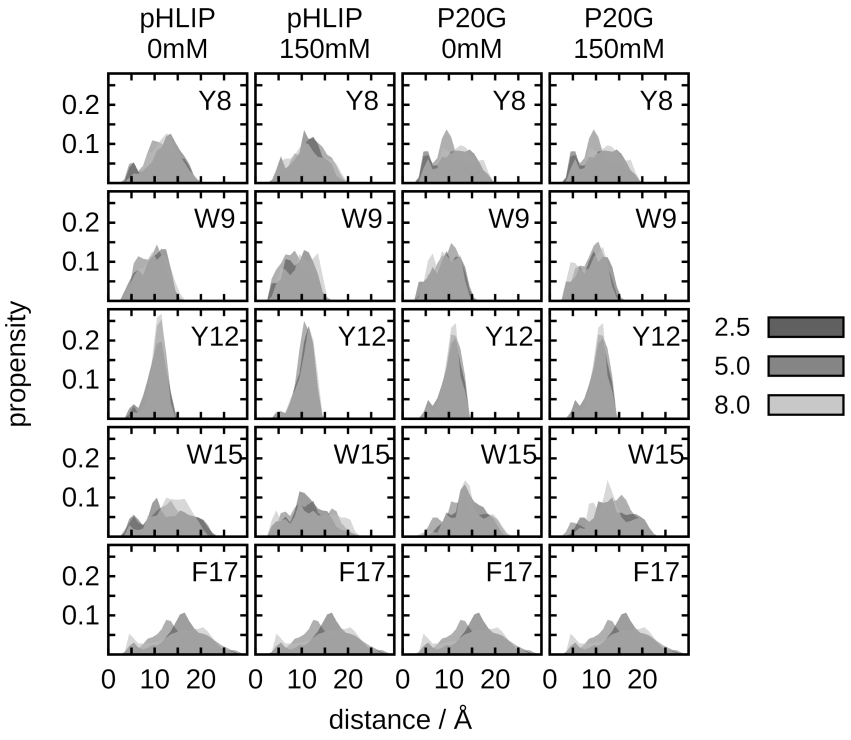

Figure 3.11: R11 can use cation-π interactions to stabilize the N-terminal half of pHLIP and P20G. Probability distribution of the distance between R11 and the aromatic residues of pHLIP and P20G as a function of pH.

Perturbations due to the P20G mutation and salt concentration also affect the overall conformation of pHLIP. In general, we observe a clear difference in behavior with respect to the radius of gyration (R_g) going from the fully protonated (pH 2.5) to the fully deprotonated state (pH 5.0) (**Fig. 3.12-3.13**). Addition of salt clearly drives pHLIP and P20G to sample smaller R_g values, with the most noticeable effect occurring at low pH. For pHLIP at no salt, we see a general decrease in free conformational sampling with an increase in pH (**Fig. 3.12**, first column). Upon adding 150 mM NaCl, pHLIP samples more compact conformations at low pH (**Fig. 3.12**, second column). P20G tends to adopt a small R_g, most likely because of the helical propensity imbued by the point mutation at position 20; this behavior is more pronounced at low pH and 150 mM NaCl (**Fig. 3.12**, third and fourth columns).

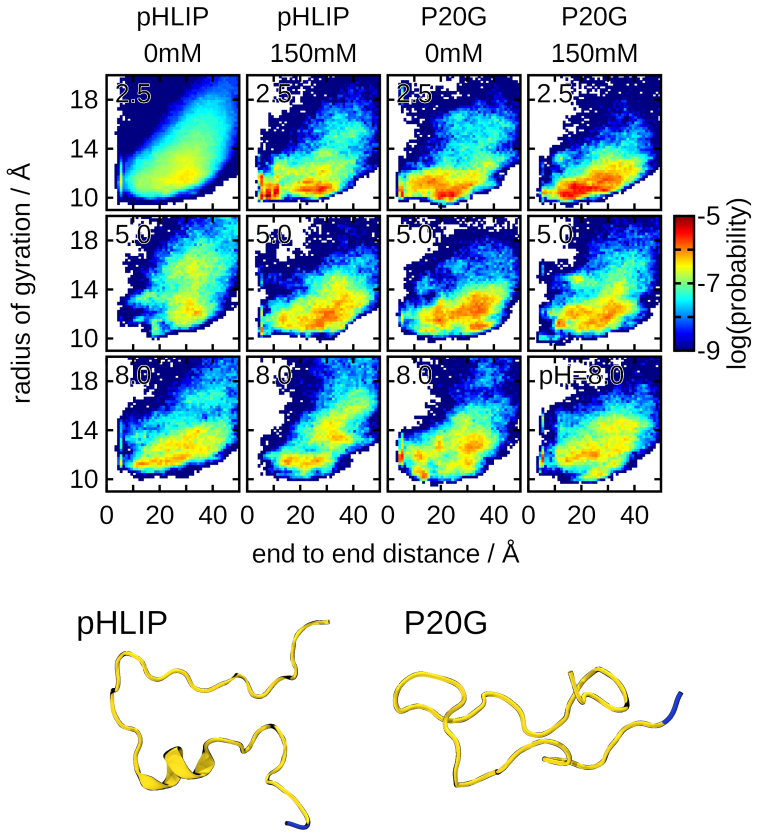

Figure 3.12: Physiological salt and acidic pH lead to more localized conformational sampling in pHLIP and P20G. Distribution of radius of gyration versus end-to-end distance of pHLIP and P20G as a function of salt and pH.

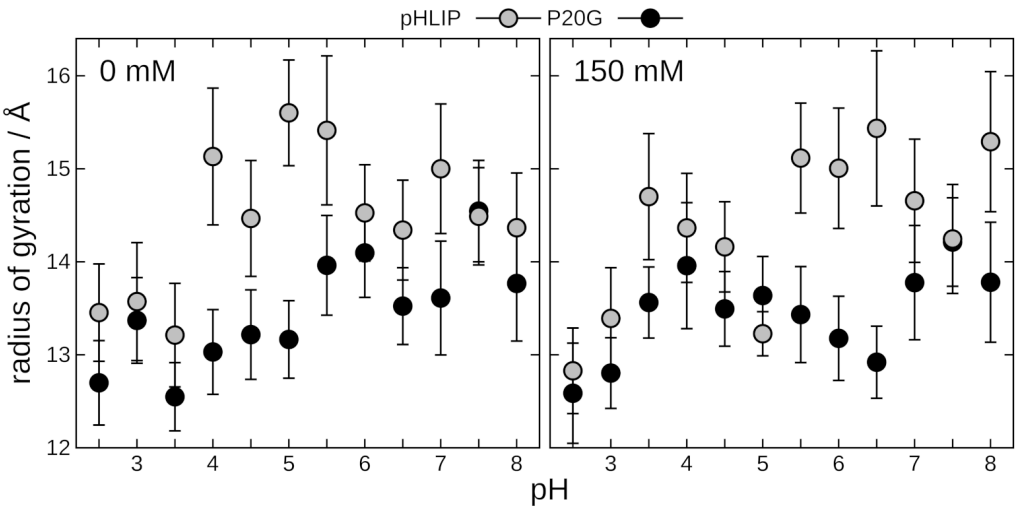

Figure 3.13: P20G possesses a more compact conformation. Measurement of the radius of gyration of the peptide as a function of pH and split by salt concentration. *Left:* radius of gyration for pHLIP and P20G as a function of pH with only counterions; *right:* radius of gyration for P20G as a function of pH with 150 mM NaCl.

3.3.3 Helicity can shift based on peptide composition and salt concentration

The main distinguishing characteristic between pHLIP and the P20G variant is helicity. When P20G was first tested experimentally, it adopted some α-helical character in solution,making it distinctive from pHLIP.[35] A more recent experimental study from the Thévenin and Ladokhin groups further refined our understanding of the differences in secondary structure between pHLIP and P20G in solution, finding that pHLIP and P20G possess 12% and 21% helicity, respectively.[36] This translates to 4 and 7 residues being involved in formation of a helical turn. Our simulations qualitatively agree with these experimental studies. We don't observe any noticeable differences in overall helicity between pHLIP and P20G at 0 mM NaCl, but at 150 mM NaCl, a general increase in helicity occurs for P20G (**Fig. 3.14**). Based on our knowledge of the crystal structure of bacteriorhodopsin, from which pHLIP is derived, and previous solid-state NMR studies [34], we know that two helical regions exist in pHLIP from residues 9 to 19 and from 21 to 30. On a per-residue basis, there are distinct behaviors between pHLIP and P20G and the two salt conditions. At 0 mM NaCl, pHLIP has helical segments in both regions, usually sampled between 5-10% of the total simulation (**Fig. 3.15A**) P20G also possesses some helical character in both regions, but has more pronounced helicity in the segment from residues 9-19. Upon adding 150 mM NaCl, a shift

occurs in the behavior of both peptides. pHLIP in general undergoes folding in the segment from residues 9-19 and is most pronounced at pH 2.5 (**Fig. 3.15B**). Likewise, for P20G, folding is slightly enhanced from residues 9-19, regardless of pH. Thus, although we don't observe any trends with respect to pH, it is clear that an increase in salt concentration to physiological levels does have an effect on folding of the N-terminal half of pHLIP and P20G.

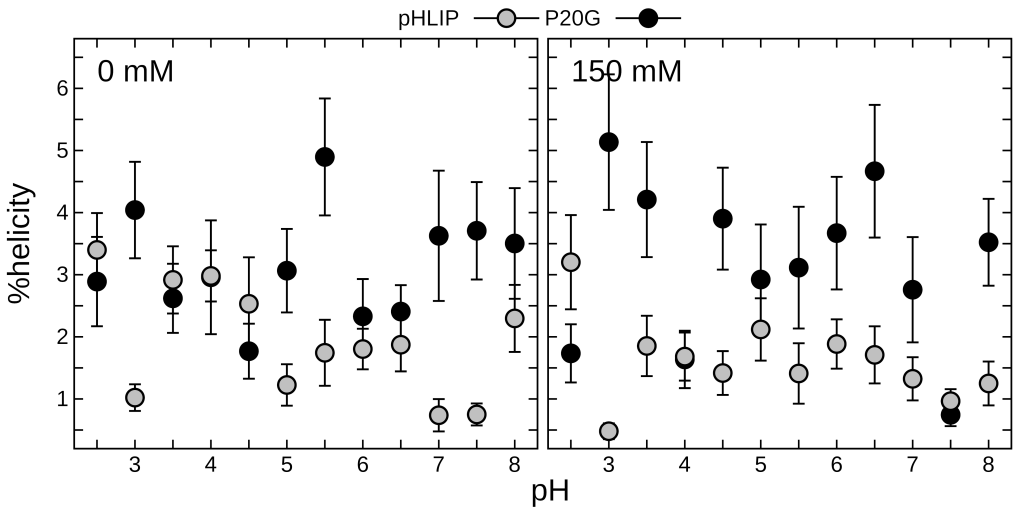

Figure 3.14: Overall helicity of pHLIP and P20G as a function of pH. *Left:* helicity of pHLIP and P20G at 0 mM NaCl. *Right:* helicity of pHLIP and P20G at 150 mM NaCl.

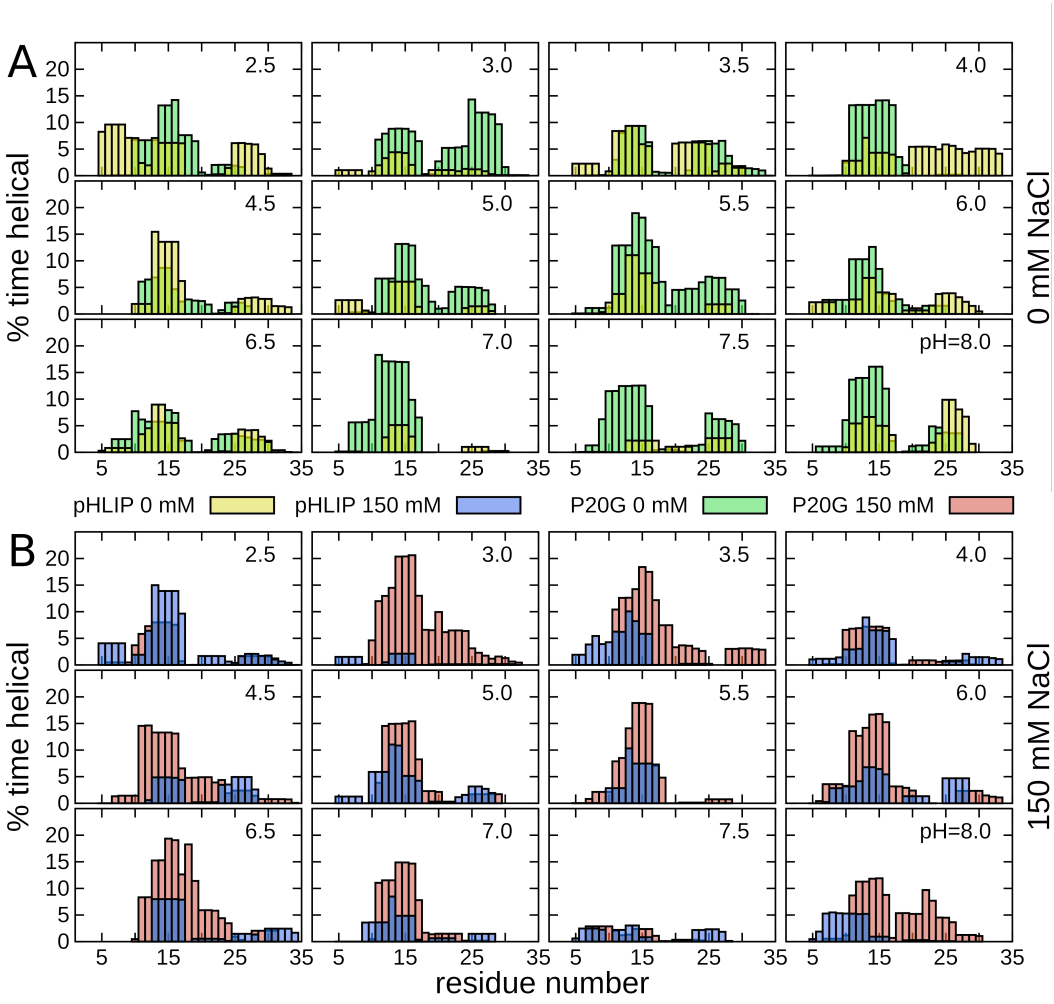

Figure 3.15: P20G mutation increases pHLIP helicity. (A) Per-residue helicity for wt-pHLIP (yellow) and P20G (green) with only counterions present. (B) Per-residue helicity for wt-pHLIP (blue) and P20G (red) with 150 mM NaCl.

3.3.4 Higher salt concentrations bias conformational selectivity of pHLIP

In general, salt bridge formation involving R11 increases with salt concentration mainly via D14 (**Fig. 3.16**). Several factors contribute to this behavior. The first is screening of specific residues in both pHLIP and P20G. Salt at physiological concentrations can screen acidic residues in pHLIP.[37] However, these interactions are non-uniform: E3, D14, and D25 have noticeably lower affinity for Na^+ cations (**Fig. 3.17**). This difference may be attributable to the proximity of clustered acidic residues at the C-terminus of pHLIP. This means that E3 and the interior aspartic acids (D14 and D25) are most likely to interact with R11. The R11-D14 interaction is more pronounced for pHLIP compared to P20G, which may be attributable to the fact that P20G has less conformational freedom due to its inherent helicity.

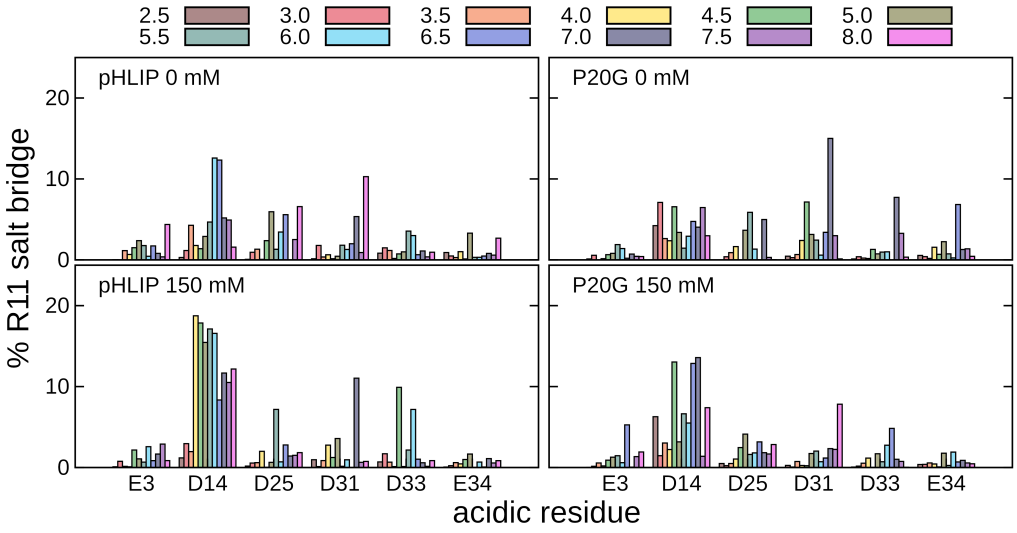

Figure 3.16: Salt bridge formation abrogated by competition with cations and conformational selectivity. Percentage of simulation time in which R11 forms a salt bridge (within 3.2 Å) with the corresponding acidic residue as a function of pH. The atoms used to define a salt bridge were the geometric centers of the terminal N atoms on R11 and the terminal O atoms on acidic residues

In almost every condition we tested, helicity has little effect on the interaction between R11 and D14. For pHLIP, a slight increase in shorter R11-D14 interactions occurs at higher pH, regardless of salt concentration (**Fig. 3.18**). In contrast, for P20G there is a clear separation of states at 0 mM NaCl – closer R11-D14 interactions occur in the helical state. This preference is lost at 150 mM NaCl, indicating that the dependence on helicity is very slight and can be easily perturbed.

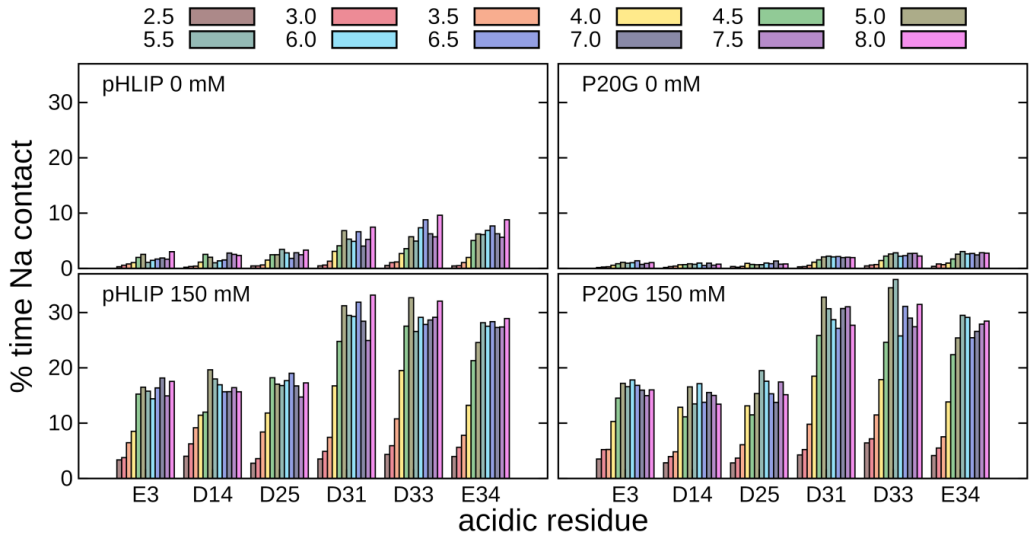

Figure 3.17: Physiological salt concentrations highlight lack of solvent accessibility of N-terminal residues in pHLIP and P20G. Percentage of simulation time at least one sodium cation is coordinated with acidic residues (within 5 Å).

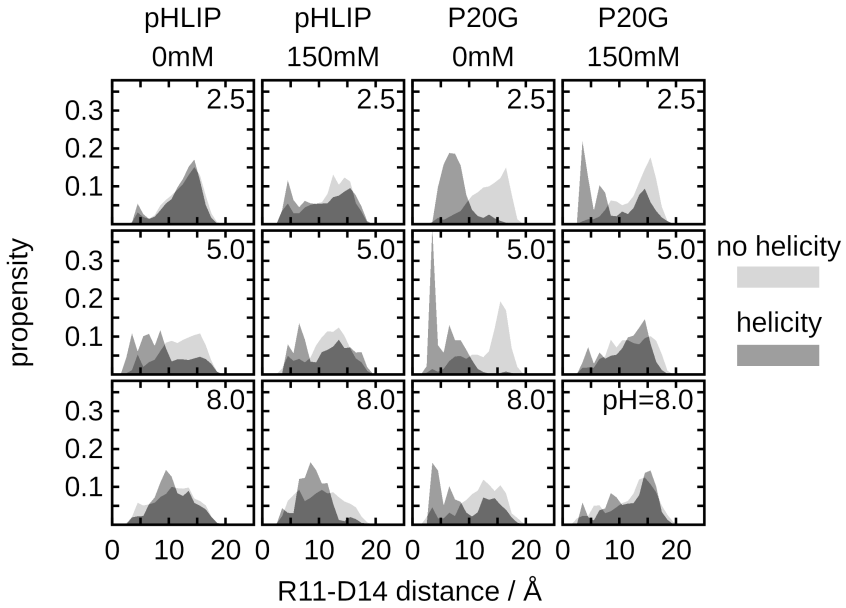

Figure 3.18: Effect of pH, salt, and helicity on the distance between R11 and D14 in pHLIP and P20G. Distribution of the R11-D14 distance when pHLIP or P20G has at least one helical turn (dark gray) versus when they have no helicity (light gray).

Another factor influencing the local environment of R11 is the proximity of aromatic residues. Cation-π interactions play an important role in stabilizing proteins,[38] and the

N-terminal half of pHLIP has several aromatic residues proximal to R11 (Y8, W9, Y12, W15, and F17). Although no single aromatic residue has preferential interactions with R11 in either pHLIP or P20G, both Y8 and W9 frequently interact with R11 at sub-5 Å distances, short enough to form a cation-π interaction (**Fig. 3.11**).

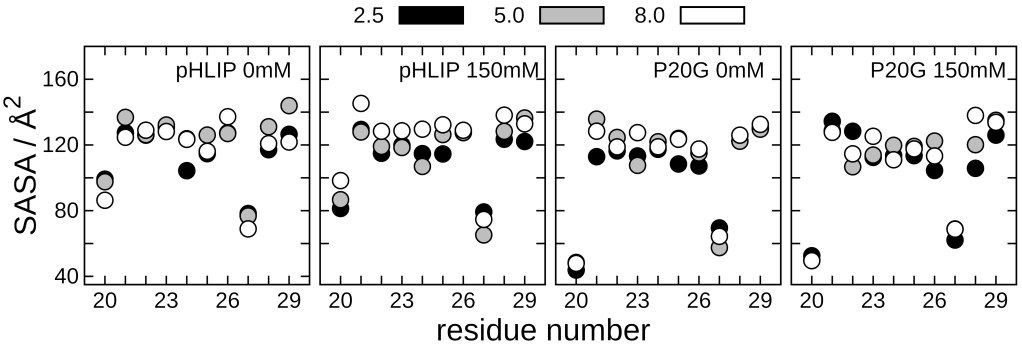

Figure 3.19: Solvent-accessible surface area of the hydrophobic stretch in pHLIP and P20G. Solvent accessible surface area (SASA) for each residue from position 20 to 29 (PLLLLDLALL) in pHLIP and P20G as a function of pH and salt concentration.

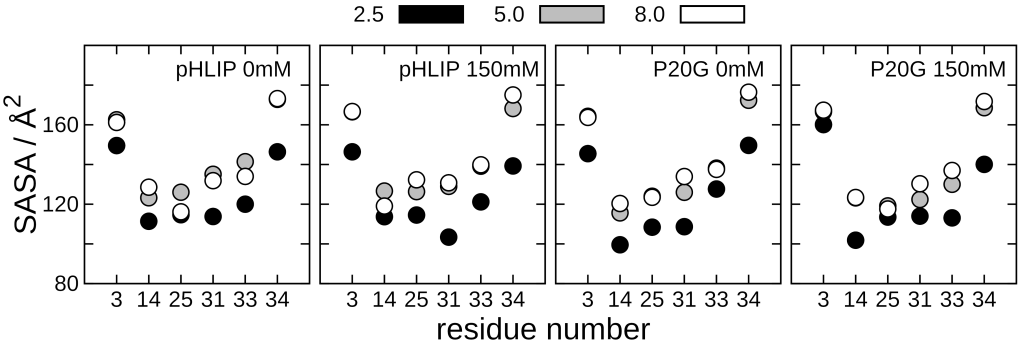

Figure 3.20: . Solvent-accessible surface area (SASA) for the acidic residues of pHLIP and P20G variant. Solvent-accessible surface area (SASA) for each acidic residue (E3, D14, D25, D31, D33, and E34) in pHLIP and P20G as a function of pH and salt concentration.

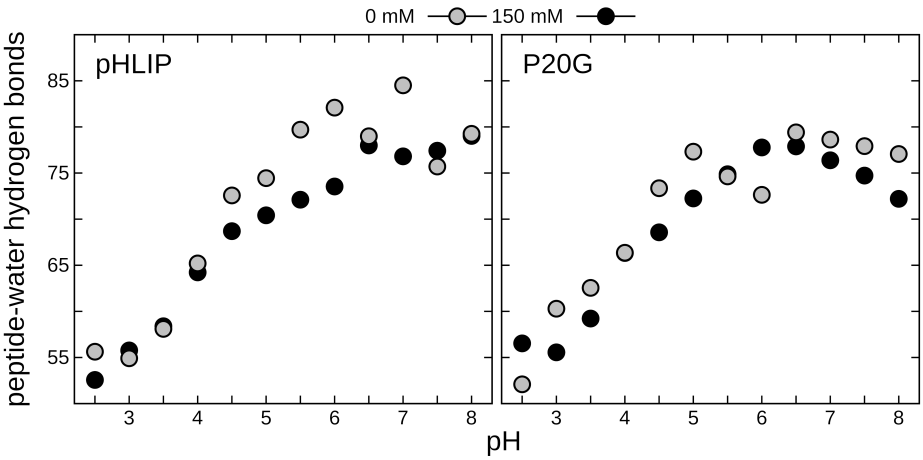

Figure 3.21: Peptide-water hydrogen bonds decrease with protonation of pHLIP and P20G. Average number of hydrogen bonds formed between pHLIP (left) and P20G (right) as a function of pH and salt concentration.

3.4 DISCUSSION

Constant pH MD simulations have been used the past two decades to gain greater understanding of the effect of pH on soluble proteins. One of the challenging aspects of CPHMD has been accurately determining the pK_a of interior residues, as both experimental [39, 40] and computational [41, 42] approaches have had varying degrees of success. SNase and ribonuclease A have been commonly used to improve our ability to accurately determine pK_a [22]. A key advance in CPHMD methods has been the ability to use explicit solvent [20, 21, 43, 19], which allows for more accurate treatment of localized conformational rearrangements upon change in titration of a protonatable site. This development has led to a rapid expansion in the systems that have been studied with CPHMD, including even membrane proteins.[44] It is clear based on our results that lower pH leads to an increase in helicity of pHLIP and P20G. In addition, lower pH depresses intramolecular interactions in the interior of pHLIP (i.e., R11, D14, and D25). Consistent with our previous study on pHLIP in state II,[37] physiological salt concentrations screen terminal acidic residues, which in turn can lead to increased interactions between interior charged residues (e.g. R11 with either D14 or D25). We realize that environmental factors contribute to conformational changes in pHLIP, but there are obviously other factors that come into play that have not been addressed in this study. For example, we know that administering pHLIP in combination with bovine serum

albumin allows for higher concentrations of pHLIP (150 µM) to be used in cell studies (Boyd, Popp, and Mertz, unpublished results).

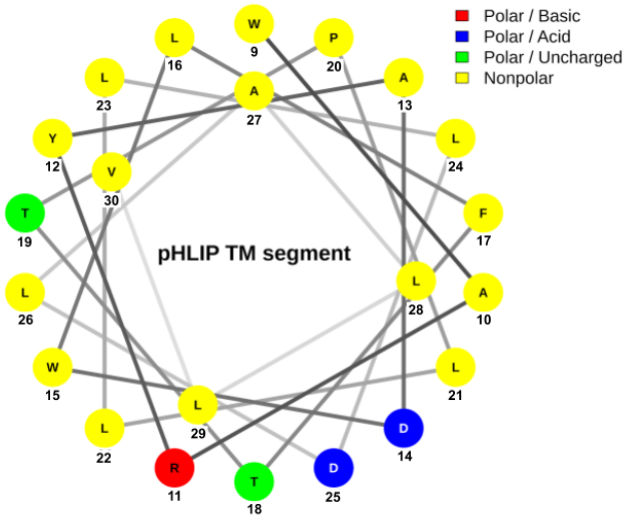

Figure 3.22: Transmembrane helical segment of pHLIP possesses an amphipathic surface. Helical wheel for the TM segment of wt-pHLIP (residues 9-30, as per Hanz et al., *Angew. Chem. Intl. Ed.* **2016** 55:12376-12381). Wheel plot was generated in the NetWheels server (doi.org/10.1101/416347).

Our simulations supply several insights into our biophysical understanding of pHLIP and their greater context in IDPs and protein folding. First, we can conclude that partitioning-folding coupling is the main driver for formation of helical conformations in pHLIP; even in highly acidic environments (i.e., pH 2.5), we do not observe a noticeable increase in helicity of pHLIP or P20G. Nevertheless, it was common in our simulations to observe transient formation of shorter helical turns, which is a necessary prerequisite for protein folding of soluble proteins and membrane-active peptides [45]. Second, pH can have a noticeable effect on intramolecular interactions in pHLIP, much like the anionic membrane-active peptide, EALA,[31] or the N-terminal fragment of exon 1 in huntingtin [46]. Charged amino acid residues, in this case a high density of anionic residues, repulse each other when deprotonated. It's only at low pH when these residues are titrated that pHLIP and P20G can adopt more compact structures (i.e., smaller R_g). This is consistent with the dependence of R_g and phase transition temperatures that leads to liquid-liquid phase separation (LLPS), a phenomena that has been linked to formation of membraneless organelles and aggregation of IDPs [6, 5, 47]. Even though pHLIP and P20G don't possess long repeats that are the

hallmark of IDPs associated with aggregation,[1] the presence of R11 and five aromatic residues in the N-terminal half of the peptide are a potential driver for phase transitions [48]. With the success of the Mittal and Pappu groups in applying coarse-grained and atomistic MD simulations of single-chain polymers to predict LLPS behavior,[6, 47] this opens the door for similar studies on membrane-active peptides like pHLIP. This type of fundamental knowledge equips us to introduce other potential modifications to the sequence of pHLIP. For example, substitution of an additional cationic residue (Arg or Lys) on the same face as D14 and D25 could make pHLIP more sensitive to changes in pH by lowering the energy barrier for helix formation (**Fig. 3.22**). By increasing the opportunity for formation of intramolecular salt bridges, it decreases the need to change the interior acidic residues (D14 and D25) topology via protonation to drive helical formation. This particular modification is very similar in nature to the dual salt bridges that give ribonuclease A a very sensitive response to changes in pH.[22] We have previously observed this behavior with R11-D14 salt bridges in state I[15] and state II.[37]

With respect to long-term development of applications with pHLIP, we anticipate that constant pH MD is another tool that can be used to probe the behavior of pHLIP variants, both in solution and at the membrane interface. As mentioned above, constant pH approaches are under continuous development, and careful consideration must be made with respect to the proper choice of a particular CPHMD implementation, as each have their strengths and weaknesses. CPHMD allows for more realistic modeling of how pHLIP will behave in an environment more consistent with the tumor microenvironment – this has mostly been missing up to this point, besides a recent study that focused on state III.[49] Unfortunately, optimization of pHLIP still remains largely unresolved – the biggest issues are determining the effect on peptide insertion due to conjugation with small molecules and properly accounting for local changes in the environment. For example, a recent study shows a drastic increase in pHLIP insertion when systems contain divalent cations.[50] A multi-faceted approach is necessary to truly address this problem.

3.5 CONCLUSIONS

In this study, we applied an advanced molecular dynamics approach to characterize the behavior of pHLIP in solution ranging from neutral pH to highly acidic pH. Our hypothesis, that more acidic environments would lead to greater sampling of folded conformations of

pHLIP, was generally not valid. However, our simulations provided a detailed view of how CPHMD could be utilized to screen pHLIP under different environmental conditions (salt concentration) and compositions (the P20G mutant) in solution.

MD simulations are capable of discerning slight differences in the behavior of pHLIP and variants of pHLIP. We could identify specific regions of the peptide that were more prone to undergo folding, based on local changes in the environment such as coordination with ions or close contacts with neighboring residues. This information is valuable from the perspective of acquiring fundamental understanding of how modifications to the local environment change the behavior of pHLIP. One of the long-standing challenges in developing variants of pHLIP that have enhanced properties (e.g., increased solubility, transport of cargo molecules) has been retention of acid-sensitivity and insertion into the plasma membrane. In many instances, these variants are incapable of being tested with lipid vesicles, due to the fact that they immediately aggregate in solution. By carrying out a two-step series of CPHMD simulations (monomeric and with multiple peptides in solution) we can quickly identify variants that can stay effectively solvated and discard those that tend to aggregate. This study represents the first step in that series; future studies will focus on applying CPHMD to systems with multiple peptides.

Even though pHLIP is not typically considered to be an IDP, the methodology and analysis used here can be applied to IDPs as well. Numerous examples exist in the literature that highlight the role of pH in the selection of conformational pathways in IDP aggregation [51]. CPHMD may prove to be an invaluable tool in providing molecular-level details about peptide interactions that are not readily apparent from experimental biophysical approaches.

3.6 BIBLIOGRAPHY

[1] P. H. Nguyen, A. Ramamoorthy, B. R. Sahoo, J. Zheng, P. Faller, J. E. Straub, L. Dominguez, J.-E. Shea, N. V. Dokholyan, A. De Simone, B. Ma, R. Nussinov, S. Najafi, S. T. Ngo, A. Loquet, M. Chiricotto, P. Ganguly, J. McCarty, M. S. Li, C. Hall, Y. Wang, Y. Miller, S. Melchionna, B. Habenstein, S. Timr, J. Chen, B. Hnath, B. Strodel, R. Kayed, S. Lesné, G. Wei, F. Sterpone, A. J. Doig, and P. Derreumaux. “Amyloid Oligomers: A Joint Experimental/Computational Perspective on Alzheimer’s Disease, Parkinson’s Disease, Type II Diabetes, and Amyotrophic Lateral Sclerosis”.

In: *Chemical reviews* 121.4 (Feb. 2021), pp. 2545–2647. ISSN: 0009-2665. DOI: 10.1021/acs.chemrev.0c01122.

[2] A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, and J. P. Taylor. “Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization”. en. In: *Cell* 163.1 (Sept. 2015), pp. 123–133. ISSN: 0092-8674. DOI: 10.1016/j.cell.2015.09.015.

[3] U. Ghosh, K. R. Thurber, W.-M. Yau, and R. Tycko. “Molecular structure of a prevalent amyloid-β fibril polymorph from Alzheimer’s disease brain tissue”. en. In: *Proceedings of the national academy of sciences* 118.4 (Jan. 2021). ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2023089118.

[4] L. Sawle and K. Ghosh. “A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins”. In: *The journal of chemical physics* 143.8 (Aug. 2015), p. 085101. ISSN: 0021-9606. DOI: 10.1063/1.4929391.

[5] J. Wang, J.-M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, I. Poser, R. V. Pappu, S. Alberti, and A. A. Hyman. “A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins”. English. In: *Cell* 174.3 (July 2018), 688–699.e16. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2018.06.006.

[6] G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal. “Relation between single-molecule properties and phase behavior of intrinsically disordered proteins”. en. In: *Proceedings of the national academy of sciences* 115.40 (Oct. 2018), pp. 9929–9934. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1804177115.

[7] S. H. White and W. C. Wimley. “Membrane protein folding and stability: Physical principles”. In: *Annual review of biophysics and biomolecular structure* 28.1 (1999), pp. 319–365.

[8] C. G. Starr, J. Ghimire, S. Guha, J. P. Hoffmann, Y. Wang, L. Sun, B. N. Landreneau, Z. D. Kolansky, I. M. Kilanowski-Doroh, M. C. Sammarco, L. A. Morici, and W. C. Wimley. “Synthetic molecular evolution of host cell-compatible, antimicrobial peptides effective against drug-resistant, biofilm-forming bacteria”. eng. In: *Proceedings of the national academy of sciences of the united states of america* 117.15 (Apr. 2020), pp. 8437–8448. ISSN: 1091-6490. DOI: 10.1073/pnas.1918427117.

[9] K. E. Burns and J. B. Delehanty. “Cellular delivery of doxorubicin mediated by disulfide reduction of a peptide-dendrimer bioconjugate”. eng. In: *International journal of pharmaceutics* 545.1-2 (July 2018), pp. 64–73. ISSN: 1873-3476. DOI: 10.1016/j.ijpharm.2018.04.027.

[10] N. A. Alhakamy, A. L. Alaofi, O. A. A. Ahmed, U. A. Fahmy, S. Md, W. H. Abdulaal, M. A. Alfaleh, A. Chakraborty, C. J. Berkland, and P. Dhar. “Development of lipid membrane based assays to accurately predict the transfection efficiency of cell-penetrating peptide-based gene nanoparticles”. eng. In: *International journal of pharmaceutics* 580 (Apr. 2020), p. 119221. ISSN: 1873-3476. DOI: 10.1016/j.ijpharm.2020.119221.

[11] P. Daumar, C. A. Wanger-Baumann, N. Pillarsetty, L. Fabrizio, S. D. Carlin, O. A. Andreev, Y. K. Reshetnyak, and J. S. Lewis. “Efficient 18F-Labeling of Large 37-Amino-Acid pHLIP Peptide Analogues and Their Biological Evaluation”. In: *Bioconjugate chemistry* 23.8 (Aug. 2012), pp. 1557–1566. ISSN: 1043-1802. DOI: 10.1021/bc3000222.

[12] M. G. V. Heiden, L. C. Cantley, and C. B. Thompson. “Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation”. en. In: *Science* 324.5930 (May 2009), pp. 1029–1033. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1160809.

[13] J. F. Hunt, P. Rath, K. J. Rothschild, and D. M. Engelman. “Spontaneous, pH-Dependent Membrane Insertion of a Transbilayer α-Helix”. In: *Biochemistry* 36.49 (Dec. 1997), pp. 15177–15192. ISSN: 0006-2960. DOI: 10.1021/bi970147b.

[14] Y. K. Reshetnyak, M. Segala, O. A. Andreev, and D. M. Engelman. “A Monomeric Membrane Peptide that Lives in Three Worlds: In Solution, Attached to, and Inserted across Lipid Bilayers”. English. In: *Biophysical journal* 93.7 (Jan. 2007), pp. 2363–2372. ISSN: 0006-3495. DOI: 10.1529/biophysj.107.109967.

[15] C. Gupta and B. Mertz. “Protonation Enhances the Inherent Helix-Forming Propensity of pHLIP”. In: *Acs omega* 2.11 (Nov. 2017), pp. 8536–8542. ISSN: 2470-1343. DOI: 10.1021/acsomega.7b01371.

[16] M.-A. Sani and F. Separovic. “How Membrane-Active Peptides Get into Lipid Membranes”. eng. In: *Accounts of chemical research* 49.6 (June 2016), pp. 1130–1138. ISSN: 1520-4898. DOI: 10.1021/acs.accounts.6b00074.

[17] Y. K. Reshetnyak, O. A. Andreev, M. Segala, V. S. Markin, and D. M. Engelman. “Energetics of peptide (pHLIP) binding to and folding across a lipid bilayer membrane”. en. In: *Proceedings of the national academy of sciences* 105.40 (Oct. 2008), pp. 15340–15345. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0804746105.

[18] X. Luo, H. Yang, H. Wang, Z. Ye, Z. Zhou, L. Gu, J. Chen, Y. Xiao, X. Liang, X. Qian, and Y. Yang. “Highly Sensitive Hill-Type Small-Molecule pH Probe That Recognizes the Reversed pH Gradient of Cancer Cells”. In: *Analytical chemistry* 90.9 (May 2018), pp. 5803–5809. ISSN: 0003-2700. DOI: 10.1021/acs.analchem.8b00218.

[19] B. K. Radak, C. Chipot, D. Suh, S. Jo, W. Jiang, J. C. Phillips, K. Schulten, and B. Roux. “Constant-pH Molecular Dynamics Simulations for Large Biomolecular Systems”. In: *Journal of chemical theory and computation* 13.12 (2017-11-22), pp. 5933–5944. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.7b00875.

[20] J. A. Wallace and J. K. Shen. “Continuous Constant pH Molecular Dynamics in Explicit Solvent with pH-Based Replica Exchange”. In: *Journal of chemical theory and computation* 7.8 (Aug. 2011), pp. 2617–2629. ISSN: 1549-9618. DOI: 10.1021/ct200146j.

[21] J. M. Swails and A. E. Roitberg. “Enhancing Conformation and Protonation State Sampling of Hen Egg White Lysozyme Using pH Replica Exchange Molecular Dynamics”. In: *Journal of chemical theory and computation* 8.11 (Nov. 2012), pp. 4393–4404. ISSN: 1549-9618. DOI: 10.1021/ct300512h.

[22] J. Khandogin, J. Chen, and C. L. Brooks. “Exploring atomistic details of pH-dependent peptide folding”. en. In: *Proceedings of the national academy of sciences* 103.49 (Dec. 2006), pp. 18546–18550. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0605216103.

[23] W. Humphrey, A. Dalke, and K. Schulten. “VMD: Visual molecular dynamics”. In: *Journal of molecular graphics* 14.1 (1996-02), pp. 33–38. ISSN: 02637855. DOI: 10.1016/0263-7855(96)00018-5. URL: http://linkinghub.elsevier.com/retrieve/pii/0263785596000185 (visited on 11/12/2018).

[24] A. G. Karabadzhak, D. Weerakkody, D. Wijesinghe, M. S. Thakur, D. M. Engelman, O. A. Andreev, V. S. Markin, and Y. K. Reshetnyak. “Modulation of the pHLIP Transmembrane Helix Insertion Pathway”. In: *Biophysical journal* 102.8 (2012-04-18),

pp. 1846–1855. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2012.03.021. pmid: 22768940. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3328699/ (visited on 11/12/2018).

[25] T. D. Romo, N. Leioatts, and A. Grossfield. “Lightweight object oriented structure analysis: Tools for building tools to analyze molecular dynamics simulations”. en. In: *Journal of computational chemistry* 35.32 (Dec. 2014), pp. 2305–2318. ISSN: 1096-987X. DOI: 10.1002/jcc.23753.

[26] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. “Scalable molecular dynamics with NAMD”. In: *Journal of computational chemistry* 26.16 (2005-12-01), pp. 1781–1802. ISSN: 1096-987X. DOI: 10.1002/jcc.20289. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20289 (visited on 11/05/2018).

[27] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell. “Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 dihedral angles”. In: *Journal of chemical theory and computation* 8.9 (2012-09-11), pp. 3257–3273. ISSN: 1549-9618. DOI: 10.1021/ct300400x. pmid: 23341755. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3549273/ (visited on 11/12/2018).

[28] J. Racine. “Gnuplot 4.0: a portable interactive plotting utility”. In: *Journal of applied econometrics* 21.1 (2006-01), pp. 133–141. ISSN: 0883-7252, 1099-1255. DOI: 10.1002/jae.885. URL: http://doi.wiley.com/10.1002/jae.885 (visited on 11/13/2018).

[29] B. Radak. *Python Tools for NAMD.* 2018-09-11. URL: https://github.com/radakb/pynamd (visited on 11/29/2018).

[30] C. Gupta, Y. Ren, and B. Mertz. “Cooperative Nonbonded Forces Control Membrane Binding of the pH-Low Insertion Peptide pHLIP”. English. In: *Biophysical journal* 115.12 (Dec. 2018), pp. 2403–2412. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2018.11.002.

[31] C. Dalgicdir, C. Globisch, C. Peter, and M. Sayar. “Tipping the Scale from Disorder to Alpha-helix: Folding of Amphiphilic Peptides in the Presence of Macroscopic and

Molecular Interfaces". en. In: *Plos computational biology* 11.8 (Aug. 2015), e1004328. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004328.

[32] R. Jonas and T. G. Ebrey. "Binding of a single divalent cation directly correlates with the blue-to-purple transition in bacteriorhodopsin". en. In: *Proceedings of the national academy of sciences* 88.1 (Jan. 1991), pp. 149–153. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.88.1.149.

[33] C. A. Castañeda, C. A. Fitch, A. Majumdar, V. Khangulov, J. L. Schlessman, and B. E. García-Moreno. "Molecular determinants of the pKa values of Asp and Glu residues in staphylococcal nuclease". en. In: *Proteins: structure, function, and bioinformatics* 77.3 (2009), pp. 570–588. ISSN: 1097-0134. DOI: 10.1002/prot.22470.

[34] S. Z. Hanz, N. S. Shu, J. Qian, N. Christman, P. Kranz, M. An, C. Grewer, and W. Qiang. "Protonation-Driven Membrane Insertion of a pH-Low Insertion Peptide". In: *Angewandte chemie international edition* 55.40 (2016-09-26), pp. 12376–12381. ISSN: 1521-3773. DOI: 10.1002/anie.201605203. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201605203 (visited on 08/14/2018).

[35] F. N. Barrera, J. Fendos, and D. M. Engelman. "Membrane physical properties influence transmembrane helix formation". en. In: *Proceedings of the national academy of sciences* 109.36 (Sept. 2012), pp. 14422–14427. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1212665109.

[36] V. Vasquez-Montes, J. Gerhart, K. E. King, D. Thévenin, and A. S. Ladokhin. "Comparison of lipid-dependent bilayer insertion of pHLIP and its P20G variant". In: *Biochimica et biophysica acta (bba) - biomembranes* 1860.2 (2018-02-01), pp. 534–543. ISSN: 0005-2736. DOI: 10.1016/j.bbamem.2017.11.006. URL: http://www.sciencedirect.com/science/article/pii/S0005273617303577 (visited on 11/14/2018).

[37] J. Westerfield, C. Gupta, H. L. Scott, Y. Ye, A. Cameron, B. Mertz, and F. N. Barrera. "Ions Modulate Key Interactions between pHLIP and Lipid Membranes". In: *Biophysical journal* 117.5 (Sept. 2019), pp. 920–929. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2019.07.034.

[38] J. P. Gallivan and D. A. Dougherty. “Cation-pi interactions in structural biology”. In: *Proceedings of the national academy of sciences* 96.17 (Aug. 1999), pp. 9459–9464. DOI: 10.1073/pnas.96.17.9459.

[39] D. G. Isom, C. A. Castañeda, B. R. Cannon, P. D. Velu, and B. G.-M. E. “Charges in the hydrophobic interior of proteins”. en. In: *Proceedings of the national academy of sciences* 107.37 (Sept. 2010), pp. 16096–16100. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1004213107.

[40] D. G. Isom, C. A. Castañeda, B. R. Cannon, and B. G.-M. E. “Large shifts in pKa values of lysine residues buried inside a protein”. en. In: *Proceedings of the national academy of sciences* 108.13 (Mar. 2011), pp. 5260–5265. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1010750108.

[41] M. Machuqueiro and A. M. Baptista. “Constant-pH Molecular Dynamics with Ionic Strength Effects: Protonation-Conformation Coupling in Decalysine”. In: *The journal of physical chemistry b* 110.6 (Feb. 2006), pp. 2927–2933. ISSN: 1520-6106. DOI: 10.1021/jp056456q.

[42] G. B. Goh, E. N. Laricheva, and C. L. Brooks. “Uncovering pH-Dependent Transient States of Proteins with Buried Ionizable Residues”. In: *Journal of the american chemical society* 136.24 (June 2014), pp. 8496–8499. ISSN: 0002-7863. DOI: 10.1021/ja5012564.

[43] Z. Yue and J. Shen. “pH-Dependent cooperativity and existence of a dry molten globule in the folding of a miniprotein BBL”. en. In: *Physical chemistry chemical physics* 20.5 (Feb. 2018), pp. 3523–3530. ISSN: 1463-9084. DOI: 10.1039/C7CP08296G.

[44] J. A. Henderson, Y. Huang, O. Beckstein, and J. Shen. “Alternative proton-binding site and long-distance coupling in Escherichia coli sodium–proton antiporter NhaA”. en. In: *Proceedings of the national academy of sciences* 117.41 (Oct. 2020), pp. 25517–25522. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2005467117.

[45] V. A. Voelz, M. S. Shell, and K. A. Dill. “Predicting Peptide Structures in Native Proteins from Physical Simulations of Fragments”. In: *Plos comput biol* 5.2 (Feb. 2009), e1000281. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1000281.

[46] Z.-M. Wang and H. A. Lashuel. “Discovery of a Novel Aggregation Domain in the Huntingtin Protein: Implications for the Mechanisms of Htt Aggregation and Toxicity”. In: *Angewandte chemie international edition* 52.2 (2013), pp. 562–567. ISSN: 1521-3773. DOI: 10.1002/anie.201206561.

[47] X. Zeng, A. S. Holehouse, A. Chilkoti, T. Mittag, and R. V. Pappu. “Connecting Coil-to-Globule Transitions to Full Phase Diagrams for Intrinsically Disordered Proteins”. en. In: *Biophysical journal* 119.2 (July 2020), pp. 402–418. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2020.06.014.

[48] F. G. Quiroz and A. Chilkoti. “Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers”. eng. In: *Nature materials* 14.11 (Nov. 2015), pp. 1164–1171. ISSN: 1476-1122. DOI: 10.1038/nmat4418.

[49] D. Vila-Viçosa, T. F. D. Silva, G. Slaybaugh, Y. K. Reshetnyak, O. A. Andreev, and M. Machuqueiro. “Membrane-Induced pKa Shifts in wt-pHLIP and Its L16H Variant”. In: *Journal of chemical theory and computation* (May 2018). ISSN: 1549-9618. DOI: 10.1021/acs.jctc.8b00102.

[50] V. Vasquez-Montes, J. Gerhart, D. Thévenin, and A. S. Ladokhin. “Divalent Cations and Lipid Composition Modulate Membrane Insertion and Cancer-Targeting Action of pHLIP”. en. In: *Journal of molecular biology* 431.24 (Nov. 2019), pp. 5004–5018. ISSN: 0022-2836. DOI: 10.1016/j.jmb.2019.10.016.

[51] M. D. Smith and M. Jelokhani-Niaraki. “pH-induced changes in intrinsically disordered proteins”. eng. In: *Methods mol. biol.* 896 (2012), pp. 223–231. ISSN: 1940-6029. DOI: 10.1007/978-1-4614-3704-8_14.

4. PHYSIOCHEMICAL PROPERTIES ALTERED BY THE TAIL GROUP OF LIPID MEMBRANES INFLUENCE HUNTINGTIN AGGREGATION AND LIPID BINDING[1]

Huntington's Disease is a neurodegenerative disorder caused by an expanded polyglutamine (polyQ) domain within the huntingtin protein (htt) that spurs the formation of toxic protein aggregates. Htt directly interact with membranes, which influence the aggregation process and can lead to membrane abnormalities. These interactions are facilitated by the 17 Nterminal residues (Nt17) which form an amphipathic α-helix (AH) implicated in both lipid binding and aggregation. Here, the impact of lipid tail unsaturation (DMPC, POPC, or DOPC) on htt/lipid interactions was investigated. The influence of the lipid vesicles on aggregation did not correlate with the degree of htt-membrane association, indicating that membrane properties alter htt aggregation. This can be linked to the initial lipid-binding mechanisms that influence Nt17 orientation on the membrane surface based on compatibility between specific hydrophobic residue and membrane defects.

4.1 INTRODUCTION

Numerous neurodegenerative diseases are characterized by formation and deposition of toxic protein aggregates in tissues. Huntington's Disease (HD) is caused by an expanded polyglutamine (polyQ) domain within the huntingtin protein (htt) that promotes aggregation.

[1]This article/chapter was published in *J. Phys. Chem. B*, **126**, Beasley, M., Frazee, N., Groover, S., Valentine, S., Mertz, B., Legleiter, J., Physicochemical Properties Altered by the Tail Group of Lipid Membranes Influence Huntingtin Aggregation and Lipid Binding, Copyright 2022 American Chemical Society. All experimental laboratory work (Thioflavin T assays, AFM, PDA assays, and mass spectrometry) was completed by Maryssa Beasley.

PolyQ expansion beyond a threshold of 35 residues is associated with disease and correlates with the extent of aggregation, age of onset, and severity of symptoms [1, 2]. Htt aggregation is a complex, multipathway process that results in a heterogeneous mixture of aggregates ranging from oligomers and annular aggregates [3, 4] to β-sheet-rich fibrils and inclusion bodies [5, 6, 7]. The heterogeneity of htt aggregation complicates assigning specific modes of toxicity to discreet aggregate species. While aggregation leads to the formation of toxic aggregates, other species may prove benign or even protective. The diffuse fraction of htt within cells appears particularly toxic [7, 8]; however, this fraction is populated by a complex mixture of monomers, oligomers, amorphous aggregates, and fibrils [9, 10, 11]. Despite this complexity, htt oligomers [8, 10], fibrils [12], and even inclusions [13] have all been linked to a variety of specific toxic mechanisms.

The polyQ domain of htt is flanked on the N-terminal end by 17 amino acids (Nt17, MATLEKLMKAFESLKSF) that act as a lipid-binding domain and govern the interaction of htt with membranes [14, 17, 16]. Nt17 is intrinsically disordered in bulk solution, but undergoes a structural transition to an amphipathic α-helix (AH) in the presence of binding partners like lipid membranes [17, 16]. The interaction between htt and phospholipid membranes via Nt17 plays a role in a number of normal htt functions [15, 16] and cytotoxic mechanisms [15, 17] by trafficking htt to mitochondria, vesicles, the ER, and the nucleus [17, 16, 15]. However, these membrane interactions can become detrimental with mutant htt as it damages and destabilizes membranes [14, 18], and this correlates with polyQ length [19]. This may result from the aggregation process, as htt aggregates incorporate membranous assemblies [20] and brain lipids [18] into their structures. These interactions between htt aggregates and membranes are linked to cell death [21], synaptic degeneration [22], and deformation of membranous organelles [23, 24].

Nt17 not only facilitates htt's interaction with phospholipids but also forms intermolecular associations that promote htt aggregation [25, 13, 14]. These interactions result in α-helical oligomers that act as nuclei for fibril formation [14]. The dual role of Nt17 in lipid binding and aggregation leads to a direct influence of membranes on the kinetics and mechanism of htt aggregation [18, 25]. 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC)/1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoserine (POPS) vesicles catalyze htt aggregation through a unique Nt17-controlled membrane-mediated mechanism [25]. Conversely, brain lipid extracts inhibit htt fibrillization [26] and stabilize specific oligomer species

[27]. Increased cholesterol [28], sphingomyelin, and GM1 [27] content in membranes also influences the morphology of htt aggregates.

The structural transition of Nt17 from intrinsically disordered to AH is a hallmark of proteins that play a role in membrane-associated cellular processes such as intracellular membrane targeting [29], curvature sensing [30], and membrane remodeling [31]. Many of these proteins function in the early secretory pathway and nuclear envelope [30, 32], while others function in the distinctly different lipid environment of the endosomal or plasma membranes [33]. The interaction between the AH and specific types of cellular membranes are crucial for proper function, but the properties contributing to their membrane affinity are poorly understood. Factors that contribute to the membrane selectivity of AH-containing proteins like Golgi tethers or α-synuclein have been identified, but they rely on specific chemical properties of the helix such as a lack of charged residues or a hydrophobic face consisting entirely of small residues [33]. These sequences result in an asymmetric helical topology where either electrostatic or hydrophobic effects drive AH binding. Few mechanistic details have been identified for the association of AHs such as Nt17 that contain both a well-defined hydrophobic face and charged residues.

Understanding the physicochemical properties of membranes that influence the interaction of Nt17 with lipids represents a promising target for the inhibition of htt aggregation and lipid binding and can provide critical insights into our fundamental understanding of the mechanism of AH binding to cellular membranes. To determine how membrane physicochemical properties influence the behavior of htt, htt aggregation and membrane association were characterized in the presence of homogeneous lipid systems with the same zwitterionic headgroup but different tails (1,2,-dimyristoyl-sn-glycero-3-phosphocholine (DMPC), POPC, and 1,2,-dioleoyl-sn-glycero-3-phosphochlline (DOPC)). Here, the impact of the lipid vesicles on htt aggregation was elucidated with Thioflavin T aggregation (ThT) assays and atomic force microscopy (AFM) , while the degree and mechanism of htt membrane association and complex formation was determined with a combination of polydiacetylene (PDA) lipid binding assays, native mass spectrometry (MS), and molecular dynamics (MD) simulations. Our results demonstrate that htt aggregation does not correlate with the degree of htt-lipid interaction, indicating that the mechanism of membrane-mediated aggregation is unique to each lipid system. In addition, the interaction of Nt17 with the membranes is driven by a combination of peptide and membrane properties; 1) shape

complementarity between amino acid sidechains and membrane defects and 2) overall binding of the entire peptide.

4.2 METHODS

4.2.1 Purification of glutathione S-transferase (GST)-htt exon 1 fusion protein

Htt-exon1 with 46 repeat glutamine residues (htt-exon1(46Q)) was expressed in e.coli as a glutathione S-transferase (GST) fusion protein and purified as previously described [34]. Briefly, the fusion proteins were purified via affinity liquid chromatography using a GST column (BioRad LPLC). Relevant fractions were tested by SDS-PAGE to assess purity. Fusion protein solutions were dialyzed for two days to remove glutathione used to elute fusion proteins from the column. To remove pre-existing aggregates, fusion proteins were subjected to high speed centrifugation (22000 ×g) prior to any experiment. Addition of Factor Xa (Promega, Madison, WI) cleaved the GST tag and initiated aggregation.

4.2.2 Lipid Vesicle Formation

DMPC, POPC, and DOPC were obtained from Avanti Polar Lipids as dried films and rehydrated in tris buffer (pH = 7.4) for 1 h at 1 mg/mL. Lipid solutions were subjected to ten freeze-thaw cycles, proceeded by bath sonication for 1 h. The size and polydispersity of lipid vesicles were measured via dynamic light scattering (DLS, NanoBrook 90plus Particle Size Analyzer, Brookhaven Instruments) to verify the formation of large unilamellar vesicles (LUVs). The mean and standard deviation of vesicle sizes were determined by assuming a log-normal distribution of the DLS data.

4.2.3 Thioflavin T Assay

To monitor fibril formation as a function of time, a Thioflavin T (ThT, Sigma-Aldrich, St. Louis, MO) aggregation assay was performed. Htt-exon1(46Q) (20 μM) was incubated with ThT (125 µM) in the presence and absence of various lipid vesicles. In the lipid conditions, lipid concentration was 100 μM, 200 μM, or 400 μM resulting in 5:1, 10:1, or 20:1 lipid:protein ratios. Reactions were run in black Costar 96-well plates with clear flat bottoms, and ThT fluorescence was monitored using a SpectraMax M2 microplate reader. Experiments were run at 37 ^{o}C with 440 nm excitation and 484 nm emission. Readings were taken every

5 min for 18 h, with shaking prior to each read. Relative maximum fluorescence was determined by normalizing the maximum fluorescence intensity of each condition to the maximum fluorescence of the huntingtin control (100%). The relative rate for each condition was calculated over a period of 3 h, beginning at the point when the fluorescence intensity reached 20% of the maximum for that condition. This setpoint was determined to ensure that the data had made a significant increase from the baseline value and that rate calculations did not begin in the noise of the lag phase. All rate values were normalized to the rate of the huntingtin control (100%).

4.2.4 Atomic Force Microscopy

Htt-exon1(46Q) (20 μM) was incubated with and without lipid vesicles (200 µM for a 10:1 lipid:protein ratio) at 37 ^{o}C and 1400 rpm using an orbital mixer. At 1, 3, 5, and 8 h time points, 2 µL aliquots of each condition were deposited on freshly cleaved mica for one min followed by a 200 μL wash with 18 MΩ water and drying with a gentle stream of clean air. These samples were imaged using a Nanoscope V Multi-Mode scanning probe microscope (VEECO) equipped with a closed loop vertical engage J-scanner. Silicon-oxide cantilevers with a nominal spring constant of 40 N/m and a resonance frequency of 300 kHz were used. Scan rates were set to 1.99 Hz with cantilever drive frequencies at 10% of resonance. All images were analyzed using the Matlab image processing toolbox (MathWorks) as previously described [35, 5]. Oligomers were classified as features at least 1.0 nm in height occupying a surface area less 4,000 nm2 with an aspect ratio less than 3.0, indicating a globular structure. The 1 nm height threshold avoided counting artifacts associated with lipid backgrounds (as determined by control images, **Fig. 4.1**). Fibrils were defined as features with aspect ratios larger than 3.0 and occupying at least 4,000 nm^2. To eliminate bias caused by varying fibril lengths and bundling, the percent surface covered by fibrils was used as a measure of fibril load.

4.2.5 Polydiacetylene Lipid Binding Assay

Polydiacetylene (PDA) assays were performed to measure the interaction between htt and lipid vesicles over time using reported protocols [36, 37]. In short, diacetylene monomers of 10,12-tricosadiynoic acid and the pure lipid system of choice were mixed at a 2:3 molar ratio in a 4:1 chloroform/ethanol solution. The organic solvents were evaporated off with a gentle

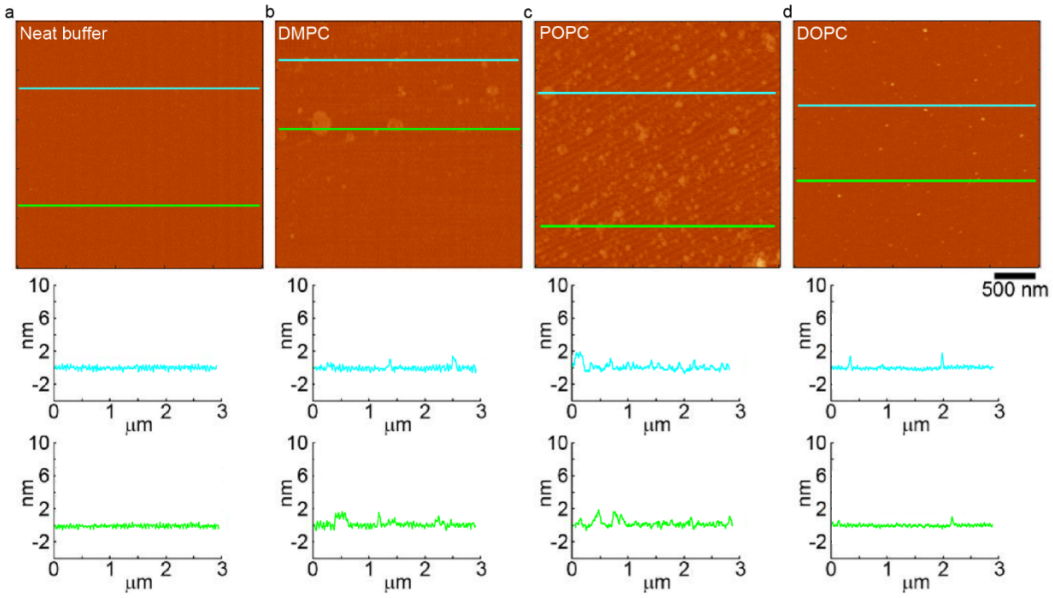

Figure 4.1: Representative AFM images of the backgrounds associated with (a) neat buffer, (b) DMPC lipid, (c) POPC lipid, (d) and DOPC lipid. The lines in each image correspond to the height profiles provided below each image.

stream of nitrogen. Films were reconstituted in tris buffer (70 ^{o}C), extensively sonicated to promote mixing, and left at 4 ^{o}C overnight to allow for self-assembly into PDA/lipid vesicles. For each experiment, the vesicle solutions were equilibrated to 25 ^{o}C and irradiated at 254 nm with constant stirring to polymerize the 10,12-tricosadiynoic acid, resulting in a royal blue solution that would display a colorimetric shift to red upon applied mechanical stress to the vesicles. PDA/lipid solutions were exposed to htt-exon1(46Q) (20 μM) for 18 h at 30 ^{o}C. This temperature ensures that all lipid vesicles are above their transition temperature (DMPC Tm = 24 ^{o}C, POPC Tm = -2 ^{o}C, DOPC Tm = -17 ^{o}C)[38] and are in the fluid phase. For each condition, the absorbance of the blue (650 nm) and red (500 nm) wavelengths were recorded by a SpectraMax M2 plate reader every 5 min with shaking between each reading for the duration of the experiment. A negative control consisted of equal volumes of neat buffer and PDA/lipid solution, while a positive control exposed vesicles to saturated NaOH (pH = 12) that stresses the system by increasing repulsion among lipid head groups, invoking a colorimetric response [39, 40]. The NaOH control can also be used to establish the sensitivity of the different lipid systems and used to normalize the colorimetric response

[41]. The % CR was calculated for each condition using the following equation:

$$\%CR = \frac{PB_0 - PB}{PB_0} 100 \tag{4.1}$$

where the PB is defined as $A_{blue}/(A_{blue} + A_{red})$ for the negative control (PB_0) and sample condition (PB). A_{blue} and A_{red} refer to the measured absorbance at the blue and red wavelengths respectively.

4.2.6 Electrospray Ionization-Mass Spectrometry (ESI-MS)

For MS experiments, lipid vesicles were formed as previously described, but using HPLC-grade water instead of tris buffer. Nt17 (10 μM) was incubated alone or with DMPC, POPC, or DOPC vesicles (10:1 lipid:peptide ratio) at 37 ^{o}C for 3 h before analysis. All ESI-MS experiments were performed using a Q Exactive Hybrid Quadruple Orbitrap mass spectrometer (Thermo Fischer, San Jose, CA). Experiments were conducted using a HESITM as the degree of lipid complexation observed is not an artifact of ionization method [42]. MS spectra were recorded in positive ion mode over a mass-to-charge (m/z) range of 400 to 4,000. Samples were infused (10 μL/min) through a needle biased at 3,500 V above the instrument inlet. The parameters for the MS instrument were: 400 ^{o}C for the capillary inlet temperature, 30 ^{o}C for the analyzer temperature, 80 V for the S-lens assembly, 400 ms for the maximum injection time, 1 × 106 for the AGC, and 70,000 for the MS resolution. Each spectrum was recorded (90 s) in triplicate. Data were analyzed using Xcalibur 2.2 software suite (Thermo Scientific). Relative abundance was determined in relation to the doubly-charged monomer peptide ion ($[M^{+2}H]^{2+}$ at m/z 988).

4.2.7 MD simulation system preparation

The structure of Nt17 was generated based on a previous solid-state NMR study (PDB 2LD2) and was modified in VMD [43]. The standard N-terminal patch was applied, and the C-terminus was amidated to mimic the absence of charge that would be present with the full huntingtin protein. The peptide was solvated using the solvate plugin in VMD [43]. Randomly-coiled conformations were generated by heating from 300 K to 700 K in the NVT ensemble in NAMD 2.13 [44] over 20 ps followed by 1 ns at 700 K. A starting structure was randomly selected from among frames with the proper omega dihedral angles. Separately, three pure lipid bilayers of DMPC, POPC, and DOPC containing 150 lipids each were

created, solvated, and ionized using the Optimal Membrane Generator in LOOS [45, 46]. The systems were equilibrated using an in-house protocol consisting of concurrent slow heating and release of restraints on the lipids, resulting in a planar lipid bilayer after 100 ns of simulation time. The randomly-coiled peptide was merged with each of the respective membrane systems, initially placing the peptide 20 Å above the surface of the bilayer. Ion concentrations were adjusted to 150 mM using the ionize plugin in VMD [43]. The merged systems were equilibrated in the NPT ensemble for 1 ns at 310 K with the Langevin thermostat and 1 atm with the Langevin barostat in NAMD 2.13 [44]. System sizes ranged from 49,000 to 53,000 atoms.

4.2.8 MD simulations

To facilitate sufficient sampling of binding and dissociation events of Nt17, we used the weighted ensemble approach via Weighted Ensemble Simulation Toolkit with Parallelization and Analysis (WESTPA) [47]. WESTPA increases the sampling of rare events in MD while maintaining equilibrium conditions. This is accomplished via a user-defined progress coordinate which describes the configurational space of the system. The progress coordinate is further sub-divided into user-defined bins. WESTPA maintains a fixed number of simulations per bin tracking the probability that the bin will be populated over time. WESTPA spawns several short simulations of time, τ, after which it will evaluate the progress coordinate and add or remove simulations to maintain a constant number of simulations per bin. This approach facilitates a significantly increased fraction of simulations sampling the relevant space (i.e., proximal to the bilayer) and less time sampling the irrelevant space (i.e., diffusion in bulk solution). Each lipid system was run in triplicate with 100 iterations per replica, a dynamics period (τ) of 20 ps, and five simulations per bin. NAMD 2.13 [44] was used for dynamics with the same parameters as the final equilibration step in MD simulation system preparation. No methods of any kind were used to restrain the peptide or bilayer in a particular conformation. A one-dimensional progress coordinate was defined as a function of the number of contacts between Nt17 and the bilayer. The progress coordinate, C_{AB}, was calculated as such

$$C_{AB} = \sum_{i \in A} \sum_{j \in B} \frac{1}{1 + (\frac{r_{ij}}{r_0})^6} \tag{4.2}$$

where A is the set of atoms in the lipids, B is the set of atoms in the peptide, r_{ij} is the distance between the ith lipid atom and jth peptide atom, and r_0 is the cutoff distance of

3 Å. The binning scheme was designed to evenly divide up the z-distance between the starting position of the peptide and the center of the bilayer. The bin values are as follows: 0.0, 0.4, 0.8, 1, 1.5, 2, 3, 4, 5, 7, 10, 15, 25, 40, 70, 120, 200, 300, 500, 650, 800, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3250, 3500, 3750, 4000, 4250, 4500, 4750, 'inf'. As a comparison to the weighted ensemble simulations, we also conducted continuous equilibrium MD simulations with longer timescales (500 ns each). Each system was run in triplicate for each lipid type with the same starting configurations and parameters as the WESTPA simulations. 500 ns was necessary to obtain systems that were sufficiently decorelated from their starting configurations (**Fig. 4.2**). Nt17 remained primarily dissociated from all bilayer systems consistent with our WESTPA simulations (**Fig. 4.3**). Despite the long timescales, we did not observe significant folding of Nt17 into an α-helix (**Fig. 4.4**). The following simulation parameters were used for both the WESTPA and equilibrium simulations. We used the CHARMM 36 forcefield including the 36m parameters for proteins and the TIP3P water model. The switchdist, cutoff, and parlistdist were set to 12 Å, 14 Å, and 16 Årespectively. We used a Langevin Piston barostat set to 1.01325 bar and the Stochastic Rescaling thermostat set to 310 K. PME was turned on with an interpolation order of 6 and grid spacing set to 1.0. Simulations were run at a 2 fs timestep.

4.2.9 Molecular Dynamics Analysis

Analysis was performed on the subset of each weighted ensemble simulation that had a progress coordinate of 120 or greater at either the beginning or end of the iteration. A progress coordinate of 120 was chosen as a threshold based on visual inspection of the simulations where Nt17 was near the bilayer surface. These snapshots are referred to as "bound" throughout the paper. The distribution of bound frames can be seen in **Fig. 4.5**. *Peptide-bilayer contacts.* The number of contacts per frame was determined using a custom python script and pyLOOS [45, 46] by iterating through every atom in the peptide paired with every atom in the bilayer; those that were less than 4 Åapart were considered a contact. *Distance plots.* dz is defined as the projection along the z-axis of the vector between the geometric average position of the atoms in each respective residue and the surface of the bilayer. This involved two calculations: 1) calculation of the length of the projection along the z-axis of the vector between the geometric average position of the atoms in each respective residue and the P atoms of the bilayer, 2) calculation of the average thickness of

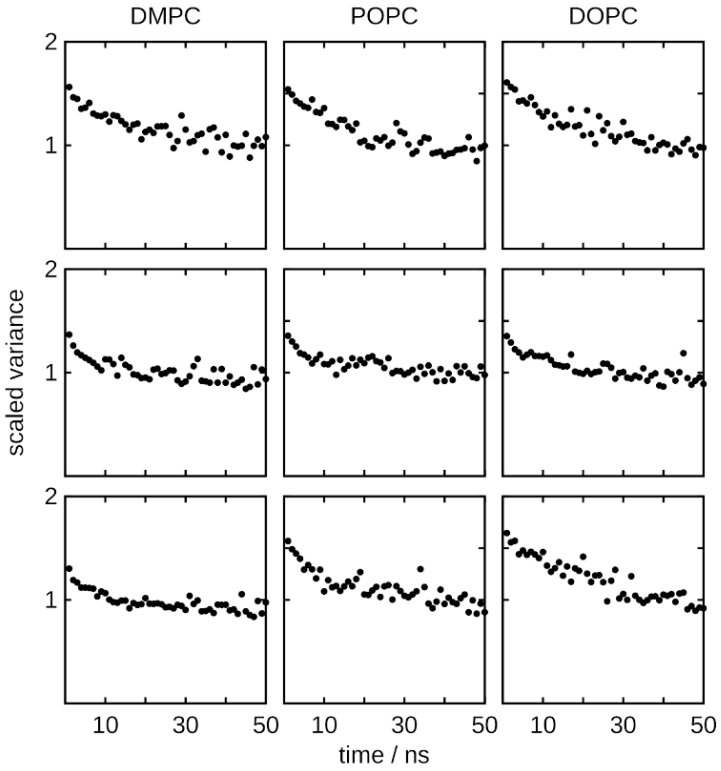

Figure 4.2: Decorrelation times for Nt17 with respect to starting position in equilibrium MD. Decorrelation of Nt17 from initial position in equilibrium MD simulations occurs within roughly 25 ns (scaled variance = 1) for all systems regardless of lipid type.

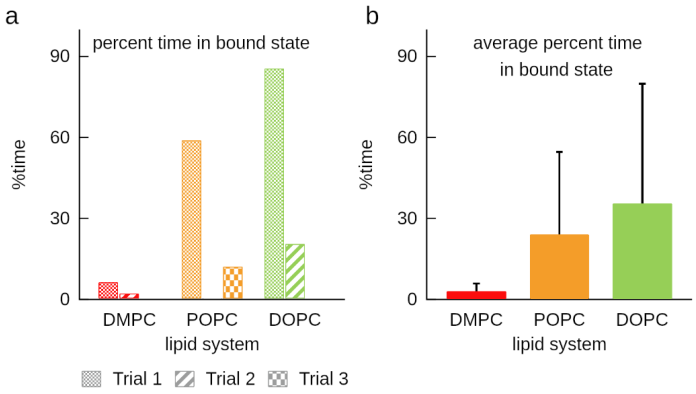

Figure 4.3: Propensity of binding events in equilibrium simulations. a) The percent time from each equilibrium trial by lipid system where Nt17 is considered to be in the "bound" state (where the same pcoord measurement used in the WESTPA simulations is greater than 120). b) Average percent of time that Nt17 is in the bound state.

the bilayer by measuring the distance along the membrane normal between the P atoms

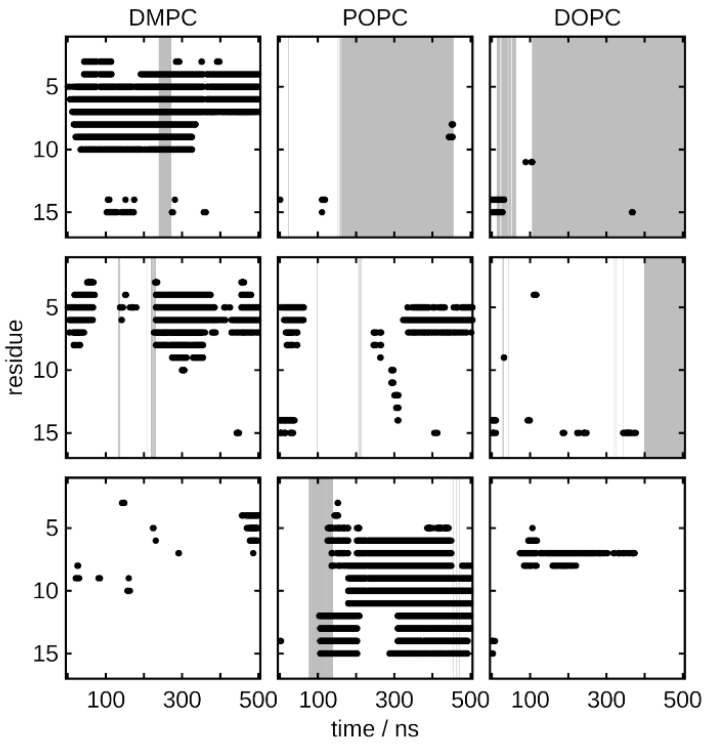

Figure 4.4: Binding and folding of Nt17 is often transient and uncorrelated. Equilibrium MD simulations of Nt17 interacting with the three different lipid types were conducted for 500 ns. Events where Nt17 is near the bilayer (pcoord of 120 or above: gray-shaded background) do not necessarily correlate with the event a particular residue is helical (black dot) for each trial and each lipid. It appears that partitioning-folding coupling of Nt17 to a lipid bilayer occurs on timescales ¿ 500 ns.

in each respective leaflet of the bilayer. Half the bilayer thickness was subtracted from the z-projection to obtain dz. Both measurements were made using the *interdist* tool in LOOS [45, 46]. *Orientation of Nt17.* The first principal axis of the peptide was measured against the bilayer normal using a custom python script and pyLOOS [45, 46]. Per residue distance dependent orientation: Distance was calculated similarly to dz, however only the terminal heavy atom of the sidechain was considered for each respective residue. To measure the angle of the sidechain with respect to the bilayer normal, a vector was drawn from the Cα carbon to the terminal heavy atom of each respective residue. Membrane maps: The *membrane_map* tool in LOOS [45, 46] was applied to each trial calculating the windowed height of the P atoms in the upper leaflet of the bilayer. Similar to dz, half the bilayer thickness was subtracted from each windowed height to determine the displacement from

the average. *Membrane defects.* Measurements of defects were based on the approach used by Voth and coworkers [34]. With the trajectory centered in the xy-plane on the peptide, the measure sasa tool in VMD [43] was used with a 3 Åseparation, -samples 300, -restrict to a selection of $x/y < -13$ and > 13 and $z > 10$ (restricting the points to being on the surface of the bilayer), and -points to create a point-based surface of the bilayer. Points were saved every 100 frames. Points were categorized based on their z-position with respect to the plane of the N atoms in the upper leaflet (protrusions were above the plane; defects were below the plane). Networks were constructed by grouping points within 0.6 Åof one another in the xy-plane. Areas were determined by first finding the maxima and minima of all the points for the frame to determine the rectangular area of the bilayer surface and then dividing the rectangular area into 2500 pixels of equal size. After this, the actual area was calculated by identifying how many pixels each network (i.e., defect/protrusion) occupied. *Helicity.* The phi and psi angles of the backbone were measured using the *ramachandran* tool in LOOS [45, 46]. A residue was considered helical if the previous and subsequent residues had a phi angle between -90 and -30 and a psi angle between -77 and -17 [48]. *Salt bridges.* Salt bridges were measured using the salt bridge plugin with default values in VMD [43] with modifications to accommodate measuring the N and P atoms of the lipid head groups.

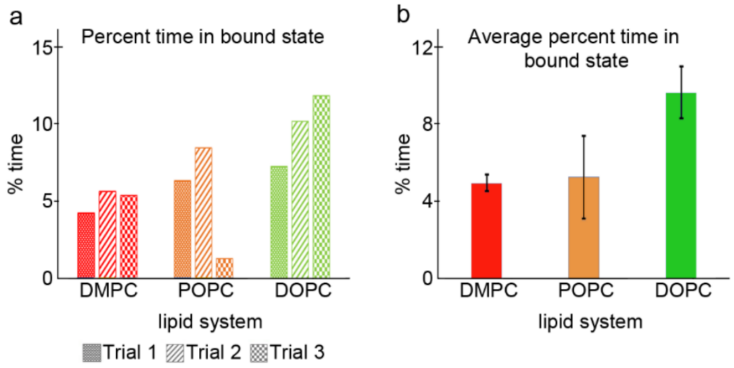

Figure 4.5: Propensity of binding events in WE simulations. a) The percent time from each WESTPA trial by lipid system where Nt17 is considered to be in the "bound" state (pcoord greater than 120). b) Average percent of time that Nt17 is in the bound state.

4.3 RESULTS AND DISCUSSION

4.3.1 Lipid membranes impact htt aggregation

Experiments were performed with a glutathione S-transferase (GST)-htt-exon1(46Q) fusion protein. Factor Xa cleaves GST, initiating aggregation. The impact of DMPC, POPC, or DOPC lipid vesicles (5:1, 10:1, and 20:1 lipid:htt) on fibrillization was determined by ThT assays. As determined by DLS, the preparation protocol produces lipid vesicles similar in size across all three lipid systems with a mean diameter in the range of 150-300 nm and polydispersity indexes (PDI) of ∼0.3–0 (representative DLS distributions are presented in **Fig. 4.6**).While vesicles alone don't invoke a ThT signal (**Fig. 4.7**), each lipid system impacted htt fibrillization (**Fig. 4.8**). Compared to aggregation in the absence of lipids, DMPC accelerated fibrillization in a dose-dependent manner, as the rate of elongation increased by 29%, 88%, and 114% at DMPC:htt ratios of 5:1, 10:1, and 20:1 respectively. The final fibril load, indicated by relative maximum fluorescence, increased in the presence of DMPC, with a maximum increase of 70% at the 20:1 DMPC:htt ratio. POPC did not significantly alter fibrillization until the lipid:htt ratio increased to 20:1, which invoked a 26% increase in aggregation rate and a 10% increase in maximum fluorescence. DOPC inhibited aggregation in a dose dependent manner, with the highest lipid:htt ratio reducing the maximum fluorescence 75%. The aggregation rate decreased with increasing DOPC:htt ratios (39% decrease at 5:1, 75% decrease at 10:1, 82% decrease at 20:1).

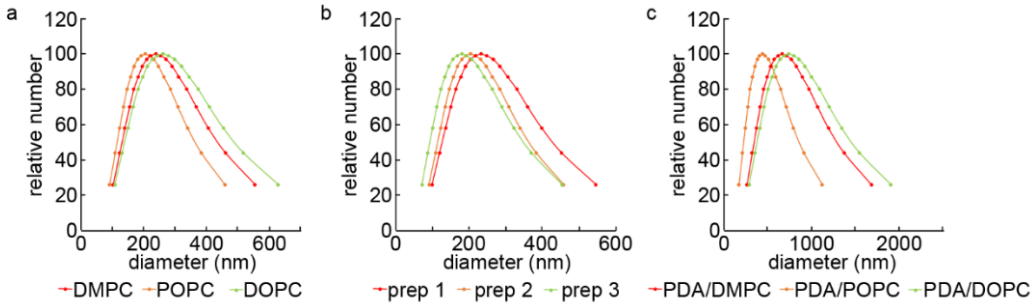

Figure 4.6: ThT aggregation assays for htt-exon1 (46Q) in the presence of (a) DMPC, (b) POPC, or (c) DOPC lipid vesicles. Htt-exon1 concentration was 20 µM and the lipid to htt molar ratio was 5:1, 10:1, or 20:1. (d) Direct comparison of the 20:1 ratio for each lipid system is shown. (e) The initial rate of aggregation and (f) the relative maximum fluorescence were determined with respect to the htt control. Error bars are provided for every sixth data point (30 min) and represent the standard error of the mean. One asterisk represents a p value of ¡0.05, and two asterisks represent a p value of ¡0.01.

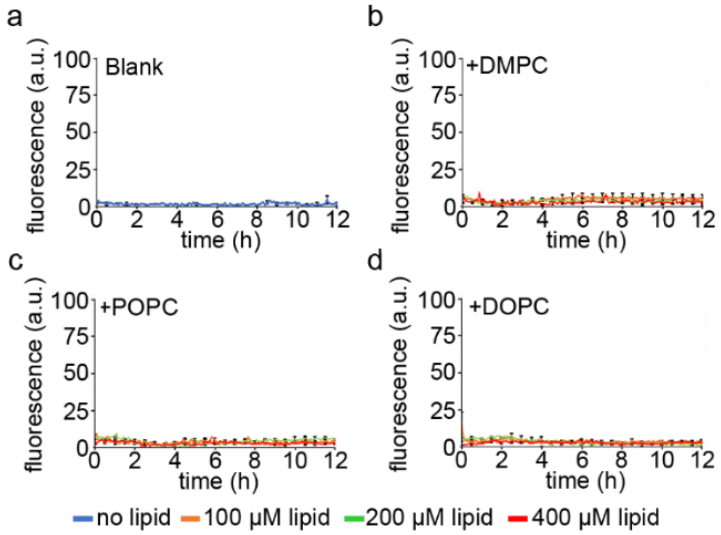

Figure 4.7: ThT fluorescence assay controls. (a) neat buffer, (b) DMPC vesicles, (c) POPC vesicles, and (d) DOPC vesicles.

As ThT predominately probes fibrillization, AFM assays were performed to analyze oligomer formation and aggregate morphology associated with aggregation in the presence of the different lipid vesicles (**Fig. 4.9**). A height threshold was utilized to avoid counting artifacts associated with lipid backgrounds (as determined by control images, **Fig. 4.1**), and aggregates were classified based on parameters defined in the Methods. Htt-exon1(46Q) (no lipid) forms oligomers within 1 h with a peak population after 3 h (**Fig. 4.10a**). A few fibrils were present at 1 h, but these became longer and more numerous with time, occupying a larger surface area. The oligomer population decreased with increasing fibril content. With DMPC, there was a slight increase in oligomers at 1 h compared with the control. After peaking at 3 h, the number of oligomers observed with DMPC was reduced at 5 and 8 h compared to the control ($p < 0.05$). DMPC enhanced fibrillization, with a clear increase in coverage at 5 h compared to the control ($p < 0.05$, **Fig. 4.10b**). The addition of POPC vesicles significantly reduced the htt oligomer population at 3, 5, and 8 h (**Fig. 4.10a**). Despite this reduction in oligomers, the extent of fibrillization was comparable to the control (**Fig. 4.10b**). DOPC altered aggregate populations, promoting and stabilizing oligomers as their population surpassed the control at 1 h and remained constant over 8 h (**Fig. 4.10a-b**). This oligomer stability was accompanied by a significant decrease in fibril content. Fibril content observed with each lipid system was consistent with ThT assays

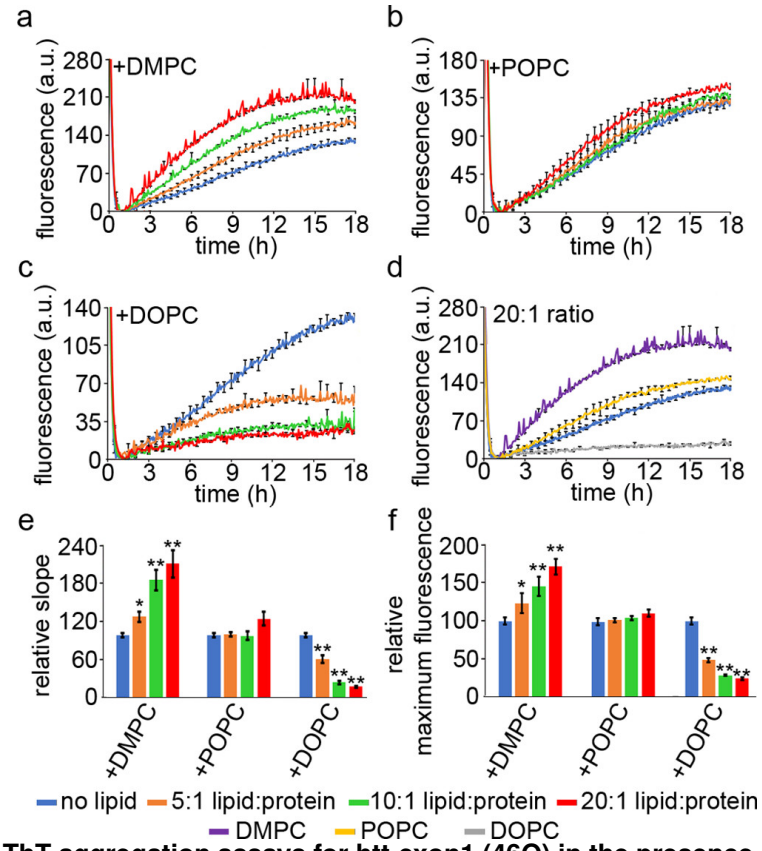

Figure 4.8: ThT aggregation assays for htt-exon1 (46Q) in the presence of (a) DMPC, (b) POPC, or (c) DOPC lipid vesicles. Htt-exon1 concentration was 20 μM and the lipid to htt molar ratio was 5:1, 10:1, or 20:1. (d) Direct comparison of the 20:1 ratio for each lipid system is shown. (e) The initial rate of aggregation and (f) the relative maximum fluorescence were determined with respect to the htt control. Error bars are provided for every sixth data point (30 min) and represent the standard error of the mean. One asterisk represents a p value of ¡0.05, and two asterisks represent a p value of ¡0.01.

in that with increasing amounts of lipid DMPC enhanced fibril formation, POPC did not significantly alter the extent of fibrillization, and fibrils decreased with DOPC.

Next, the morphologies of htt aggregates were compared. While varying rates of fibrillization result in different fibril contour lengths, the average height along the fibril contour was unchanged, suggesting that the underlying structure of fibrils was not altered by each lipid (**Fig. 4.11**). However, lipids altered oligomer morphology (**Fig. 4.10c-f**). Without lipids, oligomers increased in size with time (**Fig. 4.10c**). Specifically, oligomers had a mode height of 3-4 nm after 1 h, but this grew to 6-8.5 nm at 8 h with an increase in heterogeneity. A shift from smaller to larger oligomers occurred in the presence of DMPC, but the 8 h

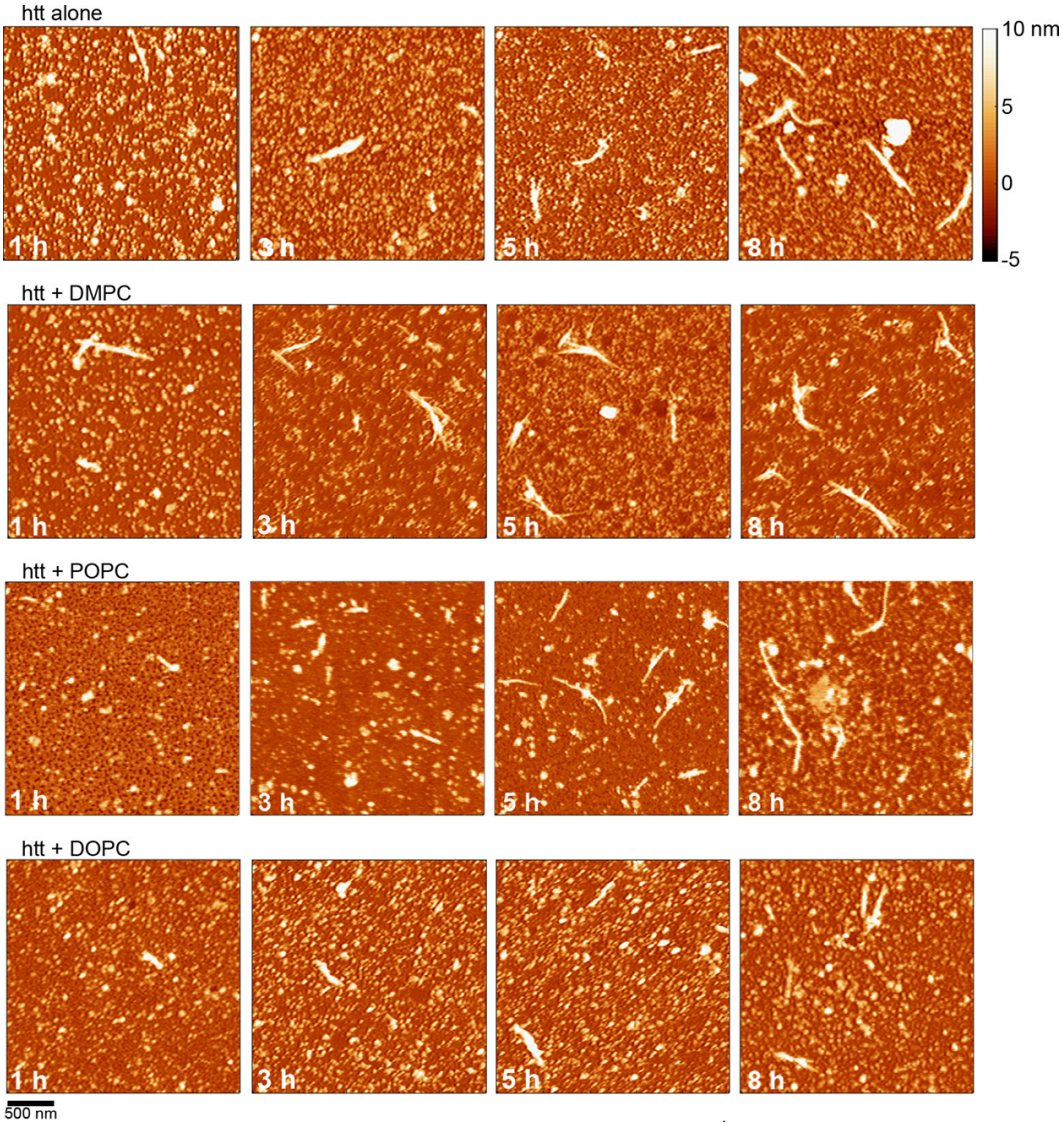

Figure 4.9: Representative AFM images of 20 µM htt-exon1(46Q) incubated alone, with DMPC, with POPC, or with DOPC lipid vesicles as a function of time (10:1 lipid:htt ratio). The color map is the same for all images.

population displayed less heterogeneity and a smaller mode height (4.5-6.5 nm, **Fig. 4.10d**). With both POPC and DOPC, oligomers were initially larger (mode height of 4-5 nm for both conditions); however, oligomer morphology was stable, displaying extensive overlap in the correlation plots between 1 and 8 h with no shift in oligomer mode height (**Fig. 4.10e-f**). In both cases, the stabilization of early oligomeric species is accompanied by a slight increase in the number of larger oligomers by 8 h. Specifically, the presence of POPC resulted in

the stabilization of smaller oligomers (4-5 nm) accompanied by the formation of a small population of large oligomers (8-9 nm) by 8 h. With regard to POPC, we have reported a similar reduction in oligomer size compared to htt incubated without lipids previously [49, 50]. However, the oligomer size distribution associated with aggregation in the presence of POPC is more heterogenous resulting in a larger population of bigger oligomers at 8 h of incubation (but still smaller than no lipid controls [49]); whereas, the oligomer size distribution in the presence of POPC remains tighter at 8 h here and in the other study [50]. Nevertheless, POPC vesicles consistently promote smaller htt oligomers compared to oligomers formed in the absence of lipid.

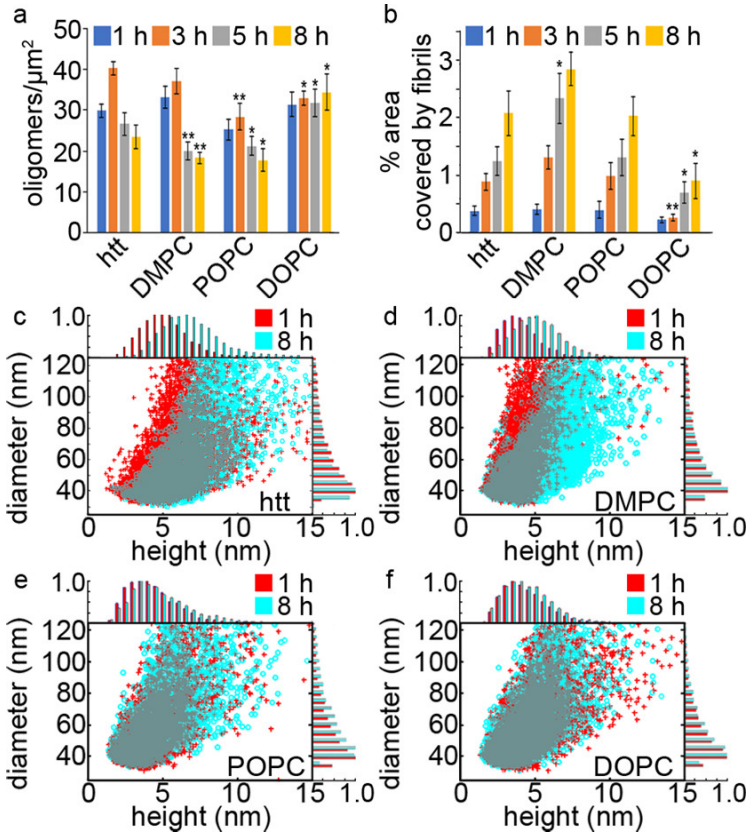

Figure 4.10: AFM analysis of the impact of DMPC, POPC, or DOPC lipid vesicles on htt-exon1(46Q) aggregation (10:1 lipid:htt ratio). (a) Analysis of the number of oligomers per unit area as a function of time for each condition. (b) Statistical analysis of the percent of each image area covered by fibrils as a condition of time. (c-f) Comparison of the height and diameter of oligomers formed at 1 h (red) and 8 h (blue) in the presence of (c) no lipid, (d) DMPC, (e) POPC, and (f) DOPC vesicles. One asterisk represents a p value of ¡0.05, and two asterisks represent a p value of ¡0.01.

4.3.2 Saturation in lipid tails influence htt/membrane interactions

The varying impact of the lipid systems on aggregation of Nt17 suggests differences in htt/lipid interaction, so polydiacetylene (PDA)/lipid affinity assays were performed as a first assessment of htt/lipid interaction (**Fig. 4.12**). PDA/lipid vesicles undergo a colorimetric shift (blue to red) as protein interaction with the vesicles results in a mechanical strain on the PDA polymer backbone. By monitoring blue (640 nm) and red (500 nm) absorbances of the vesicles upon exposure to htt-exon1(46Q), a percent colorimetric response (% CR) can be calculated that directly correlates to the extent of htt/lipid interaction. To enable direct comparison between the different lipid systems, %CR was normalized against a standard dose of NaOH. It should be noted that, based on DLS measurements, PDA/lipid vesicles are larger than the simple lipid vesicles used throughout the rest of this study, with mean diameters ranging from 440-750 nm and PDI of $\sim$0.4 (**Fig. 4.6**). Based on this assay, the relative affinity of htt-exon1(46Q) for the different lipid systems was POPC $>$ DMPC $>$ DOPC. Htt quickly bound POPC vesicles, invoking a greater %CR within 3 h than was observed for the other lipid systems over 18 h. After this initial jump, the %CR continued to increase. With DMPC vesicles, the %CR steadily rose for 9 h until a steady state interaction of $\sim$20 %CR was observed. It took over 3 h of htt exposure before DOPC vesicles elicited an observable %CR; however, once initiated, the %CR continued to gradually increase. The overall trend did not correlate with degree of lipid tail unsaturation, and the differences between the %CR curves suggests mechanistic differences in the interaction of htt with these lipid systems.

4.3.3 Nt17 forms a variety of complexes with lipids

As Nt17 facilitates htt-exon1 lipid interactions, ESI-MS was utilized for semi-quantitative assessment of the degree of peptide-lipid interactions between Nt17 peptides and the lipid systems (**Fig. 4.13**, **Tab. 4.1**). Without lipids, several Nt17 ions were present (**Fig. 4.13a**). The $[M^{+2}H]^{2+}$ ions at m/z 988 were the predominant species, followed by the $[M^{3+}H]^{3+}$ ions at m/z 658 and the $[M^{+}H]^{+}$ ions at m/z 1974. Dimer, trimer, and tetramer ions for the Nt17 peptide (m/z 1317, 1482, and 1580, respectively) were observed at significantly lower abundances. The lyophilized synthetic peptide contained sodium, resulting in detectable sodium-containing ions.

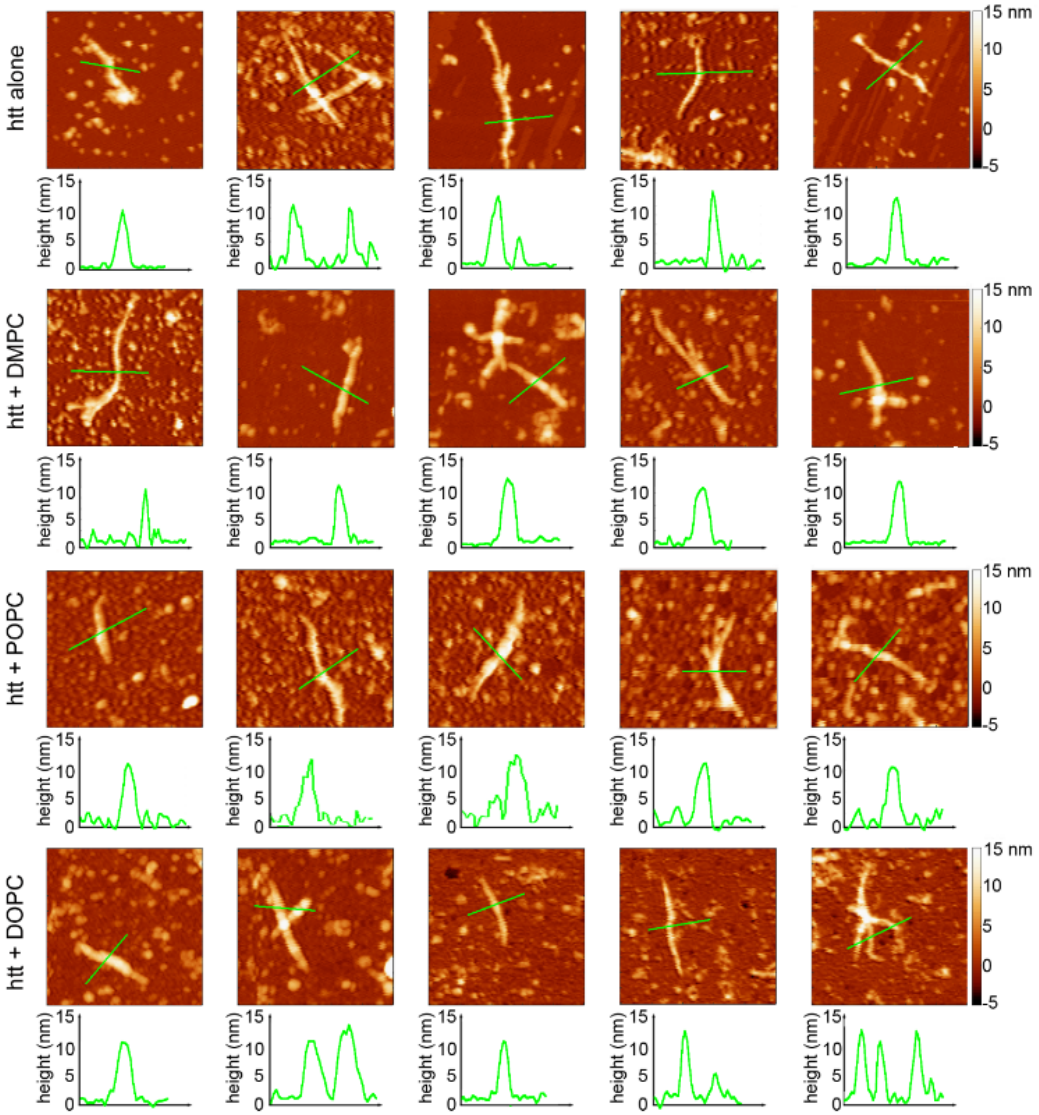

Figure 4.11: AFM images comparing the morphology of htt-exon1(46Q) fibrils formed in the absence of lipid, or when incubated with DMPC, POPC, or DOPC lipid vesicles. The color lines in each image correspond to the height profiles directly below each image. The color map is the same for all images.

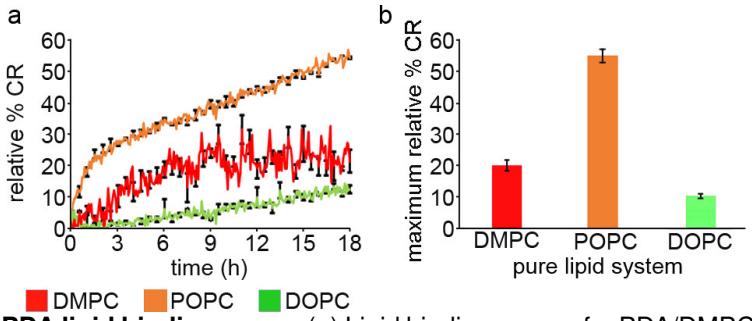

Figure 4.12: PDA lipid binding assay. (a) Lipid binding assays for PDA/DMPC, PDA/POPC, and PDA/DOPC vesicles exposed to 20 μM htt-exon1(46Q). Error bars are provided for every sixth data point and represent the standard error of the mean. (b) The maximum relative % CR obtained for each PDA/lipid system upon exposure to htt.

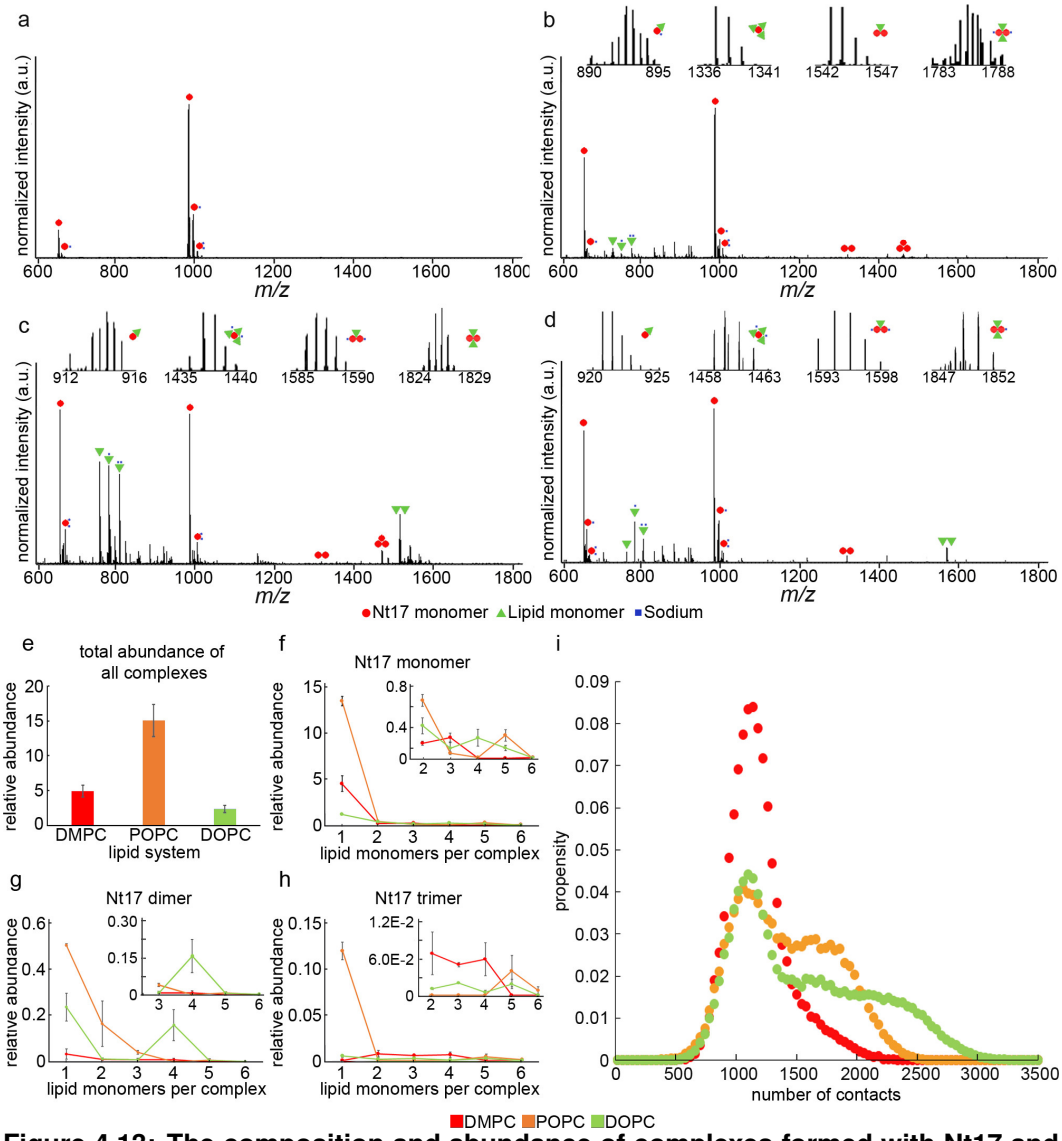

Figure 4.13: The composition and abundance of complexes formed with Nt17 and different lipids was analyzed via ESI-MS. Nt17 was incubated (a) alone or with (b) DMPC, (c) POPC, or (d) DOPC lipid vesicles. (e) The total abundance of peptide-lipid complexes was compared across all systems, as well as the abundance of complexes by lipid content with Nt17 (f) monomers, (g) dimers, and (h) trimers. Insets (f-h) illustrate an enlargement of select portions of the given plots. (i) The total molecular contacts made between Nt17 and DMPC, POPC, or DOPC lipid bilayers in MD simulations.

Ion[a]	m/z	Charge State	Ion[a]	m/z	Charge State	Ion[a]	m/z	Charge State
Nt17 + 1DMPC	2652.99	1	Nt17 + 1POPC	2734.63	1	Nt17 + 1 DOPC	2760.64	1
Nt17 + 1DMPC	1327	2	Nt17 + 1POPC	912.21	3	Nt17 + 1DOPC	1380.82	2
Nt17 + 1DMPC	885	3	Nt17 + 1POPC	1367.82	2	Nt17 + 2DOPC	1773.62	2
Nt17 + 2DMPC	1665.47	2	Nt17 + 2POPC	1747.11	2	Nt17 + 3DOPC	2166.42	2
Nt17 + 2DMPC	1110.98	3	Nt17 + 2POPC	1165.41	3	Nt17 + 3DOPC	1444.61	3
Nt17 + 3DMPC	2004.45	2	Nt17 + 3POPC	2126.91	2	Nt17 + 4DOPC	2559.22	2
Nt17 + 3DMPC	1336.96	3	Nt17 + 3POPC	1418.6	3	Nt17 + 5DOPC	2952.02	2
Nt17 + 4DMPC	2343.42	2	Nt17 + 4POPC	2506.7	2	Nt17 + 6DOPC	2230.21	3
Nt17 + 5DMPC	2682.4	2	Nt17 + 4POPC	1671.8	3	2Nt17 + 1DOPC	2367.84	2
Nt17 + 6DMPC	3021.37	2	Nt17 + 5POPC	2897.99	2	2Nt17 + 1DOPC	1578.89	3
Nt17 + 6DMPC	2014.91	3	Nt17 + 5POPC	1925	3	2Nt17 + 2DOPC	2760.64	2
2Nt17 + 1DMPC	2314.02	2	2Nt17 + 1POPC	2354.84	2	2Nt17 + 3DOPC	3153.44	2
2Nt17 + 1DMPC	1543.01	3	2Nt17 + 1POPC	1570.22	3	2Nt17 + 4DOPC	2364.49	3
2Nt17 + 2DMPC	2652.49	2	2Nt17 + 2POPC	2734.13	2	2Nt17 + 5DOPC	2626.36	3
2Nt17 + 3DMPC	2991.47	2	2Nt17 + 2POPC	1823.42	3	2Nt17 + 6DOPC	2888.22	3
2Nt17 + 3DMPC	1994.98	3	2Nt17 + 3POPC	3113.93	2	3Nt17 + 1DOPC	2236.9	3
2Nt17 + 4DMPC	3330.44	2	2Nt17 + 3POPC	2076.62	3	3Nt17 + 2DOPC	2498.77	3
2Nt17 + 4DMPC	2220.96	3	2Nt17 + 5POPC	2583.01	3	3Nt17 + 3DOPC	2760.64	3
3Nt17 + 2DMPC	2427.01	3	3Nt17 + 1POPC	2235.9	3	3Nt17 + 4DPOC	3022.51	3
3Nt17 + 3DMPC	2652.99	3	3Nt17 + 5POPC	3241.02	3	3Nt17 + 5DOPC	3284.37	3
3Nt17 + 4DMPC	2878.97	3	3Nt17 + 6POPC	3494.22	3			

Table 4.1: Assigned ions for Nt17 incubated with DMPC, POPC, or DOPC lipid vesicles along with their mass to charge ratios.

[a] The following average masses have been used to assign the peaks: 1974.04 for Nt17 peptide, 677.95 for DMPC lipid, 759.59 for POPC lipid, and 785.60 for DOPC lipid.

A heterogeneous mixture of ions corresponding to Nt17-lipid complexes (**Fig. 4.13b-d**, **Tab. 4.1**) and pure lipid complexes (**Fig. 4.14** and **Tab. 4.2**) were observed with all three lipid systems. Complexes contained monomeric to trimeric Nt17 and up to 6 lipids, with the makeup and abundance of complexes varying between lipid systems (**Fig. 4.13e-h**). The relative intensity of ions for each lipid system decreased with multimers of increasing size. However, lipid content in complexes with each system varied. The number of distinct complexes identified was similar for each lipid system (**Tab. 4.1**), but the total relative abundance paralleled the PDA assay (POPC $>$ DMPC $>$ DOPC, **Fig. 4.13e**). That said, complexes of the same composition but different charge states were observed. POPC had the lowest number of identified Nt17-lipid combinations despite having the highest relative abundance of ions while DOPC had the most distinct combinations but the lowest relative abundance.

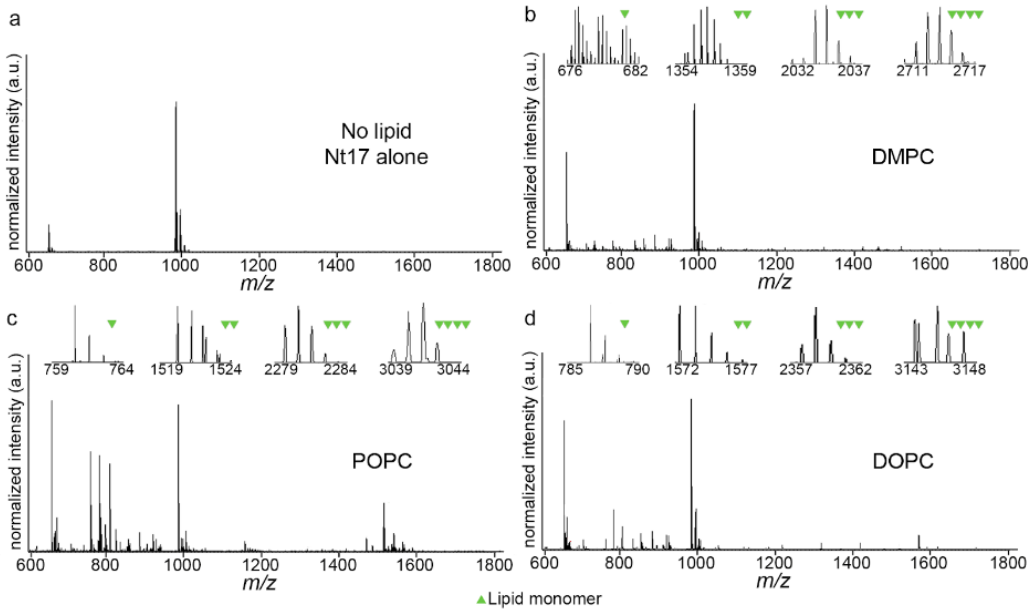

Figure 4.14: Identification of lipid only species by ESI-MS. (a) A control spectra of Nt17 with no lipids, and spectra of Nt17 with (b) DMPC, (c) POPC, or (d) DOPC lipid vesicles are presented. In each spectra, lipid only species are identified.

Monomeric Nt17 complexes were observed with up to six DMPC lipid monomers, with the most abundant complexes containing one DMPC lipid (relative abundance of 4.44%), three DMPC lipids (0.22%), and then two DMPC lipids (0.16%, **Fig. 4.13f**). Dimeric and trimeric Nt17-DMPC complexes were significantly less abundant, containing only up to

Ion[a]	m/z	Charge State	Ion[a]	m/z	Charge State	Ion[a]	m/z	Charge State
1DMPC	678.95	1	1POPC	760.59	1	1DOPC	786.6	1
2DMPC	1356.9	1	2POPC	1520.18	1	2DOPC	1572.2	1
3DMPC	2034.85	1	3POPC	2279.77	1	3DOPC	2357.8	1
4DMPC	2712.8	1	4POPC	3039.36	1	4DOPC	3143.4	1
3DMPC	1017.93	2	2POPC	760.59	2	2DOPC	786.6	2
4DMPC	1356.9	2	4POPC	1520.18	2	3DOPC	1179.4	2
3DMPC	678.95	3	3POPC	760.59	3	3DOPC	524.73	3
3DMPC	509.46	4	4POPC	570.69	4	4DOPC	590.2	4

Table 4.2: Assigned ions for DMPC, POPC, or DOPC lipid multimeric species along with their mass to charge ratios.

[a] The following average masses have been used to assign the peaks: 677.95 for DMPC lipid, 759.59 for POPC lipid, and 785.60 for DOPC lipid.

four DMPC lipids (**Fig. 4.13g-h**). Complexes of trimeric Nt17 with one DMPC lipid were nonexistent. Instead, there was a high abundance of trimeric Nt17 complexes with multiple DMPC lipids compared to other systems. The total relative intensities of Nt17-DMPC complex ions was 4.92%. In contrast, the total relative intensity of Nt17-POPC complex ions was much higher (15.05%, **Fig. 4.13e**). This large population was predominately comprised of monomeric Nt17-POPC complexes (relative intensity of 14.21%, **Fig. 4.13f**). While eleven distinct monomeric Nt17-lipid ions were observed with both DMPC and POPC, monomeric Nt17-POPC complexes only contained up to five POPC lipids. Dimeric and trimeric Nt17-POPC complexes were less abundant (**Fig. 4.13g-h**). Nt17 dimers complexed with either one, two, three, or five POPC lipids; Nt17 trimers complexed with one, five, or six POPC lipids. Nt17-DOPC complexes were the least abundant, reaching a total relative abundance of only 2.38% (**Fig. 4.13e**). The most abundant complex consisted of monomeric Nt17 with one DOPC lipid (1.17%, **Fig. 4.13f**). Dimeric and trimeric Nt17-DOPC complexes were observed with up to six DOPC lipids or five DOPC lipids respectively (**Fig. 4.13g-h**). Despite the low abundance, Nt17 had the highest number of unique complex combinations with DOPC. Previously it has been shown that gas-phase Nt17-DOPC ion complexes can be weak and dissociate with little energy input prior to mass analysis; the low abundances reported here are consistent with the earlier results [42]. Therefore, even though the Nt17-DOPC ion complexes had the lowest abundance, this may be due to their stability in the gas phase. To complement these native MS studies, the interaction of Nt17 with DMPC, POPC, or DOPC was investigated using weighted ensemble molecular

dynamics (MD) simulations. Although the timescales for the MD simulations are orders of magnitude shorter than the MS experiments, they provide a much more detailed picture of how Nt17 binds to each lipid type. In order to facilitate as direct a comparison as possible, we modeled Nt17 in a coiled conformation above the bilayer surface, allowing the peptide to spontaneously bind and dissociate. This approach is consistent with the ability of Nt17 to function as a membrane-active peptide that is governed by partitioning-folding coupling, in which peptides initially bind to a bilayer surface, partition into the headgroup region to sequester from bulk solvent, and fold into a more energetically favorable α-helix [51]. The endpoint for this process with Nt17 is represented by the structures that were obtained by Bechinger and coworkers, where the hydrophobic face of the amphiphilic helix was clearly buried in the hydrophobic region of the DPC micelle [16]. Convergence analysis of our simulations shows that regardless of lipid type, Nt17 weakly binds to the bilayer surface, with a slight shift towards more tightly bound complexes in the following order from least to greatest: DMPC $>$ POPC $>$ DOPC (**Fig. 4.15**). A rough correlation exists between the abundance of Nt17/lipid complexes observed by MS and the distribution of proteolipid contacts observed in the MD simulations (**Fig. 4.13i**). Each system has a maximum around 1100 contacts. Nt17 clearly binds most weakly to DMPC, with a narrow gaussian distribution. POPC, the next-largest system by area-per-lipid, has a bimodal distribution with a second maxima at 1750 contacts. For DOPC, instead of a second maxima there is a broad plateau, ranging from 1750 to 2500 contacts. Based on the monomer data in our MS results, there is a slight difference in the trends in lipid binding to Nt17 compared to our MD simulations. This difference could be attributed to the nature of the two techniques: MS is capturing weak peptide-lipid interactions that must remain stable in the gas phase, whereas our MD simulations are capturing intermittent interactions in the liquid phase. Based on the clear separation in quality of binding between the three lipid systems, the MD simulations were further explored to better understand the binding of Nt17 to the membrane surface.

4.3.4 Effect of lipid type on binding of Nt17.

When examining overall binding, Nt17 possesses distinct orientations for each lipid type (**Fig. 4.16** and **Fig. 4.17**). Bound complexes to DMPC are dominated by partitioning of the N-terminus into the bilayer surface. This Nt17 segment is a mixture of nonpolar and polar residues (MATL), which individually are not too bulky to access the bilayer interior

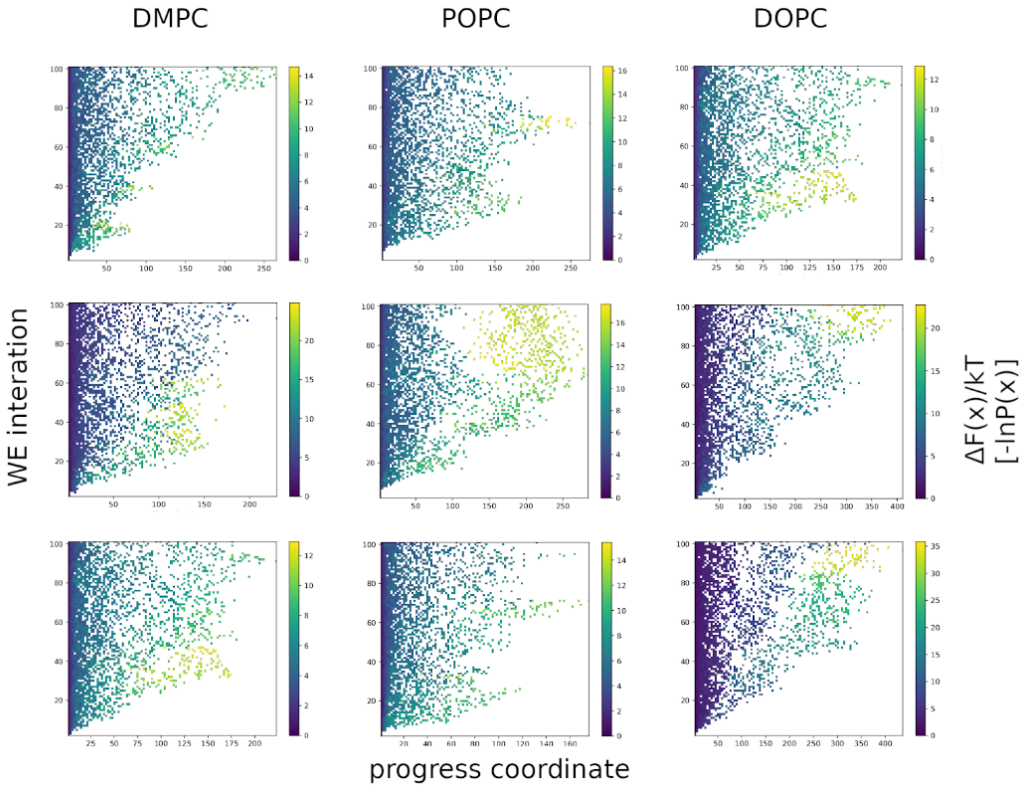

Figure 4.15: Nt17 transient binding largely persists as WESTPA simulations progress. Change in progress coordinate as a function of WESTPA iteration for each individual WESTPA run for DMPC (left), POPC (middle), and DOPC (right).

sandwiched between two relatively densely-packed headgroup regions (ADMPC = 60.6 $Å^2$ / lipid)[52] (**Fig. 4.16**). The Nt17 C-terminus is mainly dissociated from the DMPC bilayer, with a small sub-population that is less than 10 Åfrom the surface. In contrast, Nt17 strongly binds POPC via the C-terminus; F17 partitions almost 5 Åinto the headgroup region, with an extensive segment (F11, E12, S13, L14, K15, and S16) remaining less than 10 Åfrom the bilayer surface. This strong interaction potentially stems from the opportunity for a bulkier, nonpolar sidechain with conformational flexibility (i.e., F17) probing larger defects associated with a larger area per lipid (68.3 $Å^2$/ lipid) [53]. Although there is a small fraction of the middle segment of Nt17 within 15 Åof the bilayer, in general, the N-terminal portion is dissociated from the membrane, beginning with A10 and getting progressively further from the bilayer as one proceeds towards M1. For both DMPC and POPC, one of the Nt17 termini binds the membrane in a conformation that is orthogonal to the bilayer surface

(**Fig. 4.16a,c**). However, with DOPC, nearly all Nt17 residues interact with the membrane (**Fig. 4.16b**), leading to a parallel orientation of the Nt17 principal axis with respect to the bilayer. Additionally, no single residue binds most strongly, likely stemming from DOPC having the largest area per lipid of all the systems tested (72.2 $Å^2$ / lipid) [54] and leading to a larger number of defects that can be utilized by Nt17. Previous MD simulation studies on Nt17 interactions with lipid bilayers have shown that Nt17 can bind via multiple amino acid residues (e.g., L4, L7, F11, and F17) [55]. These multiple binding modes appear to be linked to orientation of the peptide: in each of our lipid systems we observe at least two populations, with bimodal distributions for both POPC and DOPC (**Fig. 4.16b,c**).

4.3.5 Relationship between sidechain orientation and binding of Nt17

Closer examination of residue-specific interactions of Nt17 with the bilayer surface reveals that, in most instances, sidechain orientation lies either parallel or away from the bilayer. For DMPC, T3 and K6 are notably parallel to the bilayer (90 deg), likely ruling out the potential for non-bonded interactions (hydrogen bonding for T3 and salt bridge formation for K6) with PC headgroups to drive formation of stable binding conformations (**Fig. 4.18** and **Fig. 4.19**). Aromatic (F11 and F17) and charged (K15) residues in the C-terminal half of Nt17 occupy two to three areas of conformational space, indicating that the dissociated portion of Nt17 is restricted to regions that allow it to bind via the N-terminus. Binding of Nt17 to POPC is much different: F17 partitions into the headgroup region nearly antiparallel to the membrane normal between acyl chains near the triglyceride backbone. This serves to restrict F11 and K15 (proximal to F17) parallel to the membrane normal (90 deg), whereas other residues (T3, K6, and K9) broadly sample orientations ranging from parallel (0 deg) to antiparallel (180 deg) (**Fig. 4.18**). For DOPC, there is a correlation between the distance of a given residue from the bilayer surface and sidechain orientation for nonpolar or slightly polar residues. Specifically, L7, A10, S13, and L14 sample conformations partitioned into the bilayer (**Fig. 4.19**). Additionally, there is a noticeable population for each lysine residue (K6, K9, and K15) oriented towards the bilayer (**Fig. 4.18**) that is close enough to form a salt bridge with the negatively charged phosphate moiety in the PC headgroup (**Fig. 4.20d**). This, in conjunction with the overall closer positioning of the peptide, suggests binding of Nt17 to DOPC is driven by a combination of charge-charge interactions and greater accessibility for peptide partitioning. With DOPC, a threshold has been reached with respect to area per

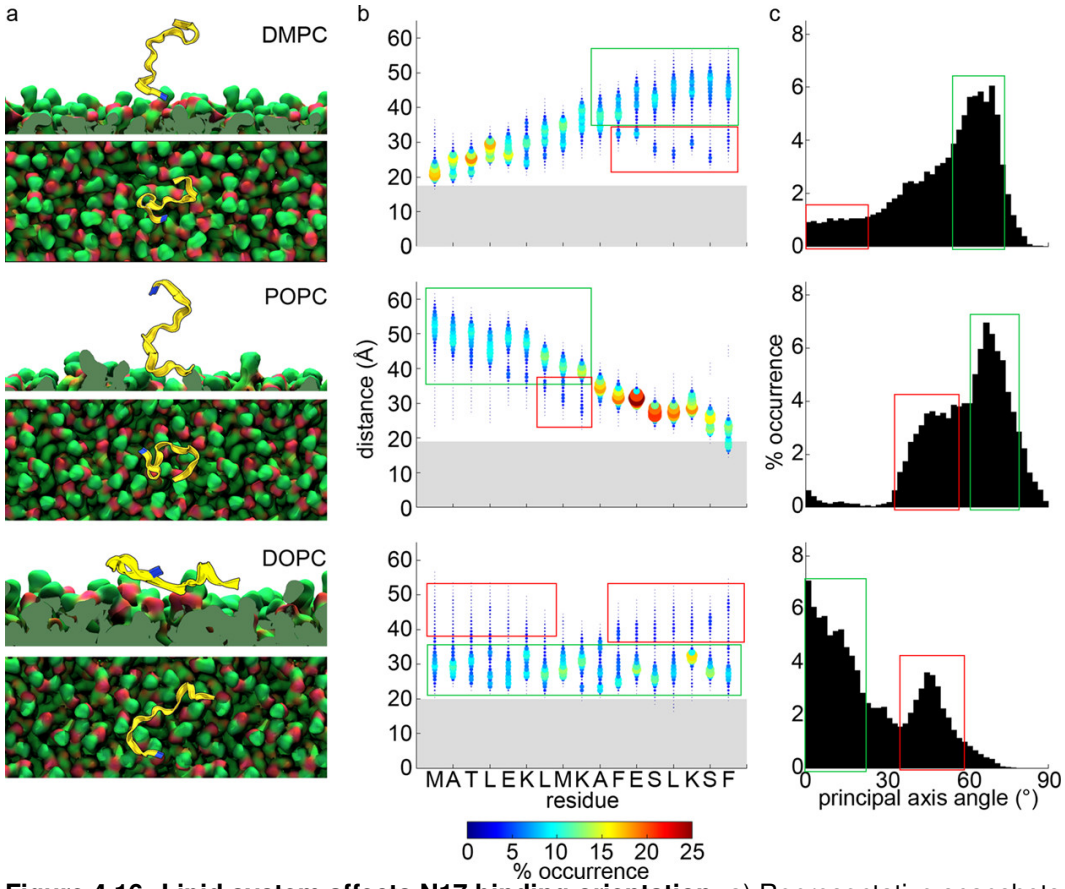

Figure 4.16: Lipid system affects N17 binding orientation. a) Representative snapshots of the bound state for each lipid system from weighted ensemble simulations. Yellow ribbon: Nt17, blue ribbon: first residue of Nt17, red/green: lipid membrane. Water and ions omitted for clarity. b) Binned z-distance between the geometric centers of the membrane and residues of Nt17 for frames in the bound state. Point size and color correspond to percent occurrence. c) Binned angle between the principal axis of the peptide and the bilayer surface. For b and c, green boxes represent data associated with the most likely orientation of Nt17 on the bilayer, shown in panel a. Red boxes represent data associated with a secondary population, shown in **Fig. 3.17**.

lipid and Nt17 binding, in that there is a sufficient population and size of defects to allow the entire peptide to partition into the bilayer.

4.3.6 The role of protrusions and defects in binding and stabilization of Nt17

When examining the behavior of the bilayer surface in the upper leaflet, each lipid system possesses regions of protrusions and defects. In general, regions of protrusions or of defects tend to group together (**Fig. 4.20a**). POPC has more well-defined and broader

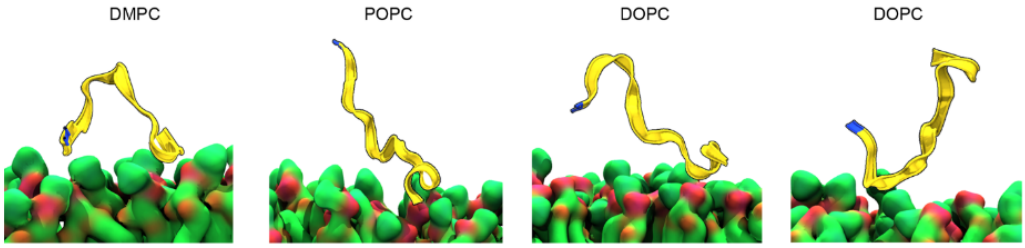

Figure 4.17: Representative snapshots of less commonly observed bound structures for each lipid system from weighted ensemble simulations. Two unique conformations are shown for DOPC.

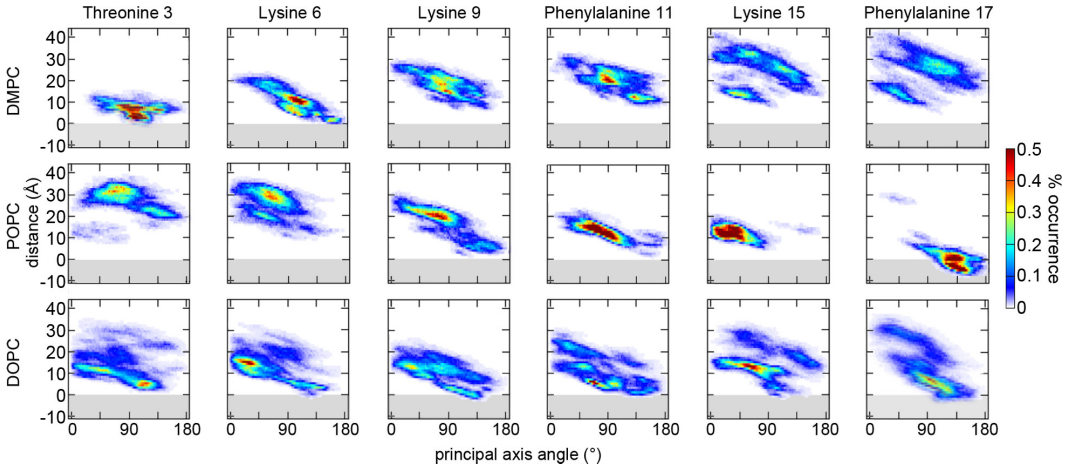

Figure 4.18: Nt17 binding can be facilitated via partitioning of amino acid sidechains into bilayer systems with higher areas per lipid. Correlation plot between the binned distribution of sidechain angles of select residues in Nt17 and the binned distribution of the z-distance between each respective residue's terminal atom and the plane of the P atoms in the lipid headgroups. Sidechain angles were defined as the angle between the vector formed by the Cα carbon and the respective terminal heavy atom and the membrane normal. 0 deg: parallel to bilayer normal, 90 deg: parallel to bilayer surface; 180 deg: antiparallel to bilayer normal (i.e., pointing into the bilayer). Top row: DMPC; middle row: POPC; bottom row: DOPC.

regions of defects compared to DMPC, which correlates with the more effective binding of Nt17. DOPC had the fewest regions of protrusions and defects, likely contributing to the global but weaker interactions of Nt17. Quantification of the total number of protrusions and defects in the bilayer surface provides a more well-defined description of what occurs when Nt17 binds to a membrane. In the absence of Nt17, the three lipid systems have similar distributions of protrusions ("above", **Fig. 4.20b**) but slightly different distribution of defects ("below", **Fig. 4.20b**). Upon Nt17 binding, two changes take place in the bilayer

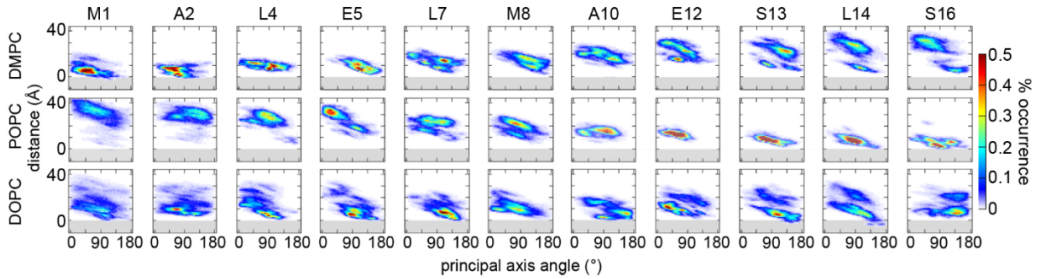

Figure 4.19: Nt17 binding can be facilitated via partitioning of amino acid sidechains into bilayer systems with higher areas per lipid. Correlation plot between the binned distribution of sidechain angles of select residues in Nt17 and the binned distribution of the z-distance between each respective residue's terminal atom and the plane of the P atoms in the lipid headgroups. (0 deg: parallel to bilayer normal, 90 deg: parallel to bilayer surface; 180 deg: antiparallel to bilayer normal (i.e., pointing into the bilayer)). Top row: DMPC; middle row: POPC; bottom row: DOPC.

surface: 1) the population of protrusions and defects decreases due to partitioning of the peptide, and 2) the distribution of protrusions and of defects between lipid systems becomes separable. With protrusions, the distributions remain consistent across all lipid systems; in the case of defects, the distributions arrange according to the corresponding areas per lipid. The difference in the shifts in the distribution of protrusions indicates that the initial interactions of Nt17 with the bilayer surface (i.e., with protrusions) is entirely non-specific with respect to lipid type. In contrast, the number and distribution of defects is most greatly affected by the ability of the bilayer to accommodate partitioning of Nt17 into the headgroup region. Since DMPC is the most tightly packed bilayer in the lateral plane of the membrane, both by area per lipid and by the number of defects, it is incapable of easily changing its lateral compressibility, a prerequisite for surface binding and partitioning of a peptide. This underscores why POPC and DOPC have progressively larger distributions of defects. Future studies will determine the energetic contributions of defects to binding of Nt17.

Non-bonded interactions also play a role in stable binding of Nt17 to the membrane. Although we did not observe the extensive formation of an α-helix that was reported by Bechinger and coworkers [16], Nt17 undergoes partial helical folding in POPC and DOPC (**Fig. 4.20c**). Folding is centered around L14 and K15, with about 20-25% helicity. Additionally, an intramolecular salt bridge frequently exists between E12 and K15. i + 3 salt bridges can contribute to α-helix formation in short peptides [56]; in this case, the salt bridge formation appears to work cooperatively with headgroup accessibility in facilitating

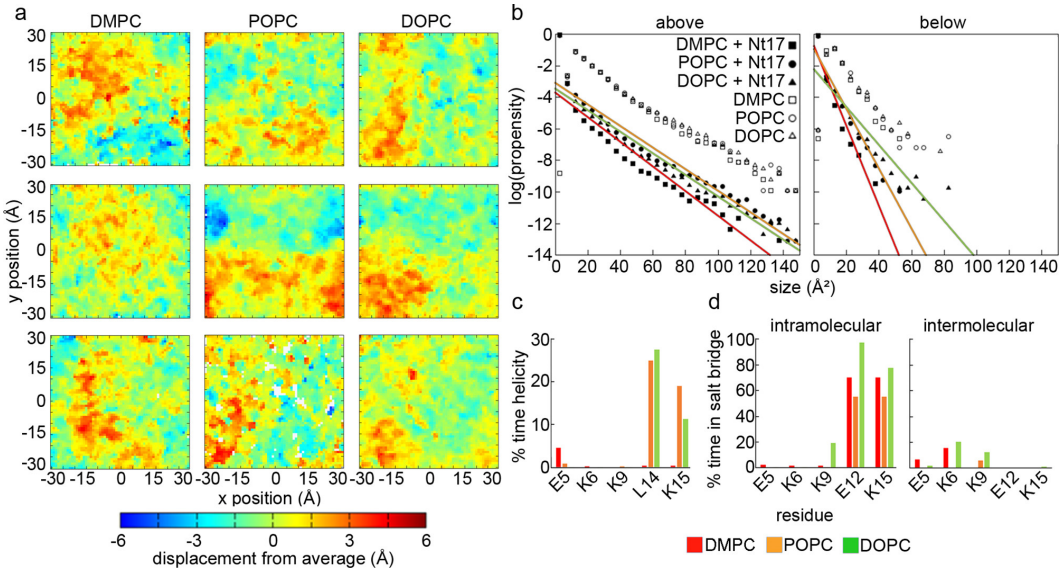

Figure 4.20: Formation of stable proteolipid complexes occurs via contributions from both redistribution of bilayer deformations and stable intermediates of Nt17. a) Heat map of the average displacement of headgroups in the upper leaflet for all three trajectories of DMPC (left), POPC (middle), and DOPC (right). b) Logarithmic distribution of protrusions ("above", left) and defects ("below", right) for lipid-only systems and corresponding lipid systems with Nt17. c) Percent helicity for Nt17 in DMPC, POPC, and DOPC. A residue was considered helical if both the prior and subsequent residues also have helical phi/psi angles. (Residues that were not helical were omitted for clarity.) d) Percent of total simulation time a charged residue in Nt17 was involved in formation of a salt bridge. Intramolecular salt bridges are those formed between residues in Nt17, while intermolecular salt bridges are those formed between charged residues in Nt17 and the phosphate moiety of the PC headgroup for DMPC, POPC, and DOPC.

Nt17 partitioning into the bilayer.

4.4 CONCLUSION

A number of cytotoxic features of HD are associated with the localization of htt to lipid membranes [23, 24, 57]. In this study, the impact of lipid tail saturation on htt-exon1(46Q) interaction with lipid membranes and subsequent consequences on aggregation were determined. The different lipids influenced htt aggregation to varying extents, with DMPC accelerating fibrillization, POPC having a minimal impact, and DOPC decreasing fibrillization. However, the extent of htt association with each lipid system did not correlate with the extent of fibrillization, indicating that the membrane environment provided by different lipid systems influences the membrane activity and aggregation of htt. The assays (ThT and AFM) used to assess htt aggregation measure it throughout the bulk of the incubations. That is, there is no

distinction made between aggregation in solution versus aggregation occurring directly on a vesicle surface. Aggregation directly on lipid membranes often exhibit unique aggregation pathways compared to those observed in aqueous solutions [25]. In addition, the interaction of the htt-exon1 lipid binding domain, i.e. Nt17, with the lipid systems demonstrated that the match between defect and hydrophobic amino acid size plays a major role in orientation of the peptide on the bilayer. These early interactions between Nt17 and each lipid system manifest in down-stream impacts on htt-exon1(46Q) aggregation. That is, these differing orientations could influence if membrane association inhibits or promotes aggregation directly on the surface. The interplay between htt associating and disassociating from membranes can also influence aggregation by potentially promoting aggregation prone or resistant conformation. In solution, the Nt17 domain of htt transiently samples α-helical conformations that are associated with early oligomer formation [14]. Binding lipids also promotes helical structure in Nt17 [25]. Brief binding may promote helical content in Nt17 upon dissociation from lipid membranes [42], enhancing aggregation efficiency. In short, there are numerous competing processes, i.e. aggregation in solution, aggregation on the lipid surface, association/disassociation from the membrane, orientation on the membrane, that all can influence net aggregation. This complexity underscores the apparent disconnect between the impact of a particular lipid on aggregation and the extent of the association of htt with that lipid system.

Components within the cellular environment, particularly lipid membranes, modify amyloid formation, promoting or stabilizing specific aggregate species [58]. Htt undergoes unique aggregation mechanisms on lipid membranes compared to bulk solution [25], and this appears dependent on membrane composition (32, 33, 65). Beyond the three simple lipid systems investigated here, more complicated lipid systems also modify htt aggregation to various extents [26, 28, 59, 60]. Total brain lipid extract (TBLE) stabilizes oligomers of truncated polyQ peptides that contain Nt17 [14] and impedes fibrillization of htt-exon1 [26], and this impact is further modified by the exogenous addition of cholesterol [28], sphingomyelin, and GM1 [27]. In contrast, htt fibrillization is accelerated in the presence of POPC/POPS (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoserine) through a unique Nt17-mediated mechanism involving membrane anchoring and two-dimensional diffusion [25]. Based on 3D HNCO NMR studies using a small peptide with Nt17 and a polyQ domain of 17 residues, the interaction of Nt17 with lipid membranes is dynamic; however, there were clear

differences in the interaction of Nt17 with 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE)/1,2,-dioleoyl-sn-glycero-3-phosphoserine (DOPS)/DOPC compared to brain extracts [61]. In this regard, the varied orientations of Nt17 observed here by MD on different lipid systems impact htt anchoring and diffusion along the membrane. These initial changes in Nt17-lipid binding manifested in altered abundances of Nt17/lipid complexes and lead to altered aggregation. Collectively, these observations indicate that the initial htt/lipid interactions profoundly impact subsequent aggregation.

While the extent of htt fibrillization correlated with lipid tail saturation (DMPC $>$ POPC $>$ DOPC; determined by ThT and AFM assays), the htt-membrane association (POPC $>$ DMPC $>$ DOPC; measured by PDA and MS) did not, suggesting a more complex relationship between htt-lipid interaction and subsequent aggregation. Through MD simulations, clear differences occur in the initial Nt17-lipid interactions that likely influence downstream aggregation. While a significant gap in timescale persists, Nt17/lipid contact density measured in MD simulations are consistent with the populations of Nt17/lipid complexes observed by MS with all three lipids. As Nt17 peptides did not contain polyQ tracts, the MD and MS studies cannot provide insight into fibrillization; however, oligomerization is driven by Nt17 [14, 62, 15]. Therefore, changes in Nt17-lipid interaction should impact downstream htt oligomerization, and all three lipid systems reduced the heterogeneity of oligomer populations compared to aggregation in the absence of lipids. The impact of lipids on htt aggregation is significant as cellular dysfunction and toxicity are linked to both htt oligomers [8, 10] and fibrils [12, 63]. Furthermore, the formation or stabilization of specific aggregate species by lipid membranes could promote the toxic disruption of membranes and organelles observed in HD [18, 64]. This suggests that targeting the initial htt/lipid interaction may represent an upstream therapeutic target.

The properties determining selectivity of AHs (like Nt17) for specific membranes are often related to curvature-induced membrane defects [30, 65, 66]. As curvature enhances the membrane binding of Nt17 [35, 67], defect sensing represents a plausible mechanism associated with htt membrane affinity. Here, the population of available membrane defects correlates well with the affinity of Nt17 for each lipid system, and the orientation of Nt17 on membranes appears to be predicated by matching residue and defect sizes. Such a scenario has been observed with other AH lipid binding motifs. For example, the amphipathic lipid-packing sensor (ALPS) inserts large hydrophobic residues preferentially into membranes

with larger membrane defects [68]. Nt17 exhibits many similarities with ALPS motifs, including a structural transition from intrinsically disordered to an AH in the presence of lipids [17, 16, 69], increased association with membranes of lower cholesterol content [28, 33] or increased curvature [65, 35] and, as shown here, a preference for insertion into bilayers comprised of monounsaturated phospholipids [33, 70]. Despite sharing functional similarities, Nt17 and ALPS motifs vary in sequence; while ALPS motifs contain a hydrophobic side comprised primarily of large amino acids (F, L, W) [30, 68], the hydrophobic face of Nt17 contains several smaller amino acids (A, T) near the N-terminus. The lack of variety in ALPS motifs hydrophobic residue sizes leads to the initial insertion of hydrophobic residues into the bilayer in a random manner [29]. The mechanism of Nt17 membrane binding appears more complex, with both hydrophobic residues size and defect size determining the initial orientation and selective binding of the motif. This defect mechanism gains further importance for selective binding of AHs given that lipid composition and membrane curvature synergistically determine defect size and distribution [71].

Upon exposure to tightly packed DMPC (saturated tails) bilayers with a relatively small area per lipid, the smaller, N-terminal residues (MATL) partitioned into the bilayer. POPC (monounsaturated) bilayers display an increased area per lipid and occurrence of defects. The interaction between Nt17 and the POPC bilayer was driven by the partitioning of the larger, C-terminal residues (LKSF). While the smaller N-terminal residues could insert into the larger defects present on POPC bilayers, Nt17 preferred a C-terminus approach, indicating a thermodynamic advantage to sequestering larger hydrophobic residues. DOPC (polyunsaturated) bilayers had a higher propensity to form defects with a broader distribution of sizes. Due to this, residues along the entire length of Nt17 exhibited more frequent, but weaker binding and Nt17 sampled a broader variety of orientations on the bilayer surface. This behavior explains why complexes observed with MS had larger DOPC content but were significantly less abundant overall than complexes formed with DMPC or POPC.

Beyond the impact of lipids on htt aggregation, htt and its aggregate forms can induce physical changes in membranes. Many amyloid-forming proteins have a deleterious impact on membrane integrity that leads to dysfunction, organelle disruption, and cell death by specifically targeting bilayers [18, 58]. For example, Aβ, IAPP, and htt all distort membrane morphology and rigidity of model lipid bilayers [72]. Specifically, mutant htt is associated with membrane degradation of the ER, nuclear envelope, and mitochondria [23, 24, 57].

With the complex nature of amyloid formation, assigning specific aggregate species to a particular toxic mechanism is challenging. This is further complicated by the dynamic nature of aggregation in which the process of aggregation potentially plays a role rather than a specific aggregate type. Fundamental understanding of the link between the initial protein/lipid interaction, subsequent aggregation, and impact on membrane integrity will provide a clearer, complete picture of these phenomena.

Author Contributions: S.V. and J.L. conceived the project. M.B. performed ThT assays, PDA assays, and ESI-MS experiments. M.B. and S.G. performed AFM experiments. N.F. performed all MD experiments. All authors contributed to analyzing the data. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Acknowledgements: This project was funded in part by the National Institutes of Health grants #R15NS090380 (J.L.), #R01GM114494 (S.V.), and #GMI20676 (B.M.). M.B. was supported by a Ruby Distinguished Doctoral Fellowship from WVU. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

4.5 BIBLIOGRAPHY

[1] J. B. Penney, J. P. Vonsattel, M. E. MacDonald, J. F. Gusella, and R. H. Myers. “CAG repeat number governs the development rate of pathology in Huntington’s disease”. eng. In: *Ann neurol* 41.5 (May 1997), pp. 689–692. ISSN: 0364-5134. DOI: 10.1002/ana.410410521.

[2] K. Kar, M. Jayaraman, B. Sahoo, R. Kodali, and R. Wetzel. “Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent”. en. In: *Nat struct mol biol* 18.3 (Mar. 2011). Number: 3 Publisher: Nature Publishing Group, pp. 328–336. ISSN: 1545-9985. DOI: 10.1038/nsmb.1992. URL: https://www.nature.com/articles/nsmb.1992 (visited on 06/23/2022).

[3] G. P. Lotz and J. Legleiter. “The role of amyloidogenic protein oligomerization in neurodegenerative disease”. eng. In: *J mol med (berl)* 91.6 (June 2013), pp. 653–664. ISSN: 1432-1440. DOI: 10.1007/s00109-013-1025-1.

[4] S. Bonfanti, M. C. Lionetti, M. R. Fumagalli, V. R. Chirasani, G. Tiana, N. V. Dokholyan, S. Zapperi, and C. A. M. La Porta. “Molecular mechanisms of heterogeneous oligomerization of huntingtin proteins”. en. In: *Sci rep* 9.1 (May 2019). Number: 1 Publisher: Nature Publishing Group, p. 7615. ISSN: 2045-2322. DOI: 10.1038/s41598-019-44151-0. URL: https://www.nature.com/articles/s41598-019-44151-0 (visited on 06/23/2022).

[5] K. A. Burke, J. Godbey, and J. Legleiter. “Assessing mutant huntingtin fragment and polyglutamine aggregation by atomic force microscopy”. eng. In: *Methods* 53.3 (Mar. 2011), pp. 275–284. ISSN: 1095-9130. DOI: 10.1016/j.ymeth.2010.12.028.

[6] M. DiFiglia, E. Sapp, K. O. Chase, S. W. Davies, G. P. Bates, J. P. Vonsattel, and N. Aronin. “Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain”. eng. In: *Science* 277.5334 (Sept. 1997), pp. 1990–1993. ISSN: 0036-8075. DOI: 10.1126/science.277.5334.1990.

[7] M. Arrasate, S. Mitra, E. S. Schweitzer, M. R. Segal, and S. Finkbeiner. “Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death”. en. In: *Nature* 431.7010 (Oct. 2004). Number: 7010 Publisher: Nature Publishing Group, pp. 805–810. ISSN: 1476-4687. DOI: 10.1038/nature02998. URL: https://www.nature.com/articles/nature02998 (visited on 06/23/2022).

[8] Y. E. Kim, F. Hosp, F. Frottin, H. Ge, M. Mann, M. Hayer-Hartl, and F. U. Hartl. “Soluble Oligomers of PolyQ-Expanded Huntingtin Target a Multiplicity of Key Cellular Factors”. eng. In: *Mol cell* 63.6 (Sept. 2016), pp. 951–964. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2016.07.022.

[9] M. A. Olshina, L. M. Angley, Y. M. Ramdzan, J. Tang, M. F. Bailey, A. F. Hill, and D. M. Hatters. “Tracking mutant huntingtin aggregation kinetics in cells reveals three major populations that include an invariant oligomer pool”. eng. In: *J biol chem* 285.28 (July 2010), pp. 21807–21816. ISSN: 1083-351X. DOI: 10.1074/jbc.M109.084434.

[10] P. Lajoie and E. L. Snapp. “Formation and toxicity of soluble polyglutamine oligomers in living cells”. eng. In: *Plos one* 5.12 (Dec. 2010), e15245. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0015245.

[11] M. Lu, L. Banetta, L. J. Young, E. J. Smith, G. P. Bates, A. Zaccone, G. S. Kaminski Schierle, A. Tunnacliffe, and C. F. Kaminski. “Live-cell super-resolution microscopy reveals a primary role for diffusion in polyglutamine-driven aggresome assembly”. eng. In: *J biol chem* 294.1 (Jan. 2019), pp. 257–268. ISSN: 1083-351X. DOI: 10.1074/jbc.RA118.003500.

[12] K. W. Drombosky, S. Rode, R. Kodali, T. C. Jacob, M. J. Palladino, and R. Wetzel. “Mutational analysis implicates the amyloid fibril as the toxic entity in Huntington’s disease”. eng. In: *Neurobiol dis* 120 (Dec. 2018), pp. 126–138. ISSN: 1095-953X. DOI: 10.1016/j.nbd.2018.08.019.

[13] F. J. B. Bäuerlein, I. Saha, A. Mishra, M. Kalemanov, A. Martínez-Sánchez, R. Klein, I. Dudanova, M. S. Hipp, F. U. Hartl, W. Baumeister, and R. Fernández-Busnadiego. “In Situ Architecture and Cellular Interactions of PolyQ Inclusions”. eng. In: *Cell* 171.1 (Sept. 2017), 179–187.e10. ISSN: 1097-4172. DOI: 10.1016/j.cell.2017.08.009.

[14] K. A. Burke, K. J. Kauffman, C. S. Umbaugh, S. L. Frey, and J. Legleiter. “The interaction of polyglutamine peptides with lipid membranes is regulated by flanking sequences associated with huntingtin”. eng. In: *J biol chem* 288.21 (May 2013), pp. 14993–15005. ISSN: 1083-351X. DOI: 10.1074/jbc.M112.446237.

[15] R. Truant, R. Atwal, and A. Burtnik. “Hypothesis: Huntingtin may function in membrane association and vesicular trafficking”. eng. In: *Biochem cell biol* 84.6 (Dec. 2006), pp. 912–917. ISSN: 0829-8211. DOI: 10.1139/o06-181.

[16] J. Velier, M. Kim, C. Schwarz, T. W. Kim, E. Sapp, K. Chase, N. Aronin, and M. DiFiglia. “Wild-type and mutant huntingtins function in vesicle trafficking in the secretory and endocytic pathways”. eng. In: *Exp neurol* 152.1 (July 1998), pp. 34–40. ISSN: 0014-4886. DOI: 10.1006/exnr.1998.6832.

[17] K. B. Kegel, E. Sapp, J. Yoder, B. Cuiffo, L. Sobin, Y. J. Kim, Z.-H. Qin, M. R. Hayden, N. Aronin, D. L. Scott, G. Isenberg, W. H. Goldmann, and M. DiFiglia. “Huntingtin associates with acidic phospholipids at the plasma membrane”. eng. In: *J biol chem* 280.43 (Oct. 2005), pp. 36464–36473. ISSN: 0021-9258. DOI: 10.1074/jbc.M503672200.

[18] J. Suopanki, C. Götz, G. Lutsch, J. Schiller, P. Harjes, A. Herrmann, and E. E. Wanker. “Interaction of huntingtin fragments with brain membranes–clues to early dysfunction

in Huntington's disease". eng. In: *J neurochem* 96.3 (Feb. 2006), pp. 870–884. ISSN: 0022-3042. DOI: 10.1111/j.1471-4159.2005.03620.x.

[19] K. A. Burke, K. M. Hensal, C. S. Umbaugh, M. Chaibva, and J. Legleiter. "Huntingtin disrupts lipid bilayers in a polyQ-length dependent manner". eng. In: *Biochim biophys acta* 1828.8 (Aug. 2013), pp. 1953–1961. ISSN: 0006-3002. DOI: 10.1016/j.bbamem.2013.04.025.

[20] Z.-H. Qin, Y. Wang, E. Sapp, B. Cuiffo, E. Wanker, M. R. Hayden, K. B. Kegel, N. Aronin, and M. DiFiglia. "Huntingtin bodies sequester vesicle-associated proteins by a polyproline-dependent interaction". eng. In: *J neurosci* 24.1 (Jan. 2004), pp. 269–281. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.1409-03.2004.

[21] Q. Ruan, M. Lesort, M. E. MacDonald, and G. V. W. Johnson. "Striatal cells from mutant huntingtin knock-in mice are selectively vulnerable to mitochondrial complex II inhibitor-induced cell death through a non-apoptotic pathway". eng. In: *Hum mol genet* 13.7 (Apr. 2004), pp. 669–681. ISSN: 0964-6906. DOI: 10.1093/hmg/ddh082.

[22] U. P. Shirendeb, M. J. Calkins, M. Manczak, V. Anekonda, B. Dufour, J. L. McBride, P. Mao, and P. H. Reddy. "Mutant huntingtin's interaction with mitochondrial protein Drp1 impairs mitochondrial biogenesis and causes defective axonal transport and synaptic degeneration in Huntington's disease". eng. In: *Hum mol genet* 21.2 (Jan. 2012), pp. 406–420. ISSN: 1460-2083. DOI: 10.1093/hmg/ddr475.

[23] M. Ueda, S. Li, M. Itoh, M.-X. Wang, M. Hayakawa, S. Islam, n. Tana, K. Nakagawa, H. Chen, and T. Nakagawa. "Expanded polyglutamine embedded in the endoplasmic reticulum causes membrane distortion and coincides with Bax insertion". eng. In: *Biochem biophys res commun* 474.2 (May 2016), pp. 259–263. ISSN: 1090-2104. DOI: 10.1016/j.bbrc.2016.04.034.

[24] K.-Y. Liu, Y.-C. Shyu, B. A. Barbaro, Y.-T. Lin, Y. Chern, L. M. Thompson, C.-K. James Shen, and J. L. Marsh. "Disruption of the nuclear membrane by perinuclear inclusions of mutant huntingtin causes cell-cycle re-entry and striatal cell death in mouse and cell models of Huntington's disease". eng. In: *Hum mol genet* 24.6 (Mar. 2015), pp. 1602–1616. ISSN: 1460-2083. DOI: 10.1093/hmg/ddu574.

[25] N. K. Pandey, J. M. Isas, A. Rawat, R. V. Lee, J. Langen, P. Pandey, and R. Langen. “The 17-residue-long N terminus in huntingtin controls stepwise aggregation in solution and on membranes via different mechanisms”. eng. In: *J biol chem* 293.7 (Feb. 2018), pp. 2597–2605. ISSN: 1083-351X. DOI: 10.1074/jbc.M117.813667.

[26] M. Beasley, A. R. Stonebraker, I. Hasan, K. L. Kapp, B. J. Liang, G. Agarwal, S. Groover, F. Sedighi, and J. Legleiter. “Lipid Membranes Influence the Ability of Small Molecules To Inhibit Huntingtin Fibrillization”. In: *Biochemistry* 58.43 (Oct. 2019). Publisher: American Chemical Society, pp. 4361–4373. ISSN: 0006-2960. DOI: 10.1021/acs.biochem.9b00739. URL: https://doi.org/10.1021/acs.biochem.9b00739 (visited on 06/23/2022).

[27] M. Chaibva, X. Gao, P. Jain, W. A. Campbell, S. L. Frey, and J. Legleiter. “Sphingomyelin and GM1 Influence Huntingtin Binding to, Disruption of, and Aggregation on Lipid Membranes”. In: *Acs omega* 3.1 (Jan. 2018). Publisher: American Chemical Society, pp. 273–285. DOI: 10.1021/acsomega.7b01472. URL: https://doi.org/10.1021/acsomega.7b01472 (visited on 06/23/2022).

[28] X. Gao, W. A. Campbell, M. Chaibva, P. Jain, A. E. Leslie, S. L. Frey, and J. Legleiter. “Cholesterol Modifies Huntingtin Binding to, Disruption of, and Aggregation on Lipid Membranes”. eng. In: *Biochemistry* 55.1 (Jan. 2016), pp. 92–102. ISSN: 1520-4995. DOI: 10.1021/acs.biochem.5b00900.

[29] M. Giménez-Andrés, A. Čopič, and B. Antonny. “The Many Faces of Amphipathic Helices”. eng. In: *Biomolecules* 8.3 (July 2018), E45. ISSN: 2218-273X. DOI: 10.3390/biom8030045.

[30] G. Drin, J.-F. Casella, R. Gautier, T. Boehmer, T. U. Schwartz, and B. Antonny. “A general amphipathic α-helical motif for sensing membrane curvature”. en. In: *Nat struct mol biol* 14.2 (Feb. 2007). Number: 2 Publisher: Nature Publishing Group, pp. 138–146. ISSN: 1545-9985. DOI: 10.1038/nsmb1194. URL: https://www.nature.com/articles/nsmb1194 (visited on 06/23/2022).

[31] M. Masuda, S. Takeda, M. Sone, T. Ohki, H. Mori, Y. Kamioka, and N. Mochizuki. “Endophilin BAR domain drives membrane curvature by two newly identified structure-based mechanisms”. In: *Embo j* 25.12 (June 2006), pp. 2889–2897. ISSN: 0261-4189.

DOI: 10.1038/sj.emboj.7601176. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1500852/ (visited on 06/23/2022).

[32] C. M. Doucet, J. A. Talamas, and M. W. Hetzer. “Cell Cycle Dependent Differences in Nuclear Pore Complex Assembly in Metazoa”. In: *Cell* 141.6 (June 2010), pp. 1030–1041. ISSN: 0092-8674. DOI: 10.1016/j.cell.2010.04.036. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2887351/ (visited on 06/23/2022).

[33] I. M. Pranke, V. Morello, J. Bigay, K. Gibson, J.-M. Verbavatz, B. Antonny, and C. L. Jackson. “α-Synuclein and ALPS motifs are membrane curvature sensors whose contrasting chemistry mediates selective vesicle binding”. In: *J cell biol* 194.1 (July 2011), pp. 89–103. ISSN: 0021-9525. DOI: 10.1083/jcb.201011118. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3135411/ (visited on 06/23/2022).

[34] L. M. Thompson, C. T. Aiken, L. S. Kaltenbach, N. Agrawal, K. Illes, A. Khoshnan, M. Martinez-Vincente, M. Arrasate, J. G. O’Rourke, H. Khashwji, T. Lukacsovich, Y.-Z. Zhu, A. L. Lau, A. Massey, M. R. Hayden, S. O. Zeitlin, S. Finkbeiner, K. N. Green, F. M. LaFerla, G. Bates, L. Huang, P. H. Patterson, D. C. Lo, A. M. Cuervo, J. L. Marsh, and J. S. Steffan. “IKK phosphorylates Huntingtin and targets it for degradation by the proteasome and lysosome”. eng. In: *J cell biol* 187.7 (Dec. 2009), pp. 1083–1099. ISSN: 1540-8140. DOI: 10.1083/jcb.200909067.

[35] K. A. Burke and J. Legleiter. “Atomic force microscopy assays for evaluating polyglutamine aggregation in solution and on surfaces”. eng. In: *Methods mol biol* 1017 (2013), pp. 21–40. ISSN: 1940-6029. DOI: 10.1007/978-1-62703-438-8_2.

[36] M. Sokolovski, T. Sheynis, S. Kolusheva, and R. Jelinek. “Membrane interactions and lipid binding of casein oligomers and early aggregates”. en. In: *Biochimica et biophysica acta (bba) - biomembranes* 1778.10 (Oct. 2008), pp. 2341–2349. ISSN: 0005-2736. DOI: 10.1016/j.bbamem.2008.07.001. URL: https://www.sciencedirect.com/science/article/pii/S0005273608002083 (visited on 06/23/2022).

[37] F. Zheng, Z. Wu, and Y. Chen. “A quantitative method for the measurement of membrane affinity by polydiacetylene-based colorimetric assay”. eng. In: *Anal biochem* 420.2 (Jan. 2012), pp. 171–176. ISSN: 1096-0309. DOI: 10.1016/j.ab.2011.09.026.

[38] *Phase Transition Temperatures for Glycerophospholipids*. en-US. URL: https://avantilipids.com/tech-support/physical-properties/phase-transition-temps (visited on 06/23/2022).

[39] S. J. Kew and E. A. H. Hall. "pH Response of Carboxy-Terminated Colorimetric Polydiacetylene Vesicles". In: *Anal. chem.* 78.7 (Apr. 2006). Publisher: American Chemical Society, pp. 2231–2238. ISSN: 0003-2700. DOI: 10.1021/ac0517794. URL: https://doi.org/10.1021/ac0517794 (visited on 06/23/2022).

[40] J. Song, Q. Cheng, S. Kopta, and R. C. Stevens. "Modulating Artificial Membrane Morphology: pH-Induced Chromatic Transition and Nanostructural Transformation of a Bolaamphiphilic Conjugated Polymer from Blue Helical Ribbons to Red Nanofibers". In: *J. am. chem. soc.* 123.14 (Apr. 2001). Publisher: American Chemical Society, pp. 3205–3213. ISSN: 0002-7863. DOI: 10.1021/ja0035046. URL: https://doi.org/10.1021/ja0035046 (visited on 06/23/2022).

[41] M. Beasley, A. R. Stonebraker, and J. Legleiter. "Normalizing polydiacetylene colorimetric assays of vesicle binding across lipid systems". en. In: *Analytical biochemistry* 609 (Nov. 2020), p. 113864. ISSN: 0003-2697. DOI: 10.1016/j.ab.2020.113864. URL: https://www.sciencedirect.com/science/article/pii/S0003269720303961 (visited on 06/23/2022).

[42] A. K. Karanji, M. Beasley, D. Sharif, A. Ranjbaran, J. Legleiter, and S. J. Valentine. "Investigating the interactions of the first 17 amino acid residues of Huntingtin with lipid vesicles using mass spectrometry and molecular dynamics". eng. In: *J mass spectrom* 55.1 (Jan. 2020), e4470. ISSN: 1096-9888. DOI: 10.1002/jms.4470.

[43] W. Humphrey, A. Dalke, and K. Schulten. "VMD: Visual molecular dynamics". en. In: *Journal of molecular graphics* 14.1 (Feb. 1996), pp. 33–38. ISSN: 0263-7855. DOI: 10.1016/0263-7855(96)00018-5. URL: https://www.sciencedirect.com/science/article/pii/0263785596000185 (visited on 05/26/2022).

[44] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. "Scalable molecular dynamics with NAMD". eng. In: *J comput chem* 26.16 (Dec. 2005), pp. 1781–1802. ISSN: 0192-8651. DOI: 10.1002/jcc.20289.

[45] T. D. Romo and A. Grossfield. “LOOS: an extensible platform for the structural analysis of simulations”. eng. In: *Annu int conf ieee eng med biol soc* 2009 (2009), pp. 2332–2335. ISSN: 2375-7477. DOI: 10.1109/IEMBS.2009.5335065.

[46] T. D. Romo, N. Leioatts, and A. Grossfield. “Lightweight Object Oriented Structure analysis: Tools for building Tools to Analyze Molecular Dynamics Simulations”. In: *J comput chem* 35.32 (Dec. 2014), pp. 2305–2318. ISSN: 0192-8651. DOI: 10.1002/jcc.23753. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4227929/ (visited on 06/23/2022).

[47] M. C. Zwier, J. L. Adelman, J. W. Kaus, A. J. Pratt, K. F. Wong, N. B. Rego, E. Suárez, S. Lettieri, D. W. Wang, M. Grabe, D. M. Zuckerman, and L. T. Chong. “WESTPA: An Interoperable, Highly Scalable Software Package for Weighted Ensemble Simulation and Analysis”. In: *J. chem. theory comput.* 11.2 (Feb. 2015). Publisher: American Chemical Society, pp. 800–809. ISSN: 1549-9618. DOI: 10.1021/ct5010615. URL: https://doi.org/10.1021/ct5010615 (visited on 06/23/2022).

[48] A. E. García and K. Y. Sanbonmatsu. “A-Helical stabilization by side chain shielding of backbone hydrogen bonds”. In: *Proceedings of the national academy of sciences* 99.5 (Mar. 2002). Publisher: Proceedings of the National Academy of Sciences, pp. 2782–2787. DOI: 10.1073/pnas.042496899. URL: https://www.pnas.org/doi/abs/10.1073/pnas.042496899 (visited on 06/23/2022).

[49] M. Beasley, S. Groover, S. J. Valentine, and J. Legleiter. “Lipid headgroups alter huntingtin aggregation on membranes”. en. In: *Biochimica et biophysica acta (bba) - biomembranes* 1863.1 (Jan. 2021), p. 183497. ISSN: 0005-2736. DOI: 10.1016/j.bbamem.2020.183497. URL: https://www.sciencedirect.com/science/article/pii/S0005273620303400 (visited on 06/23/2022).

[50] S. E. Groover, M. Beasley, V. Ramamurthy, and J. Legleiter. “Phosphomimetic Mutations Impact Huntingtin Aggregation in the Presence of a Variety of Lipid Systems”. In: *Biochemistry* 59.49 (Dec. 2020). Publisher: American Chemical Society, pp. 4681–4693. ISSN: 0006-2960. DOI: 10.1021/acs.biochem.0c00788. URL: https://doi.org/10.1021/acs.biochem.0c00788 (visited on 06/23/2022).

[51] S. H. White and W. C. Wimley. “Hydrophobic interactions of peptides with membrane interfaces”. eng. In: *Biochim biophys acta* 1376.3 (Nov. 1998), pp. 339–352. ISSN: 0006-3002. DOI: 10.1016/s0304-4157(98)00021-5.

[52] N. Kučerka, Y. Liu, N. Chu, H. I. Petrache, S. Tristram-Nagle, and J. F. Nagle. “Structure of Fully Hydrated Fluid Phase DMPC and DLPC Lipid Bilayers Using X-Ray Scattering from Oriented Multilamellar Arrays and from Unilamellar Vesicles”. In: *Biophys j* 88.4 (Apr. 2005), pp. 2626–2637. ISSN: 0006-3495. DOI: 10.1529/biophysj.104.056606. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1305359/ (visited on 06/23/2022).

[53] N. Kucerka, S. Tristram-Nagle, and J. F. Nagle. “Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains”. eng. In: *J membr biol* 208.3 (Dec. 2005), pp. 193–202. ISSN: 0022-2631. DOI: 10.1007/s00232-005-7006-8.

[54] S. Tristram-Nagle, H. I. Petrache, and J. F. Nagle. “Structure and interactions of fully hydrated dioleoylphosphatidylcholine bilayers.” In: *Biophys j* 75.2 (Aug. 1998), pp. 917–925. ISSN: 0006-3495. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1299765/ (visited on 06/23/2022).

[55] S. Côté, G. Wei, and N. Mousseau. “Atomistic mechanisms of huntingtin N-terminal fragment insertion on a phospholipid bilayer revealed by molecular dynamics simulations”. eng. In: *Proteins* 82.7 (July 2014), pp. 1409–1427. ISSN: 1097-0134. DOI: 10.1002/prot.24509.

[56] S. Marqusee and R. L. Baldwin. “Helix stabilization by Glu-...Lys+ salt bridges in short peptides of de novo design”. eng. In: *Proc natl acad sci u s a* 84.24 (Dec. 1987), pp. 8898–8902. ISSN: 0027-8424. DOI: 10.1073/pnas.84.24.8898.

[57] Y. Jiang, S. R. Chadwick, and P. Lajoie. “Endoplasmic reticulum stress: The cause and solution to Huntington’s disease?” eng. In: *Brain res* 1648.Pt B (Oct. 2016), pp. 650–657. ISSN: 1872-6240. DOI: 10.1016/j.brainres.2016.03.034.

[58] K. A. Burke, E. A. Yates, and J. Legleiter. “Biophysical insights into how surfaces, including lipid membranes, modulate protein aggregation related to neurodegeneration”. eng. In: *Front neurol* 4 (2013), p. 17. ISSN: 1664-2295. DOI: 10.3389/fneur.2013.00017.

[59] J. Zhang, J. Tan, R. Pei, and S. Ye. “Acidic Environment Significantly Alters Aggregation Pathway of Human Islet Amyloid Polypeptide at Negative Lipid Membrane”. In: *Langmuir* 36.6 (Feb. 2020). Publisher: American Chemical Society, pp. 1530–1537. ISSN: 0743-7463. DOI: 10.1021/acs.langmuir.9b03623. URL: https://doi.org/10.1021/acs.langmuir.9b03623 (visited on 06/23/2022).

[60] F. Hane, E. Drolle, R. Gaikwad, E. Faught, and Z. Leonenko. “Amyloid-β aggregation on model lipid membranes: an atomic force microscopy study”. eng. In: *J alzheimers dis* 26.3 (2011), pp. 485–494. ISSN: 1875-8908. DOI: 10.3233/JAD-2011-102112.

[61] G. R. Levy, K. Shen, Y. Gavrilov, P. E. S. Smith, Y. Levy, R. Chan, J. Frydman, and L. Frydman. “Huntingtin’s N-Terminus Rearrangements in the Presence of Membranes: A Joint Spectroscopic and Computational Perspective”. eng. In: *Acs chem neurosci* 10.1 (Jan. 2019), pp. 472–481. ISSN: 1948-7193. DOI: 10.1021/acschemneuro.8b00353.

[62] V. N. Sivanandam, M. Jayaraman, C. L. Hoop, R. Kodali, R. Wetzel, and P. C. A. van der Wel. “The aggregation-enhancing huntingtin N-terminus is helical in amyloid fibrils”. eng. In: *J am chem soc* 133.12 (Mar. 2011), pp. 4558–4566. ISSN: 1520-5126. DOI: 10.1021/ja110715f.

[63] L. Pieri, K. Madiona, L. Bousset, and R. Melki. “Fibrillar α-Synuclein and Huntingtin Exon 1 Assemblies Are Toxic to the Cells”. In: *Biophys j* 102.12 (June 2012), pp. 2894–2905. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2012.04.050. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3379023/ (visited on 06/23/2022).

[64] Y. S. Choo, G. V. W. Johnson, M. MacDonald, P. J. Detloff, and M. Lesort. “Mutant huntingtin directly increases susceptibility of mitochondria to the calcium-induced permeability transition and cytochrome c release”. eng. In: *Hum mol genet* 13.14 (July 2004), pp. 1407–1420. ISSN: 0964-6906. DOI: 10.1093/hmg/ddh162.

[65] G. Drin and B. Antonny. “Amphipathic helices and membrane curvature”. eng. In: *Febs lett* 584.9 (May 2010), pp. 1840–1847. ISSN: 1873-3468. DOI: 10.1016/j.febslet.2009.10.022.

[66] N. S. Hatzakis, V. K. Bhatia, J. Larsen, K. L. Madsen, P.-Y. Bolinger, A. H. Kunding, J. Castillo, U. Gether, P. Hedegård, and D. Stamou. “How curved membranes recruit amphipathic helices and protein anchoring motifs”. en. In: *Nat chem biol* 5.11 (Nov. 2009). Number: 11 Publisher: Nature Publishing Group, pp. 835–841. ISSN: 1552-

4469. DOI: 10.1038/nchembio.213. URL: https://www.nature.com/articles/nchembio.213 (visited on 06/23/2022).

[67] M. Tao, N. K. Pandey, R. Barnes, S. Han, and R. Langen. “Structure of Membrane-Bound Huntingtin Exon 1 Reveals Membrane Interaction and Aggregation Mechanisms”. en. In: *Structure* 27.10 (Oct. 2019), 1570–1580.e4. ISSN: 0969-2126. DOI: 10.1016/j.str.2019.08.003. URL: https://www.sciencedirect.com/science/article/pii/S0969212619302746 (visited on 06/23/2022).

[68] S. Vanni, L. Vamparys, R. Gautier, G. Drin, C. Etchebest, P. F. Fuchs, and B. Antonny. “Amphipathic Lipid Packing Sensor Motifs: Probing Bilayer Defects with Hydrophobic Residues”. In: *Biophys j* 104.3 (Feb. 2013), pp. 575–584. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2012.11.3837. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3566459/ (visited on 06/23/2022).

[69] C. Prévost, M. E. Sharp, N. Kory, Q. Lin, G. A. Voth, R. V. Farese, and T. C. Walther. “Mechanism and Determinants of Amphipathic Helix-Containing Protein Targeting to Lipid Droplets”. eng. In: *Dev cell* 44.1 (Jan. 2018), 73–86.e4. ISSN: 1878-1551. DOI: 10.1016/j.devcel.2017.12.011.

[70] G. van Meer, D. R. Voelker, and G. W. Feigenson. “Membrane lipids: where they are and how they behave”. en. In: *Nat rev mol cell biol* 9.2 (Feb. 2008). Number: 2 Publisher: Nature Publishing Group, pp. 112–124. ISSN: 1471-0080. DOI: 10.1038/nrm2330. URL: https://www.nature.com/articles/nrm2330 (visited on 06/23/2022).

[71] S. Vanni, H. Hirose, H. Barelli, B. Antonny, and R. Gautier. “A sub-nanometre view of how membrane curvature and composition modulate lipid packing and protein recruitment”. en. In: *Nat commun* 5.1 (Sept. 2014). Number: 1 Publisher: Nature Publishing Group, p. 4916. ISSN: 2041-1723. DOI: 10.1038/ncomms5916. URL: https://www.nature.com/articles/ncomms5916 (visited on 06/23/2022).

[72] K. A. Burke, E. A. Yates, and J. Legleiter. “Amyloid-forming proteins alter the local mechanical properties of lipid membranes”. eng. In: *Biochemistry* 52.5 (Feb. 2013), pp. 808–817. ISSN: 1520-4995. DOI: 10.1021/bi301070v.

5. DEVELOPMENT OF A HIGH-THROUGHPUT MEMBRANE PERMEABILITY ASSAY FOR CYCLIC LARIAT PEPTIDES

5.1 INTRODUCTION

Membrane-permeable cyclic peptides are attractive to many scientists and pharmaceutical companies as potential drug candidates[1]. The selection of possible peptide sizes allows selectivity for the target receptor and modulates the protein-protein interaction. While linear peptides are flexible, and therefore lack affinity, selectivity, and bioavailability, cyclic peptides are more rigid and adopt structures with affinity for protein surfaces with high specificity. Paramount to the identification of drug-viable cyclic peptides is the determining the potential aqueous structures of the peptide and the membrane permeability of the peptide.

A particular class of cyclic peptides known as lariat depsipeptides make up roughly 30% of the naturally occurring cyclic peptides according to a survey[2] of the Natural Products Atlas[3]. The majority of these peptides are cyclized by a connection between the C-terminal carboxylic acid and a side-chain hydroxyl group to form an ester linkage. Of particular interest to this study is the library of peptides studied by C. Kelly, et al.[2] (**Fig. 5.1**), 4096 nine-residue peptides with an ester linkage between the C-terminus and Thr3. While C. Kelly, et al. tested the permeability of a sizeable proportion of the peptides in the library, a high-throughput method of determining the permeability as well as revealing pertinent structural information of potential candidates would be a useful tool.

MD simulations are an excellent tool to determine the aqueous solution structures of cyclic peptides; these are of particular interest as they can serve as starting structures for docking simulations to determine if the peptide could act as a drug for a particular protein-protein interaction. Since cyclic peptides often occupy several different conformations in

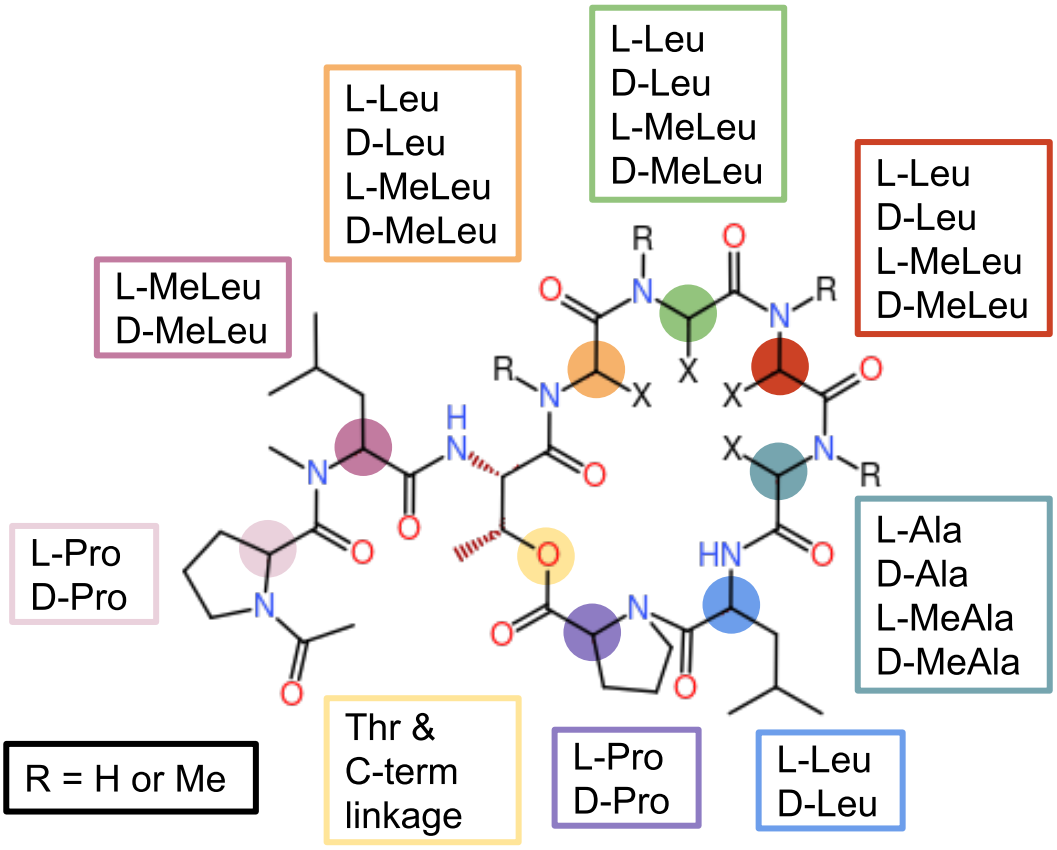

Figure 5.1: Cyclic peptide scaffold used in this study. Lariat peptides have a short tail in addition to a cyclic backbone. In this particular study, peptides had seven residues within the macrocycle connected by a depsi linkage between the sidechain of T3 and the C-terminus of P9. The peptide library consists of permutations at each residue: "L" and "D" designate the specific amino acid enantiomers and "Me" designates methylation of the backbone nitrogen.

solution, it is difficult to resolve their structure via a method like NMR spectroscopy[4]. Umbrella sampling (US) is an established method for determining the permeability of small molecules[5, 6, 7] and cyclic peptides in particular[8, 9], however, even this method can be cumbersome requiring upwards of 20 simulations on the order of 200 ns per peptide with the added system size of a membrane.

The overall question that we wanted to answer was, "Is it possible to use MD simulations of cyclic peptides in different solvents to predict membrane permeability?" In order to do this, it required screening of a large number of peptides and running simulations in solvents with large enough physical differences to provide resolving power in the conformational behavior of these peptides in solution. For this study, we chose water (i.e., aqueous) and

octanol (organic), as they are the de facto solvents[10] used to calculate partition coefficients of small molecules.

5.2 METHODS

5.2.1 Force field parameterization

Parameters for the depsilinkage (ester linkage) between Thr3 and Pro9 were generated via FFTK[11]. Parameters for methylated Leucine, methylated D-Leucine, and methylated D-Alanine were determined by comparison to existing residues in the CHARMM36 force field[12].

5.2.2 Structure generation

15 peptides were chosen randomly from each library of those tested by Kelly et al.[2] resulting in 240 total peptides. Linear structures were generated and the depsilinkage formed with psfgen in VMD[13]. Peptide structures were minimized in NAMD 2.14[14] for 5000 steps with a 10 kcal/mol/Å^2 restraint on the omega dihedrals of the residues in the ring using colvars[15]. Each peptide was then solvated in water and in octanol using *solvate* in VMD[13].

5.2.3 Molecular dynamics

Each of the 480 systems were minimized for 7500 steps then equilibrated in NPT for 100 ps, and finally run in NVT for 10 ns of conventional MD followed by 40 ns of gaussian-accelerated MD[16, 17, 18, 19]. Each step was carried out in NAMD 2.14[14]. Four peptides were run under the same conditions in both conventional and gaussian-accelerated MD for 250 ns to confirm convergence of the 50 ns runs.

5.2.4 Permeability Assay

To predict the permeability of the peptides, we used a modified version of the following resistivity (R) and permeability (P) equation

$$R = \frac{1}{P} = \int_{z_1}^{z_2} \frac{exp[\beta W(z)]}{D(z)} dz, \tag{5.1}$$

where $W(z)$ is the potential of mean force (PMF), $D(z)$ is the local diffusivity coefficient, β is the thermodynamic beta ($\beta = 1/k_B T$), and z is a collective variable the describes the relative position as the solute transitions through solvents[6]. We will instead use the following

$$R = \frac{1}{P} = \frac{exp[\beta W(z_{aq})]}{D(z_{aq})} - \frac{exp[\beta W(z_{oct})]}{D(z_{oct})}, \tag{5.2}$$

to estimate the permeability of a peptide given the PMF and local diffusivity in water (z_{aq}) and octanol (z_{oct}). The PMF ($W(z)$) was estimated with the Boltzmann distribution,

$$p_i \propto e^{-\varepsilon_i/(kT)} \tag{5.3}$$

where p_i is the probability of the system being in state i, ε_i is the energy of that state, k is the Boltzmann constant, and T is the temperature. We applied this to the five most representative structures of each of the 250 ns GaMD runs found via k-means clustering; this generates a dictionary of relating structure and energy. To determine the PMF for each peptide, we used k-means clustering to find the 5 most representative structures then used the *rmsd2ref* tool in LOOS[20, 21] to compare the representative structures back to the dictionary assigning the energy value to the representative structure from the most similar dictionary structure. The five energy values were then averaged. The local diffusivity constant will be estimated using the generalized Einstein-Smoluchowski equation [6],

$$D(z) = \frac{\langle |z(t) - z(0)|^2 \rangle}{6t} \tag{5.4}$$

where $z(t)$ represents the position at time, t. To account for the affect of GaMD on diffusion, we compared diffusivities of the 250 ns runs to calculate a scaling factor of 2.06.

5.2.5 Analysis

Backbone dihedral angles. The backbone dihedral angles were measured using the torsion tool in LOOS[20, 21]. The linkage interrupts the measurement of the final ω dihedral. *Hbonds.* The number of hydrogen bonds was measured using the *hmatrix* tool in LOOS[20, 21]. *Principle Component Analysis.* The principle component analysis (PCA) was performed with the *svd* tool on the backbone heavy atoms in LOOS[20, 21].

5.3 RESULTS AND DISCUSSION

The first step was to directly compare the results of our permeability assay (P_{calc}) to the permeability results (P_{app}) from Kelly et al.[2]. Although the range of $\log(P_{calc})$ is similar to that of $\log(P_{app})$, there is no correlation between the two sets of values (**Fig 5.2**). This was confirmed by a Pearson correlation coefficient of -0.0059; correlated or anti-correlated data sets have values that exceed +/- 0.5. This indicates that calculation of partitioning coefficients based on the PMF between two solvents is insufficient to correctly predict behavior of permeability of cyclic peptides. The next step was to look for other observables from our simulations that could indicate differences in behavior of cyclic peptides in water versus octanol.

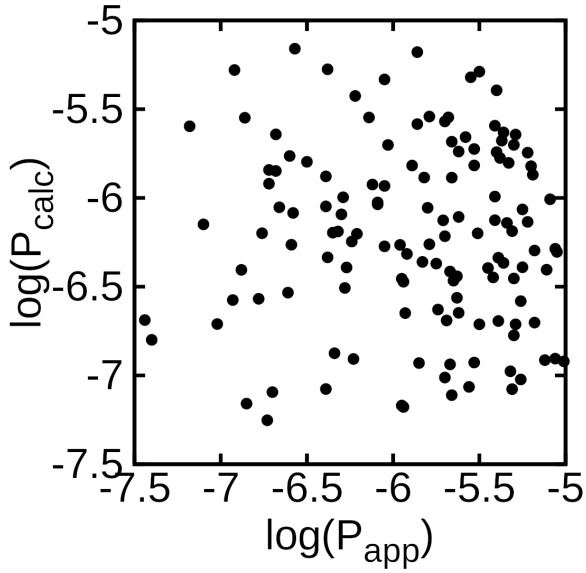

Figure 5.2: Calculated log(P) values from MD simulations are in poor agreement with experimental data. Plot of the log of apparent partitioning coefficients (P_{app}) obtained from Kelly et al.[2] against partitioning coefficients calculated from simulations conducted in this study ($\log(P_{calc})$). Approximately 200 peptides were tested.

Next, we looked to see if the trend in methylation was conserved: increase in methylation correlated to an increase of permeability. We observe fairly good agreement with the experimental results of Kelly et al.[2], in which there is a noticeable increase in permeability from N = 0 to N =1 and slight increase from N = 1 to N = 4 (**Fig. 5.3**). The conventional

hypothesis with respect to N-methylation in cyclic peptides is that for each N atom that is methylated, it removes a potential hydrogen donor from the molecule. This effectively makes the cyclic peptide less soluble in aqueous solution and more prone diffuse across the membrane.

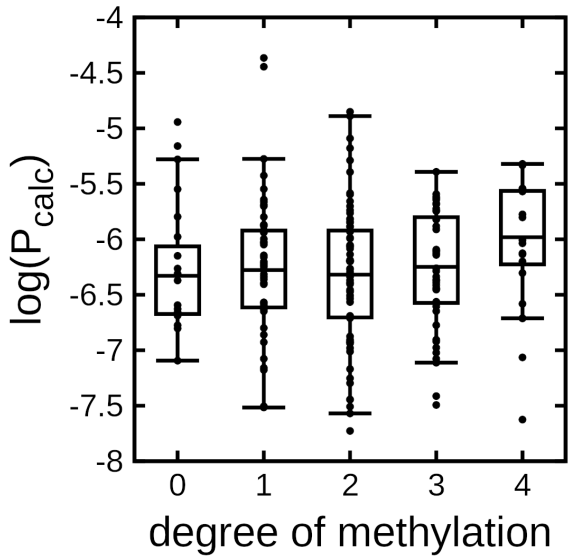

Figure 5.3: N-methylation generally correlates with permeability. Box plot of partition coefficients based on MD simulations (P_{calc}) as a function of the number of N-methylations on the cyclic peptide backbone. Mid-points in each box is the mean for each particular data set. Boxes represent the two intermediate quartiles of each data set. Boxes and means were calculated by leaving out the two largest outliers.

In order to more appropriately represent the distribution of log(P) values for each methylation state, we dropped the two farthest outliers from each set of data points. Careful analysis of the lariat peptide library from Kelly et al. shows two things: 1) N-methylation is a gaussian distribution (most highly populated value is N = 2), and 2) actual experimental data on the peptide library is 23-40% of the total theoretical number in the peptide library[2] (**Fig. 5.4**). N = 0 and N = 4 were by far the most poorly represented peptides in the study of Kelly et al., and so it was necessary for them to over-sample these two populations in order to capture accurate permeability behavior. In our case, we did not have the computational resources to run an equivalent number of cyclic peptides (this would have translated to an additional 30-40 peptide systems), and so the permeability behavior for N = 0,4 will be noisier than the other three data sets. By removing the two largest outliers, we more

effectively capture the trends across the entire library tested that is in greater agreement with the experimental data as well.

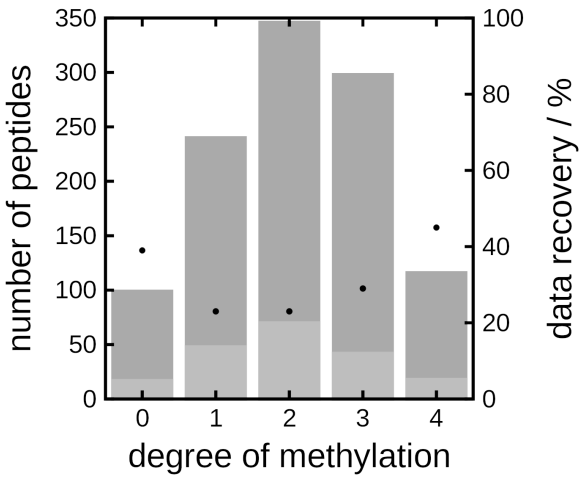

Figure 5.4: Distribution of peptide library by N-methylation is roughly gaussian and indicates how oversampling of N = 0,4 groups is beneficial to understanding overall trends. Bar plot of the number of cyclic peptides that were tested according to N-methylation for the experimental approach of Kelly et al.[2] (dark gray) and the simulation-based approach (light gray). Data recovery percentage is based on the expected number of cyclic peptides that could theoretically be produced during library generation versus the actual number of cyclic peptides that were synthesized[2].

Comparison to Kelly et al.[2] shows that permeability is loosely correlated with heterochirality. However, this may have been an optimistic interpretation of their results in light of our data (**Fig. 5.5**). For the more largely populated portions of heterochirality (N = 2,3,4,5, (**Fig. 5.6**)), the distributions were quite broad ($\log(P_{app})$ -5.0 to -7.5); our data is roughly consistent with these results. (Comparison between experimental and computational results for N = 0, 1, 7, and 8 would largely be speculative because of the smaller number of samples for each of these states). Given these conditions, the most likely conclusion to draw from our results is that there is little to no correlation between heterochirality and permeability. This underscores the need to employ additional approaches to effectively identify promising candidates.

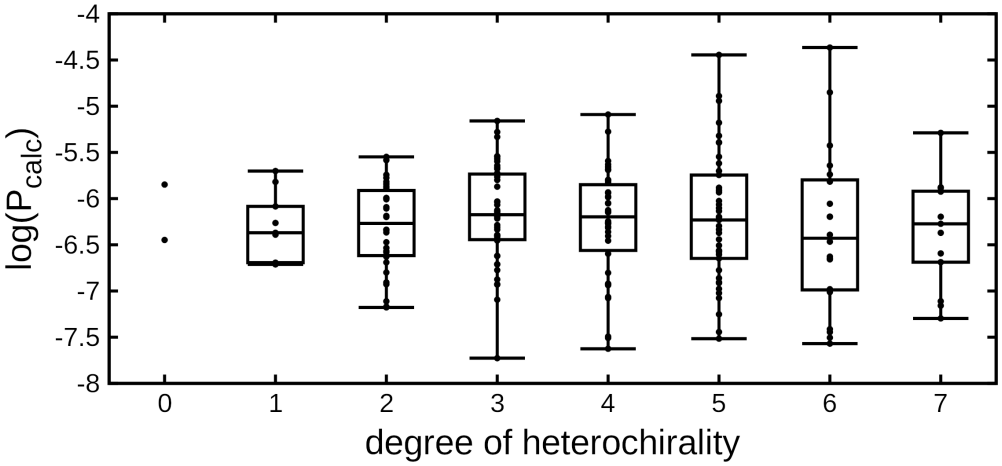

Figure 5.5: Heterochirality in general has no effect on calculated permeability of cyclic peptides. Box plot of partition coefficients based on MD simulations (P_{calc}) as a measure of the degree of heterochirality (i.e., number of D-amino acid residues). Mid-points in each box is the mean for each particular data set. Boxes represent the two intermediate quartiles of each data set.

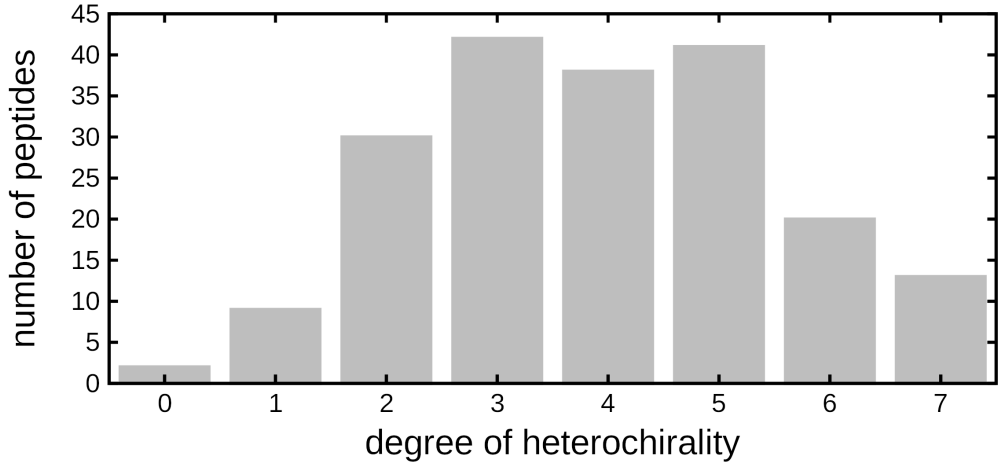

Figure 5.6: Cyclic peptides tested by heterochirality is skewed towards peptides with several D-amino acids.

After comparing our results to the experimental results of Kelly, et al.[2], we looked to determine the ability of the assay to provide relevant results about the structure of the peptides. First, we calculated the propensity of each residue pair to form intramolecular hydrogen bonds (**Fig. 5.8**). In general, the propensity for hydrogen bond formation decreases with an increase in degree of N-methylation. In particular, Leu8 consistently participates as a hydrogen bond acceptor regardless of solvent, indicating its importance as a key residue for stabilization of the cyclic peptides. Quite often, Leu4 and Leu5 are the hydrogen bond donors to Leu8 which can stabilize particular conformations of the cyclic peptide backbone. A noticeable difference between solvents occurs in octanol, where Leu4 and Leu5 participate as hydrogen bond acceptors as well, forming hydrogen bonds with Ala7 and Leu8. This additional non-bonded interaction may be necessary to facilitate the conformational transition that allows for effective diffusion of these specific cyclic peptides from a polar to a nonpolar environment.

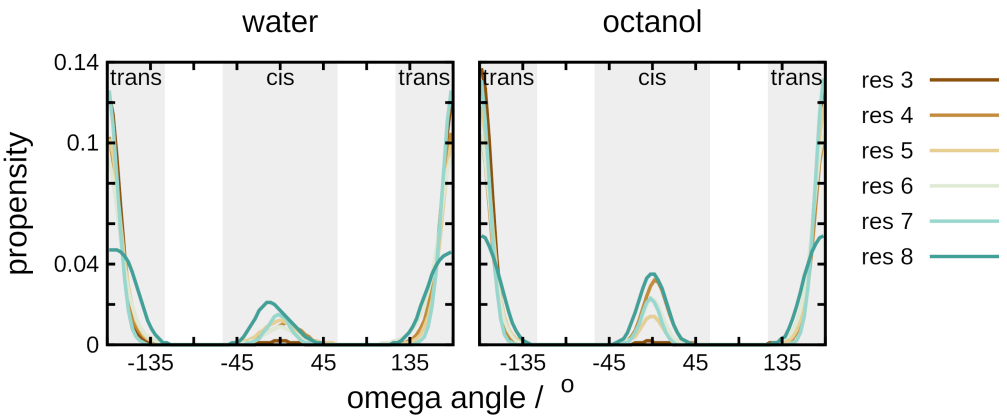

Figure 5.7: Several residues in the lariat ring frequently have an cis ω dihedral angle. The propensity for each residue to occupy the phase space of the ω dihedral angle based on solvent. The residue number corresponds to the ω dihedral that follows that residue (i.e., res 3 corresponds to the ω dihedral between residues 3 and 4).

The ω bond in peptides is the bond that joins the carbonyl carbon of the previous residue to the nitrogen of the next residue; very rarely is the bond not trans (around $\pm 180 \mathrm{deg}$). However, in short cyclic peptides, cis omega dihedral angles are more common[22], especially those involving proline residues[23]. We observed a similar trend here (**Fig. 5.7**) with the Leu8 to Pro9 having a cis configuration for 24% and 31% of the time in water and octanol respectively. We also observed an increased percentage of cis ω dihedrals for nearly every

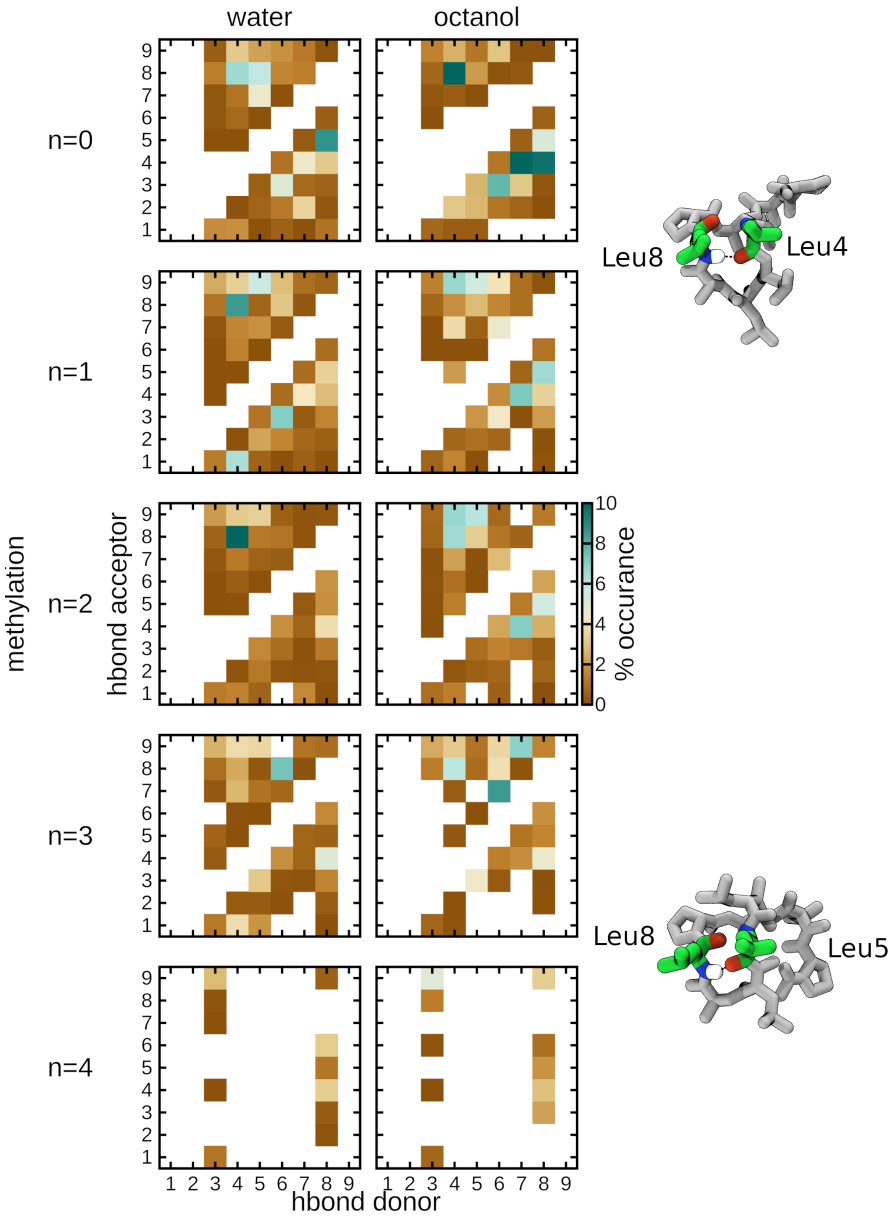

Figure 5.8: Octanol solvent is a major driver of formation of intramolecular hydrogen bonds. Fraction of intramolecular hydrogen bond formation between the donor (NH) and acceptor (C=O) groups in cyclic peptides as a function of N-methylation (n = 0, 1, 2, 3, 4) and solvent (water or octanol).

residue in octanol with respect to water; this is most likely due to the more hydrophobic nature of octanol causing the trans to cis flip to better bury the hydrophilic carbonyl on the peptide backbone.

Finally, we performed principle component analysis (PCA) on the heavy atoms of the peptide backbones in attempts to simplify the predominant motions of the peptide. Plotting PC1 vs PC2 (**Figs. 5.11-5.26**), 16 of the roughly 200 total peptides showed a distinct difference in the distributions between water and octanol (**Fig. 5.9**. We used the 16 peptides (6 of which had values that corresponded with the data from Kelly et al.[2]) to recompare our $\log(P_{calc})$ and their $\log(P_{app})$ values. Filtering of data using PCA clearly shows an improvement in agreement between experiment and simulation and provides potential validation for using PCA as a metric for identification of cyclic peptides with distinct conformational and physical behavior.

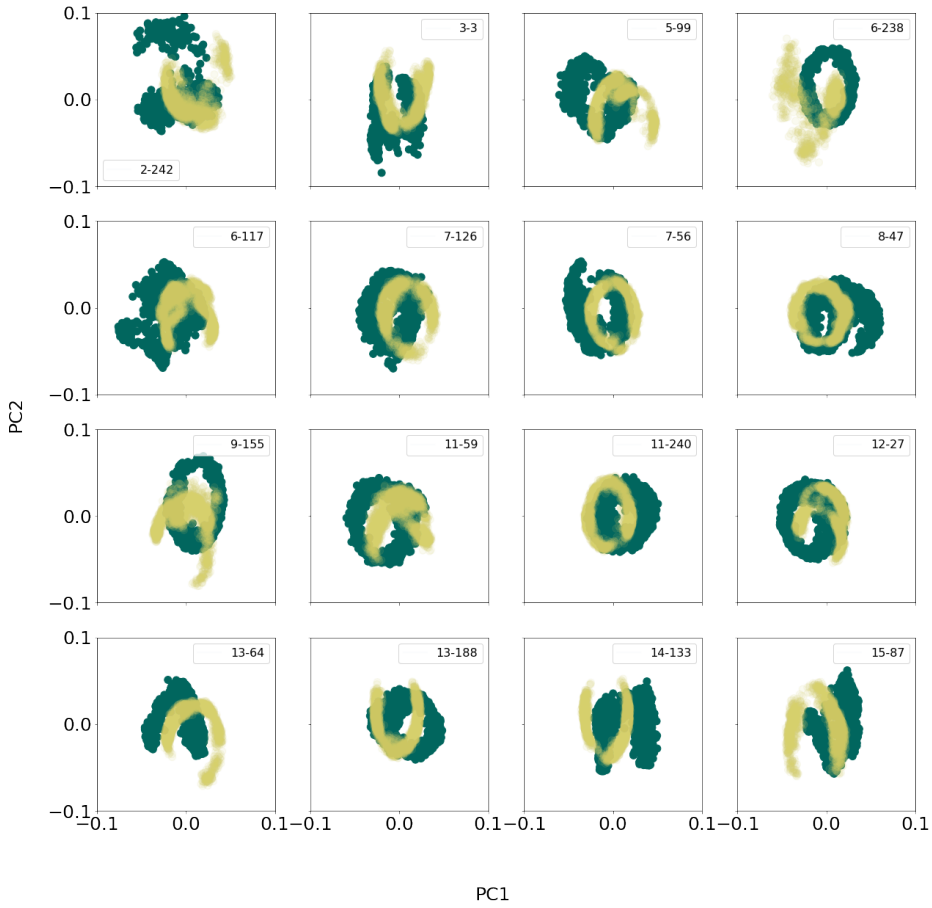

Figure 5.9: PCA can be used to identify peptides with distinct conformational behavior between aqueous and organic solvent. Principal component analysis (PCA) was applied to the heavy atoms in cyclic peptide backbones, plotting the first (PC1) versus second (PC2) principal component in the PCA series. *Teal:* water; *Cream:* octanol. Visual inspection of the distribution of PC1 versus PC2 was carried out for all cyclic peptides in this study; the 16 peptides with the most noticeable difference between water and octanol are presented here. The full library of PCA plots (**Figs. 5.11-5.26**) are divided by into the sublibraries from Kelly et al.[2].

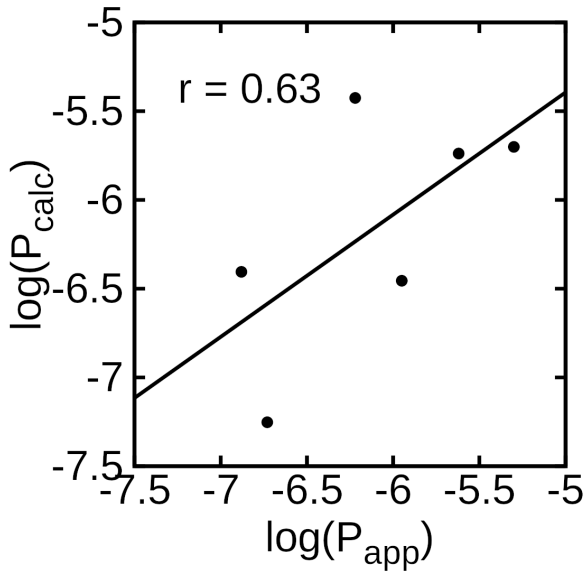

Figure 5.10: Filtering of data via PCA leads to reliable correlation between experimental and predicted peptide permeability. Plot of the log of apparent partitioning coefficients (Papp) obtained from Kelly et al. [2] against partitioning coefficients calculated from simulations conducted in this study ($\log(P_{calc})$) using only cyclic peptides that displayed distinct behavior from PCA (**Fig. 5.9**). Of the 16 peptides that were selected from PCA, only six had corresponding experimental data from [2].

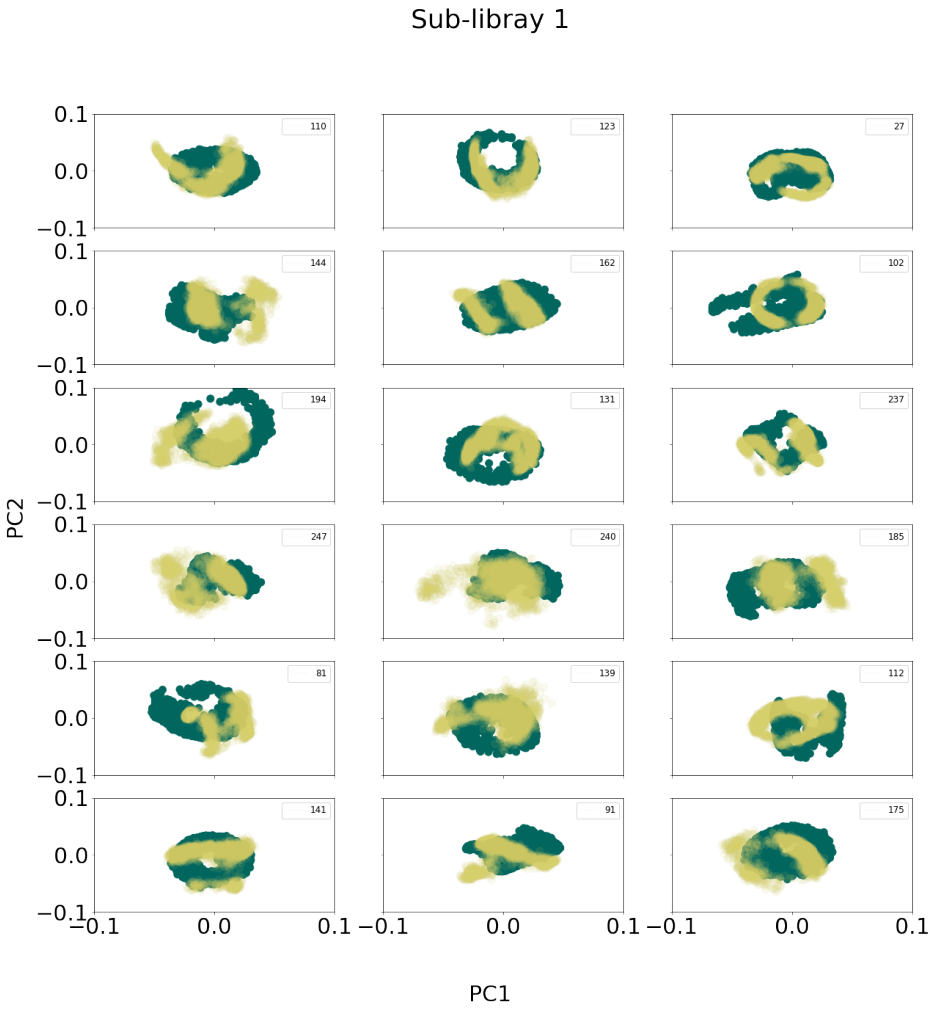

Figure 5.11: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

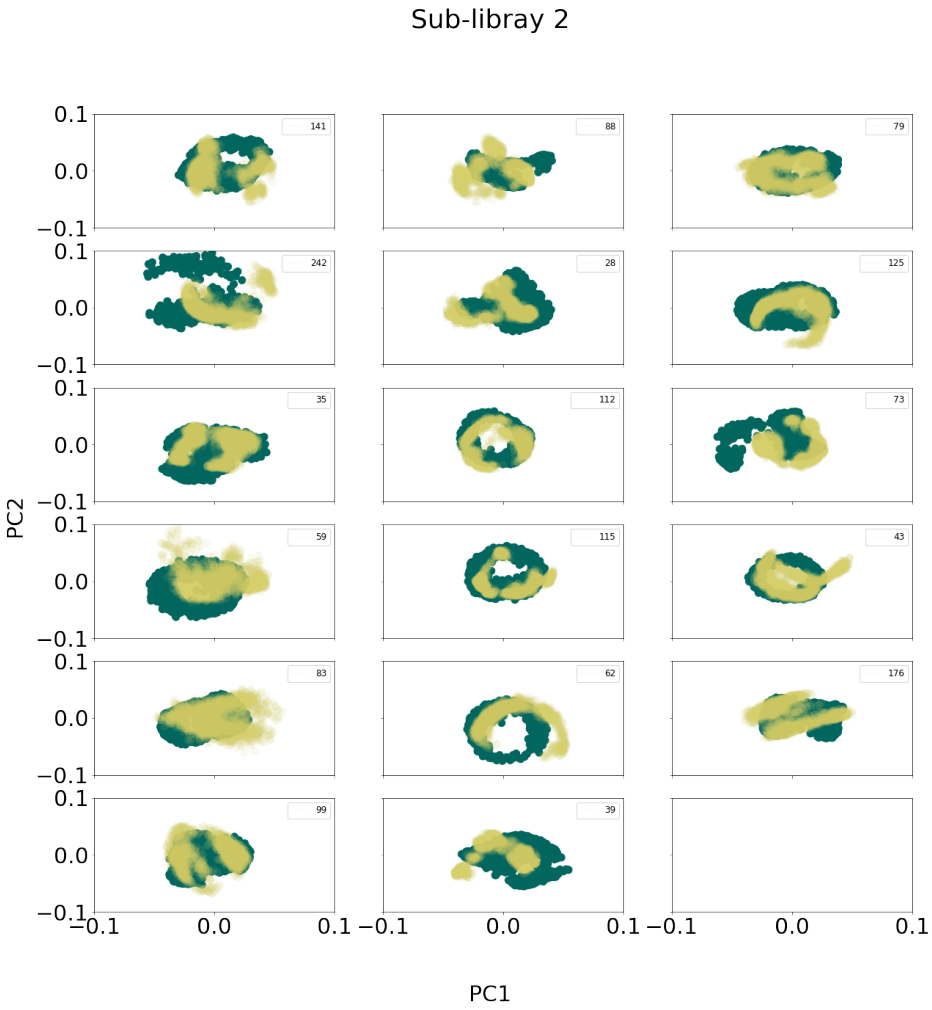

Figure 5.12: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

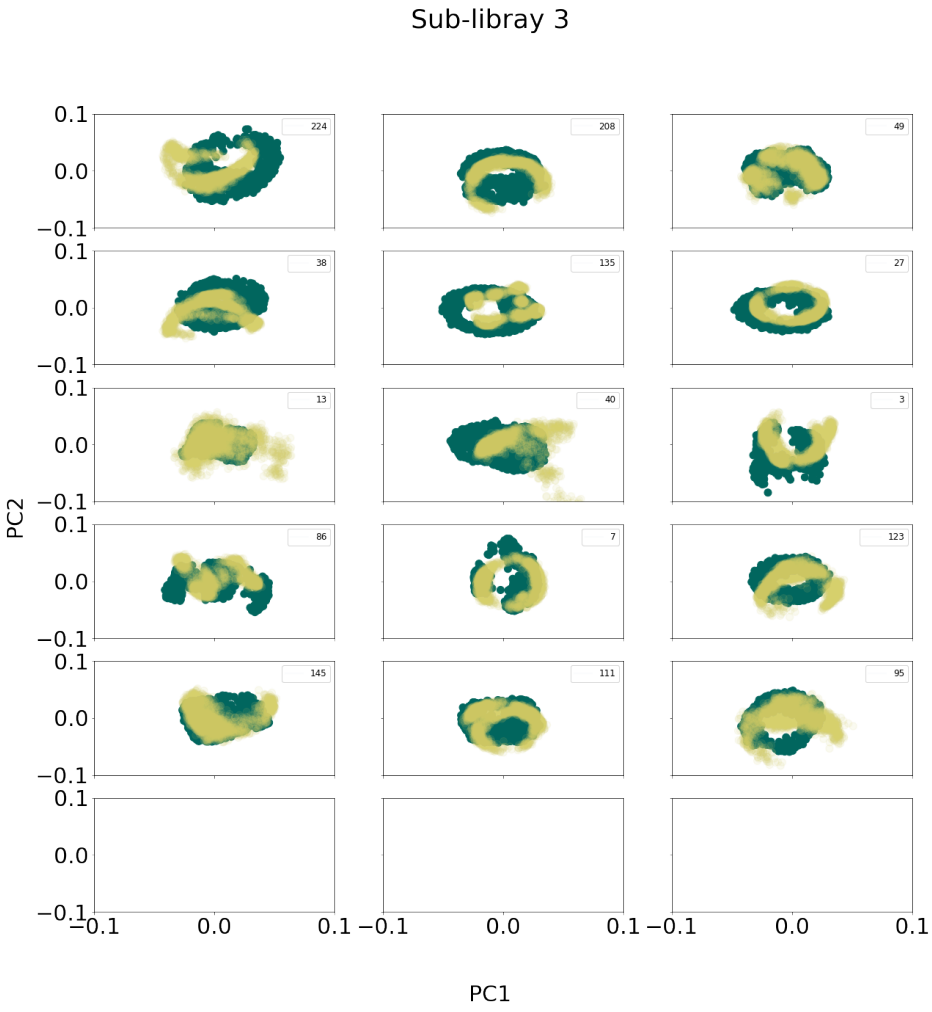

Figure 5.13: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

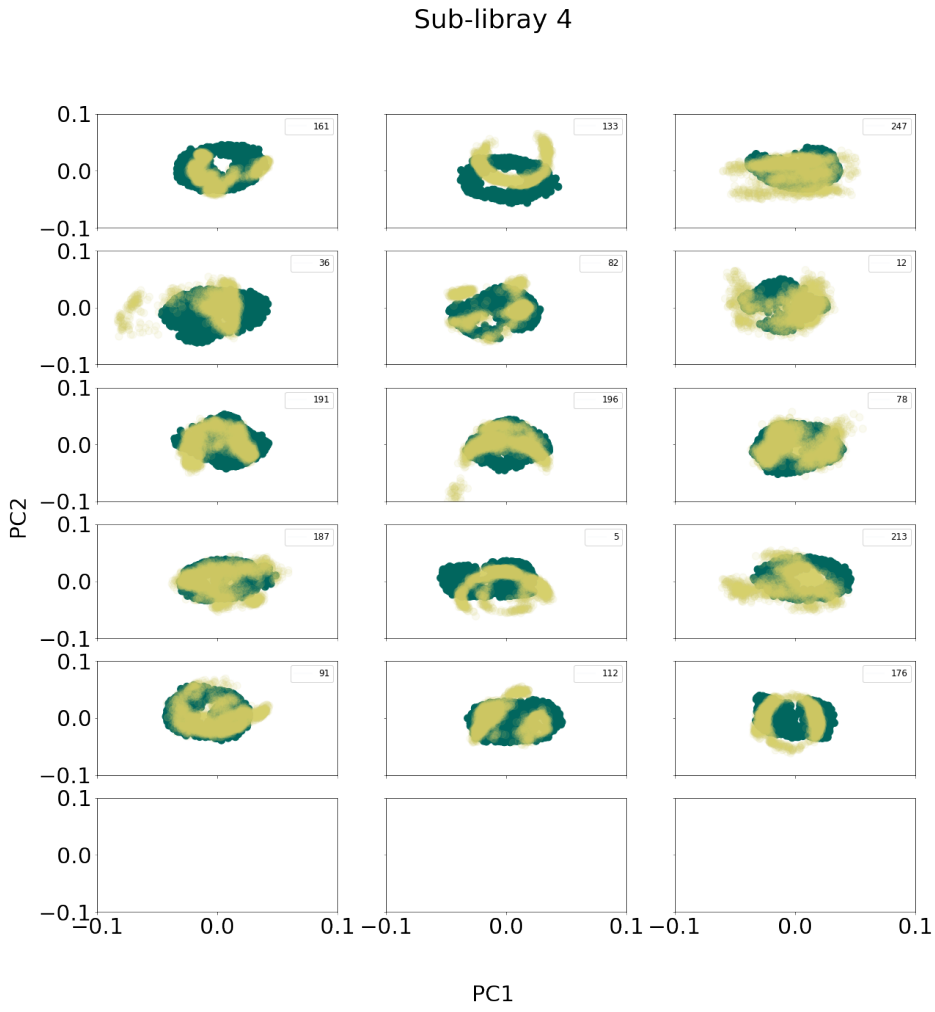

Figure 5.14: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

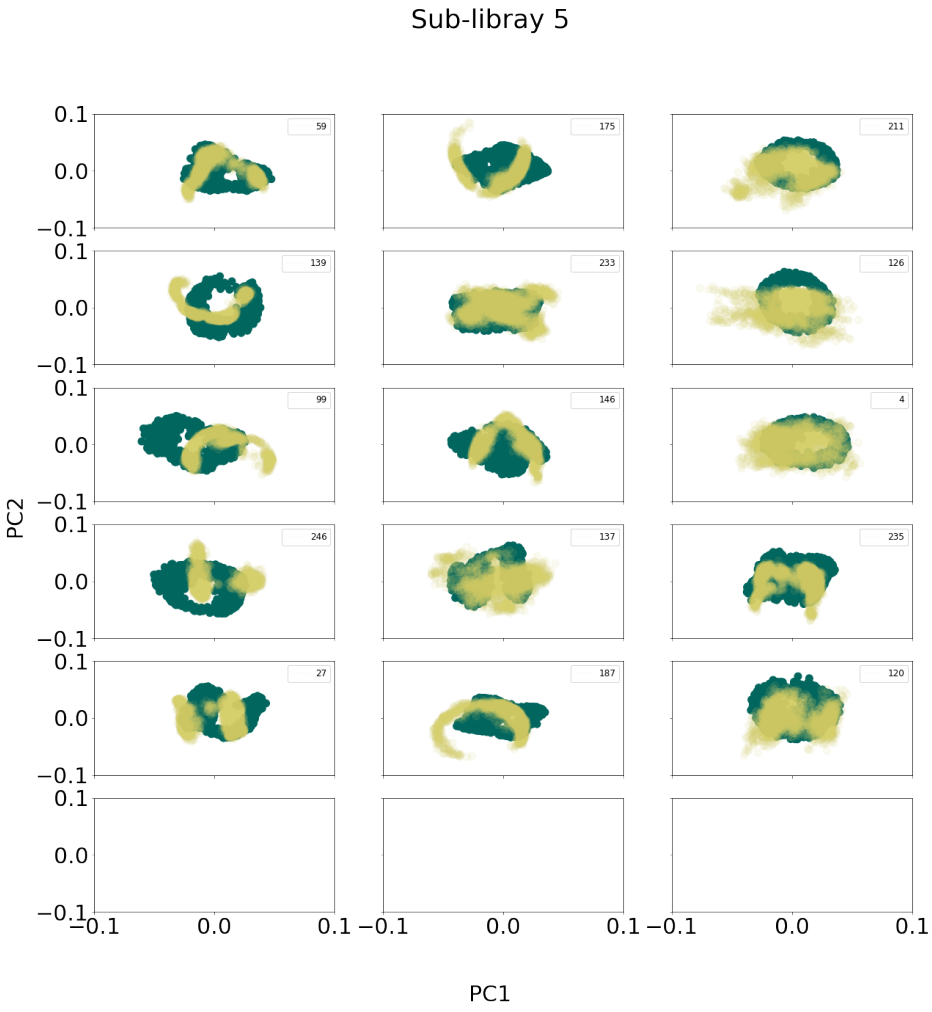

Figure 5.15: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

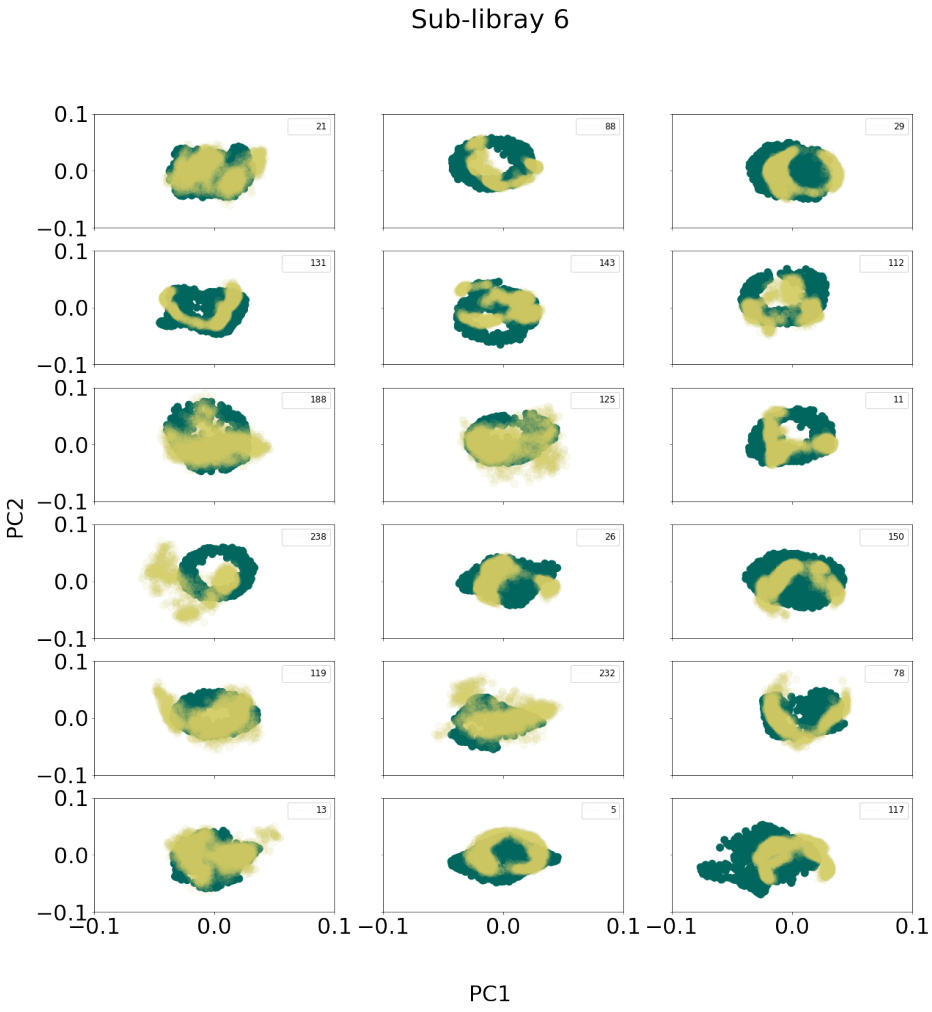

Figure 5.16: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

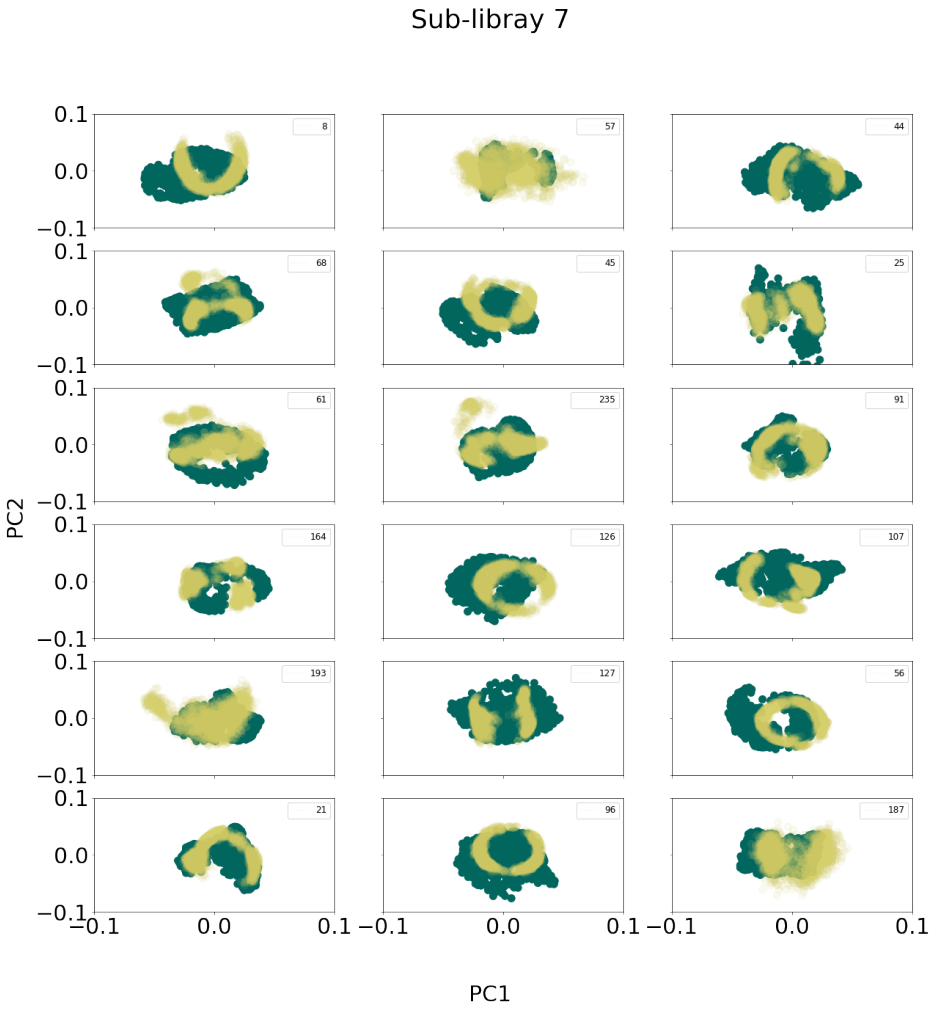

Figure 5.17: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

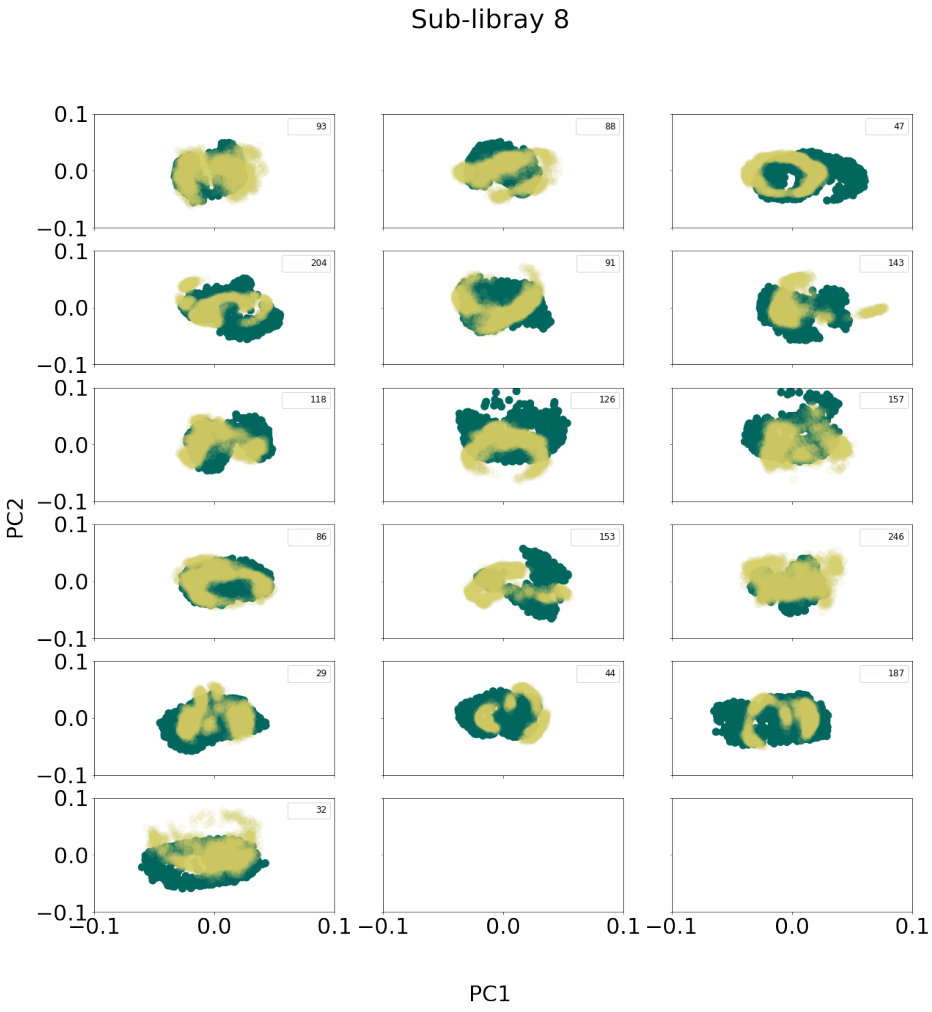

Figure 5.18: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

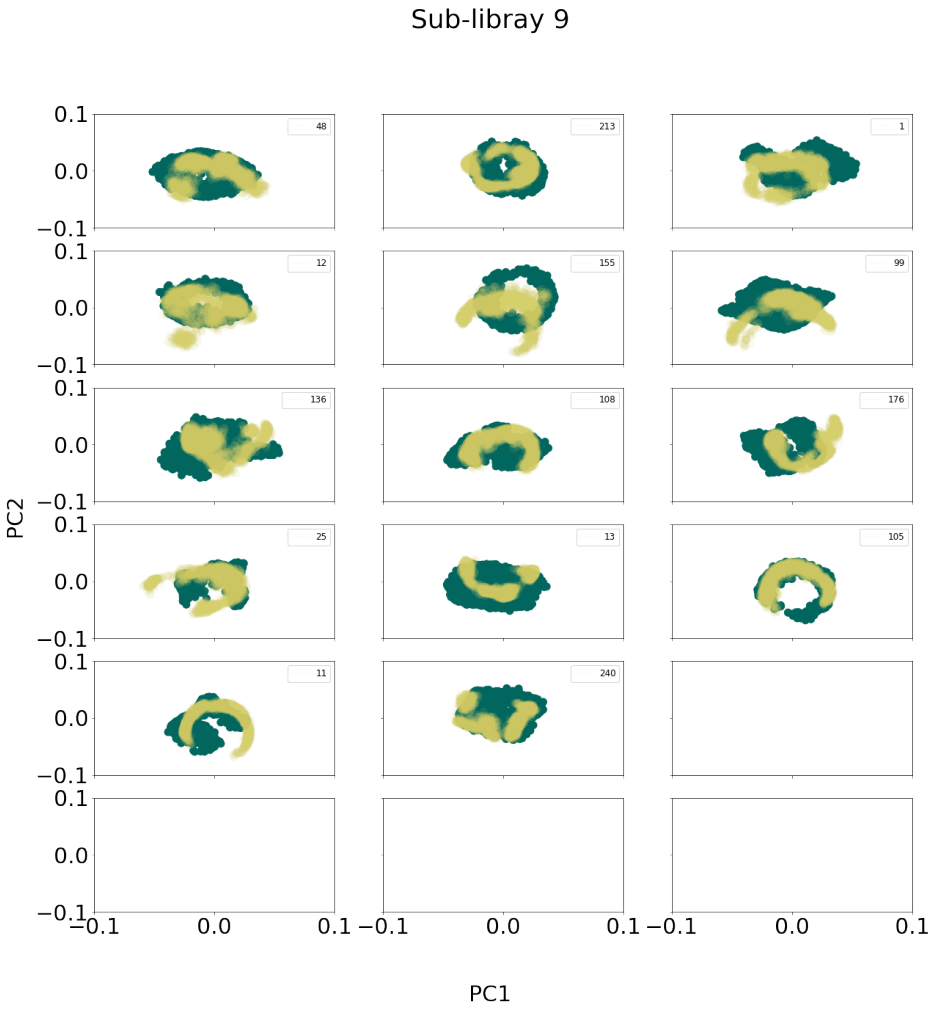

Figure 5.19: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

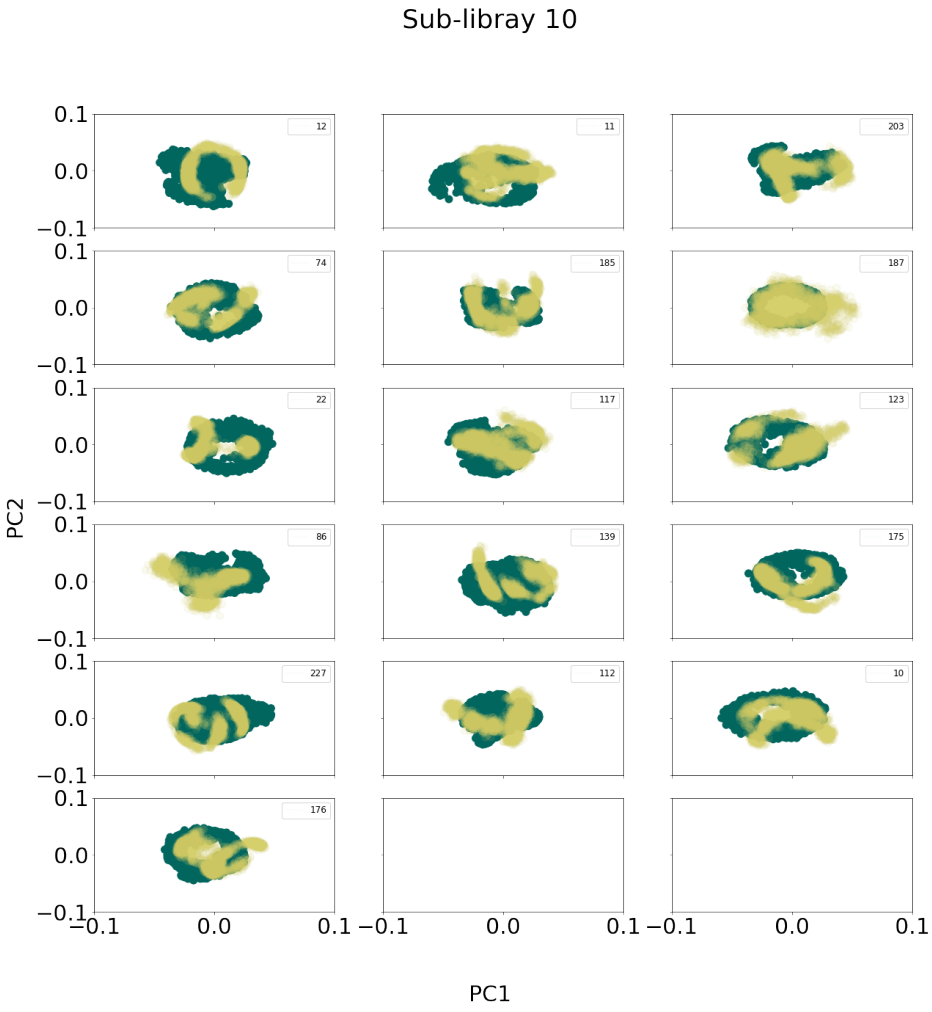

Figure 5.20: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

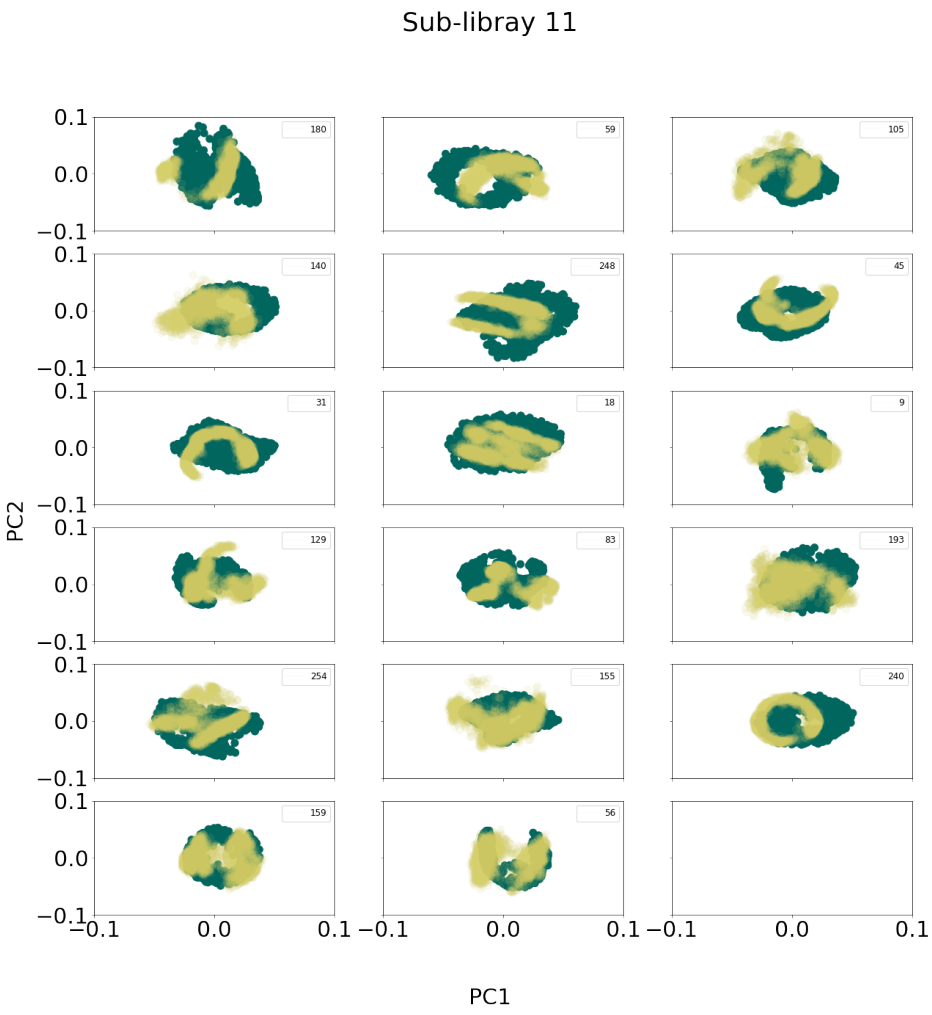

Figure 5.21: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

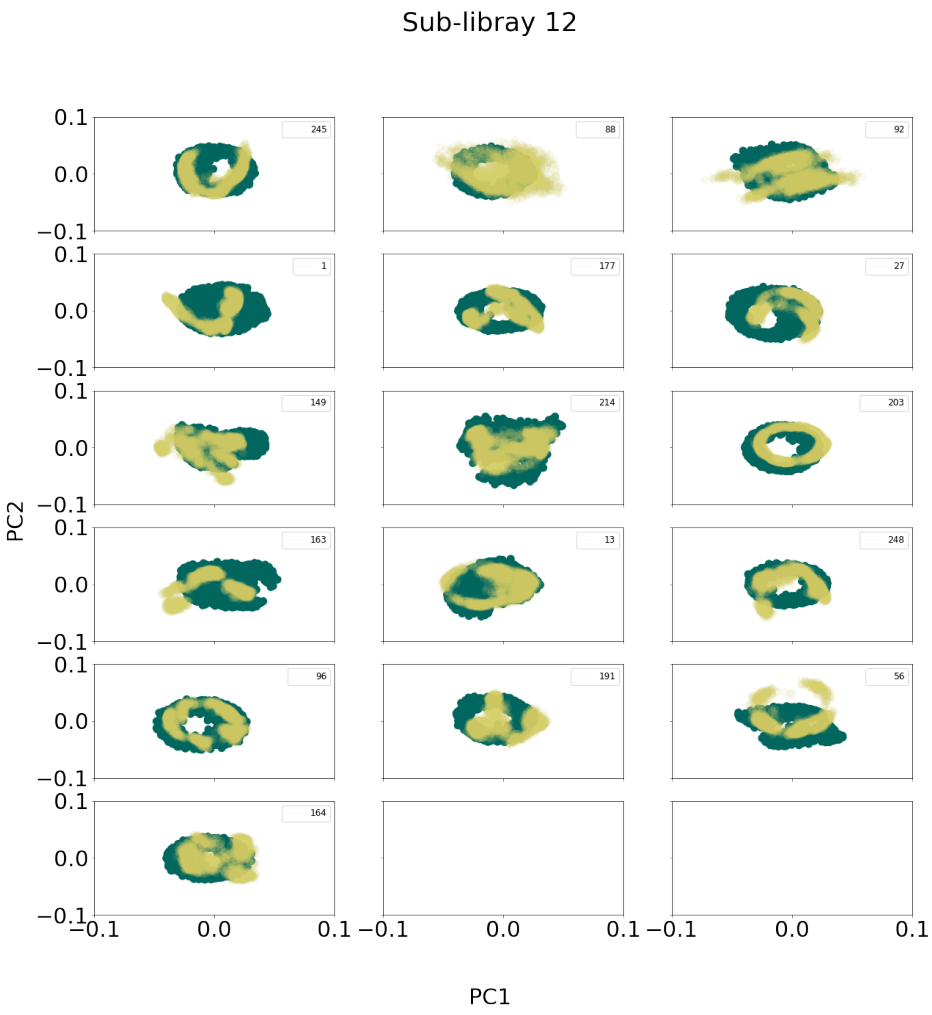

Figure 5.22: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

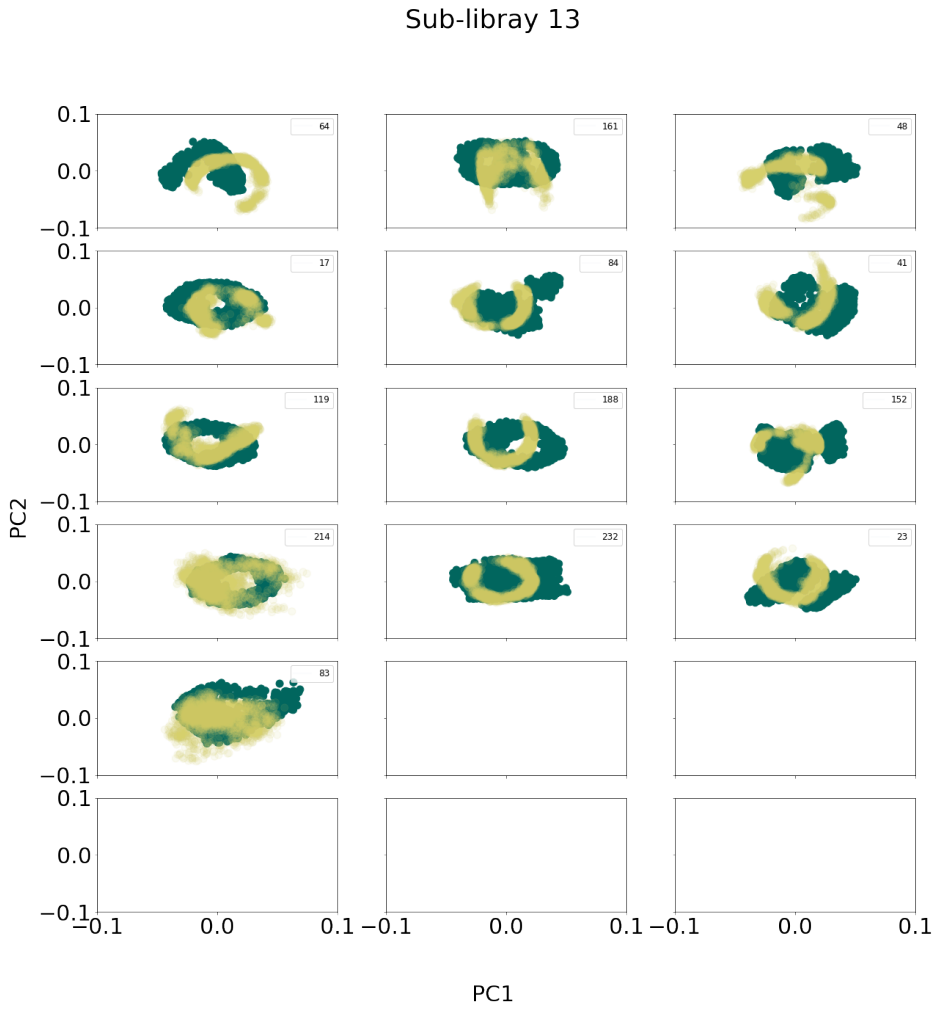

Figure 5.23: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

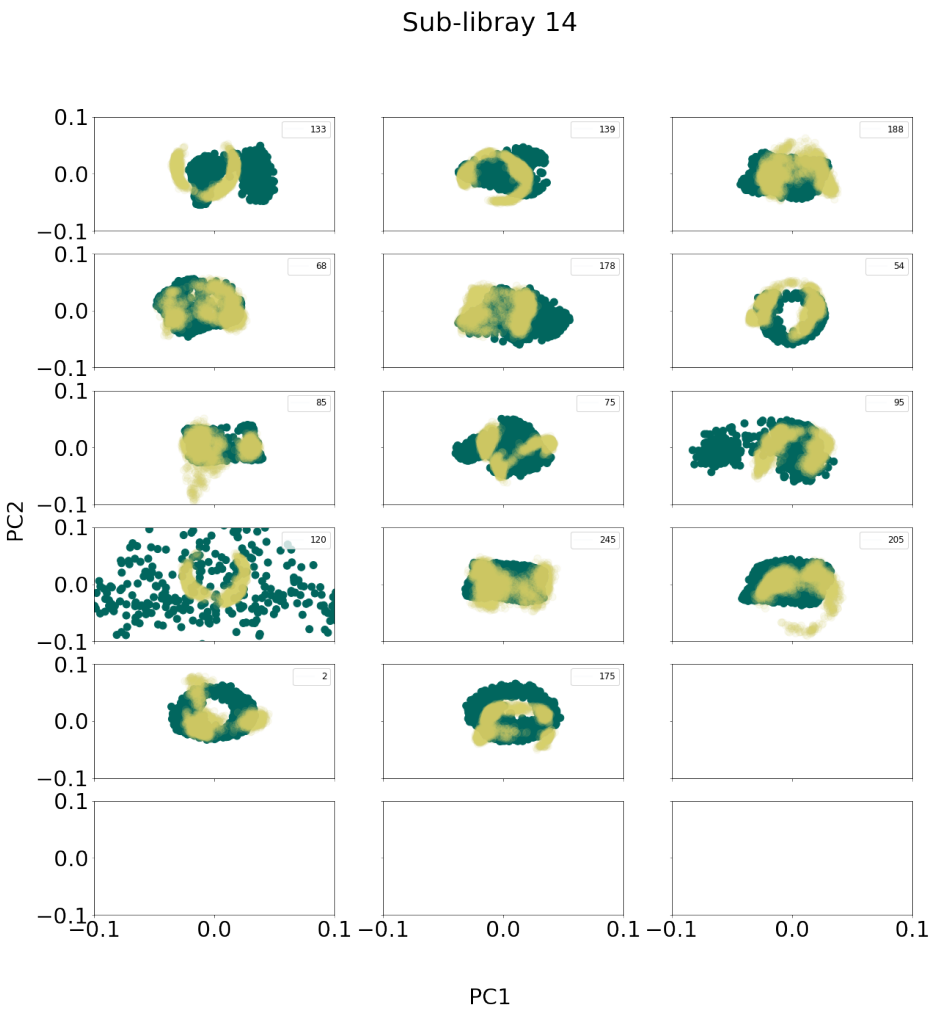

Figure 5.24: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

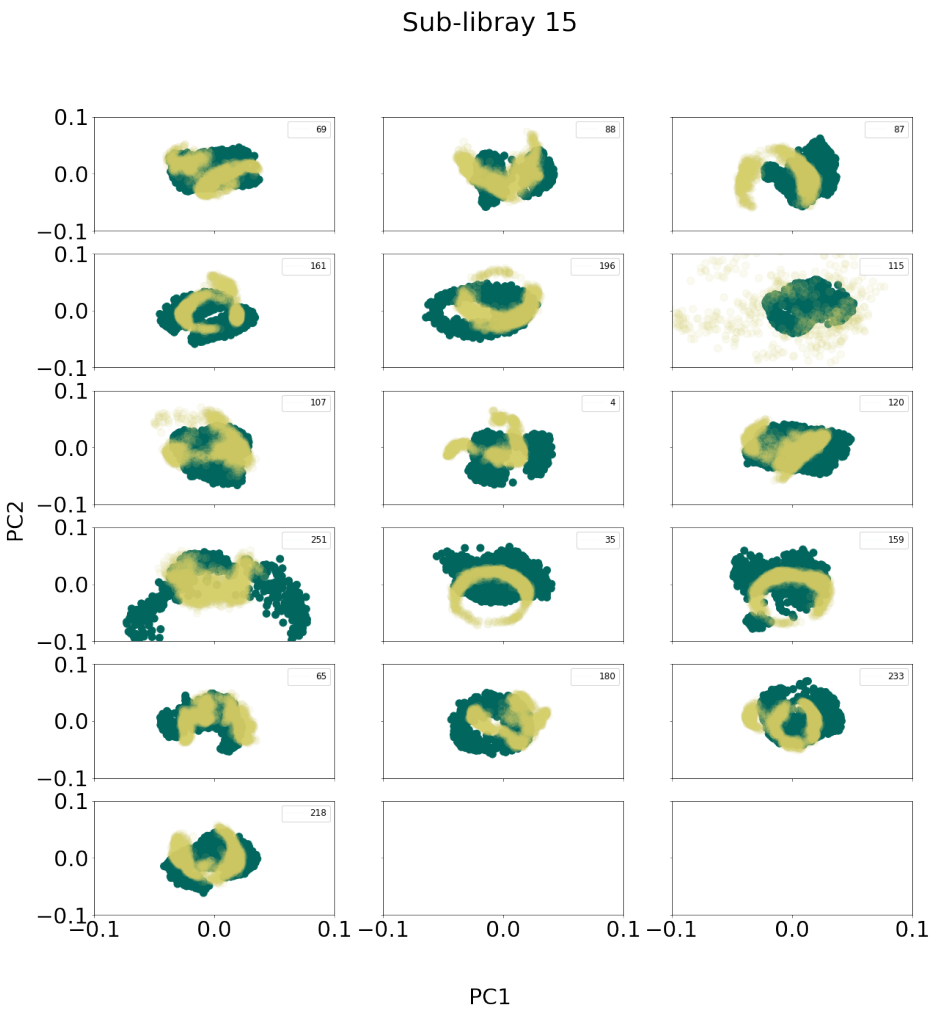

Figure 5.25: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

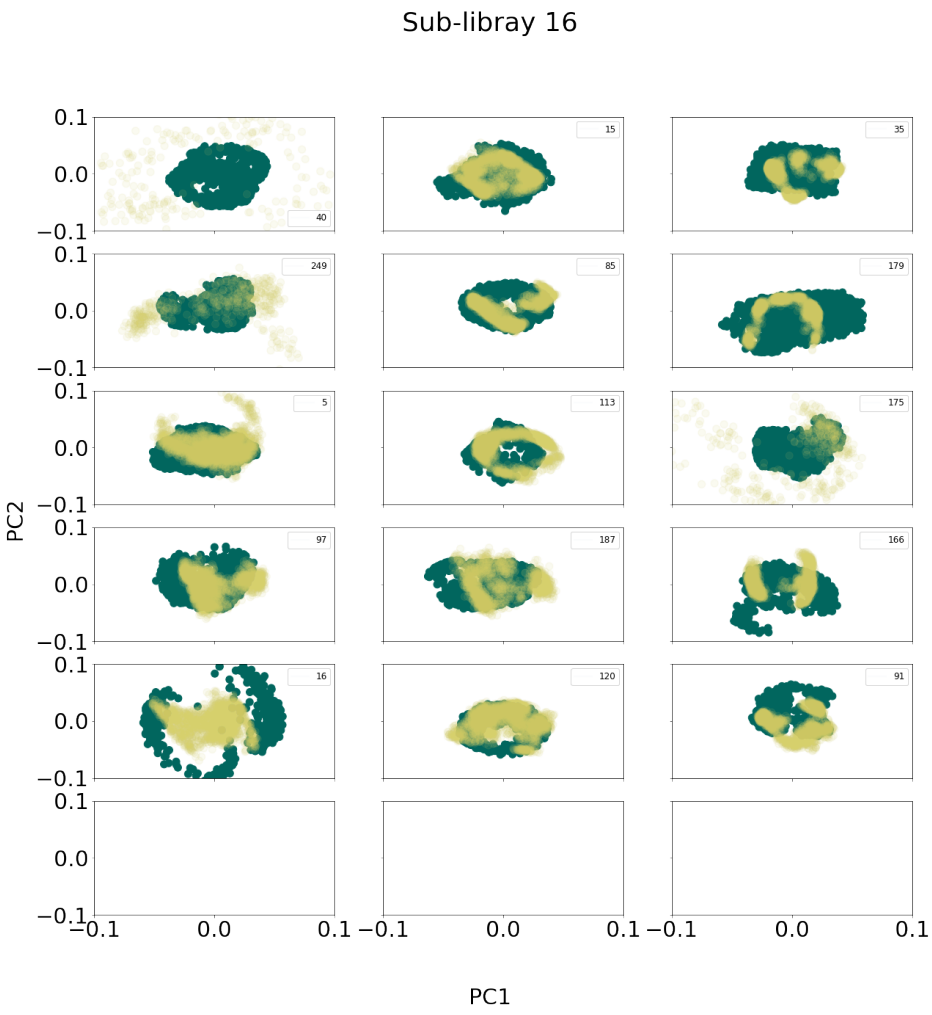

Figure 5.26: Representative plot of the first versus second principal components for cyclic peptide backbones (heavy atoms). *Teal:* water; *Cream:* octanol.

5.4 CONCLUSION

Although our permeability data did not directly compare with experiment, we did see a conserved trend with regards to methylation (permeability increases with methylation). We successfully determined several residues that were key in hydrogen bonding; important to know when trying to design a peptide that needs to fit a protein binding pocket of a particular size. Given the success of PCA to identify peptides that agree with the permeability from experiment, adding PCA to the assay seems promising, perhaps using voronoi tessellation to compare the overlap of PCA from each solvent or applying a machine learning technique to determine differences in PC populations.

5.5 BIBLIOGRAPHY

[1] C. Morrison. “Constrained peptides’ time to shine?” en. In: *Nat rev drug discov* 17.8 (Aug. 2018), pp. 531–533. ISSN: 1474-1776, 1474-1784. DOI: 10.1038/nrd.2018.125. URL: http://www.nature.com/articles/nrd.2018.125 (visited on 04/01/2022).

[2] C. N. Kelly, C. E. Townsend, A. N. Jain, M. R. Naylor, C. R. Pye, J. Schwochert, and R. S. Lokey. “Geometrically Diverse Lariat Peptide Scaffolds Reveal an Untapped Chemical Space of High Membrane Permeability”. en. In: *J. am. chem. soc.* 143.2 (Jan. 2021), pp. 705–714. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.0c06115. URL: https://pubs.acs.org/doi/10.1021/jacs.0c06115 (visited on 03/09/2022).

[3] J. A. van Santen, G. Jacob, A. L. Singh, V. Aniebok, M. J. Balunas, D. Bunsko, F. C. Neto, L. Castaño-Espriu, C. Chang, T. N. Clark, J. L. Cleary Little, D. A. Delgadillo, P. C. Dorrestein, K. R. Duncan, J. M. Egan, M. M. Galey, F. J. Haeckl, A. Hua, A. H. Hughes, D. Iskakova, A. Khadilkar, J.-H. Lee, S. Lee, N. LeGrow, D. Y. Liu, J. M. Macho, C. S. McCaughey, M. H. Medema, R. P. Neupane, T. J. O’Donnell, J. S. Paula, L. M. Sanchez, A. F. Shaikh, S. Soldatou, B. R. Terlouw, T. A. Tran, M. Valentine, J. J. J. van der Hooft, D. A. Vo, M. Wang, D. Wilson, K. E. Zink, and R. G. Linington. “The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery”. en. In: *Acs cent. sci.* 5.11 (Nov. 2019), pp. 1824–1833. ISSN: 2374-7943, 2374-7951. DOI: 10.1021/acscentsci.9b00806.

URL: https://pubs.acs.org/doi/10.1021/acscentsci.9b00806 (visited on 04/01/2022).

[4] J. Damjanovic, J. Miao, H. Huang, and Y.-S. Lin. “Elucidating Solution Structures of Cyclic Peptides Using Molecular Dynamics Simulations”. en. In: *Chem. rev.* 121.4 (Feb. 2021), pp. 2292–2324. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/acs.chemrev.0c01087. URL: https://pubs.acs.org/doi/10.1021/acs.chemrev.0c01087 (visited on 03/29/2022).

[5] J. D. Bowman, W. H. Coldren, and S. Lindert. “Mechanism of cardiac troponin c calcium sensitivity modulation by small molecules illuminated by umbrella sampling simulations”. In: *Journal of chemical information and modeling* 59.6 (June 24, 2019). Publisher: American Chemical Society, pp. 2964–2972. ISSN: 1549-9596. DOI: 10.1021/acs.jcim.9b00256. URL: https://doi.org/10.1021/acs.jcim.9b00256 (visited on 09/26/2022).

[6] C. T. Lee, J. Comer, C. Herndon, N. Leung, A. Pavlova, R. V. Swift, C. Tung, C. N. Rowley, R. E. Amaro, C. Chipot, Y. Wang, and J. C. Gumbart. “Simulation-Based Approaches for Determining Membrane Permeability of Small Compounds”. en. In: *J. chem. inf. model.* 56.4 (Apr. 2016), pp. 721–733. ISSN: 1549-9596, 1549-960X. DOI: 10.1021/acs.jcim.6b00022. URL: https://pubs.acs.org/doi/10.1021/acs.jcim.6b00022 (visited on 03/25/2022).

[7] A. Shoji, C. Kang, K. Fujioka, J. P. Rose, and R. Sun. “Assessing the intestinal permeability of small molecule drugs via diffusion motion on a multidimensional free energy surface”. In: *Journal of chemical theory and computation* 18.1 (Jan. 11, 2022). Publisher: American Chemical Society, pp. 503–515. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.1c00661. URL: https://doi.org/10.1021/acs.jctc.1c00661 (visited on 09/26/2022).

[8] M. Sugita, S. Sugiyama, T. Fujie, Y. Yoshikawa, K. Yanagisawa, M. Ohue, and Y. Akiyama. “Large-Scale Membrane Permeability Prediction of Cyclic Peptides Crossing a Lipid Bilayer Based on Enhanced Sampling Molecular Dynamics Simulations”. en. In: *J. chem. inf. model.* 61.7 (July 2021), pp. 3681–3695. ISSN: 1549-9596, 1549-960X. DOI: 10.1021/acs.jcim.1c00380. URL: https://pubs.acs.org/doi/10.1021/acs.jcim.1c00380 (visited on 03/18/2022).

[9] P. W. J. M. Frederix, J. Idé, Y. Altay, G. Schaeffer, M. Surin, D. Beljonne, A. S. Bondarenko, T. L. C. Jansen, S. Otto, and S. J. Marrink. “Structural and spectroscopic properties of assemblies of self-replicating peptide macrocycles”. In: *ACS nano* 11.8 (Aug. 22, 2017). Publisher: American Chemical Society, pp. 7858–7868. ISSN: 1936-0851. DOI: 10.1021/acsnano.7b02211. URL: https://doi.org/10.1021/acsnano.7b02211 (visited on 09/26/2022).

[10] W. C. Wimley, T. P. Creamer, and S. H. White. “Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides”. In: *Biochemistry* 35.16 (Apr. 23, 1996), pp. 5109–5124. ISSN: 0006-2960. DOI: 10.1021/bi9600153.

[11] C. G. Mayne, J. Saam, K. Schulten, E. Tajkhorshid, and J. C. Gumbart. “Rapid parameterization of small molecules using the force field toolkit”. In: *Journal of computational chemistry* 34.32 (2013). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.23422, pp. 2757–2770. ISSN: 1096-987X. DOI: 10.1002/jcc.23422. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.23422 (visited on 09/18/2022).

[12] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell. “CHARMM36m: an improved force field for folded and intrinsically disordered proteins”. In: *Nature methods* 14.1 (Jan. 2017). Number: 1 Publisher: Nature Publishing Group, pp. 71–73. ISSN: 1548-7105. DOI: 10.1038/nmeth.4067. URL: https://www.nature.com/articles/nmeth.4067 (visited on 09/18/2022).

[13] W. Humphrey, A. Dalke, and K. Schulten. “VMD: Visual molecular dynamics”. en. In: *Journal of molecular graphics* 14.1 (Feb. 1996), pp. 33–38. ISSN: 0263-7855. DOI: 10.1016/0263-7855(96)00018-5. URL: https://www.sciencedirect.com/science/article/pii/0263785596000185 (visited on 05/26/2022).

[14] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. “Scalable molecular dynamics with NAMD”. eng. In: *J comput chem* 26.16 (Dec. 2005), pp. 1781–1802. ISSN: 0192-8651. DOI: 10.1002/jcc.20289.

[15] G. Fiorin, M. L. Klein, and J. Hénin. “Using collective variables to drive molecular dynamics simulations”. In: *Molecular physics* 111.22 (Dec. 1, 2013). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00268976.2013.813594, pp. 3345–3362.

ISSN: 0026-8976. DOI: 10.1080/00268976.2013.813594. URL: https://doi.org/10.1080/00268976.2013.813594 (visited on 09/18/2022).

[16] Y. Miao and J. A. McCammon. “Chapter six - gaussian accelerated molecular dynamics: theory, implementation, and applications”. In: *Annual reports in computational chemistry*. Ed. by D. A. Dixon. Vol. 13. Elsevier, Jan. 1, 2017, pp. 231–278. DOI: 10.1016/bs.arcc.2017.06.005. URL: https://www.sciencedirect.com/science/article/pii/S1574140017300087 (visited on 09/18/2022).

[17] J. Wang, P. R. Arantes, A. Bhattarai, R. V. Hsu, S. Pawnikar, Y.-m. M. Huang, G. Palermo, and Y. Miao. “Gaussian accelerated molecular dynamics: principles and applications”. In: *WIREs computational molecular science* 11.5 (Sept. 2021). ISSN: 1759-0876, 1759-0884. DOI: 10.1002/wcms.1521. URL: https://onlinelibrary.wiley.com/doi/10.1002/wcms.1521 (visited on 09/18/2022).

[18] Y. Miao, V. A. Feher, and J. A. McCammon. “Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation”. In: *Journal of chemical theory and computation* 11.8 (Aug. 11, 2015), pp. 3584–3595. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.5b00436. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.5b00436 (visited on 09/18/2022).

[19] Y. T. Pang, Y. Miao, Y. Wang, and J. A. McCammon. “Gaussian accelerated molecular dynamics in NAMD”. In: *Journal of chemical theory and computation* 13.1 (Jan. 10, 2017), pp. 9–19. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.6b00931. URL: https://pubs.acs.org/doi/10.1021/acs.jctc.6b00931 (visited on 09/18/2022).

[20] T. D. Romo, N. Leioatts, and A. Grossfield. “Lightweight Object Oriented Structure analysis: Tools for building Tools to Analyze Molecular Dynamics Simulations”. In: *J comput chem* 35.32 (Dec. 2014), pp. 2305–2318. ISSN: 0192-8651. DOI: 10.1002/jcc.23753. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4227929/ (visited on 06/23/2022).

[21] T. D. Romo and A. Grossfield. “LOOS: an extensible platform for the structural analysis of simulations”. eng. In: *Annu int conf ieee eng med biol soc* 2009 (2009), pp. 2332–2335. ISSN: 2375-7477. DOI: 10.1109/IEMBS.2009.5335065.

[22] P. Hosseinzadeh, P. R. Watson, T. W. Craven, X. Li, S. Rettie, F. Pardo-Avila, A. K. Bera, V. K. Mulligan, P. Lu, A. S. Ford, B. D. Weitzner, L. J. Stewart, A. P. Moyer, M. Di Piazza, J. G. Whalen, P. J. Greisen, D. W. Christianson, and D. Baker. "Anchor extension: a structure-guided approach to design cyclic peptides targeting enzyme active sites". In: *Nature communications* 12.1 (June 7, 2021). Number: 1 Publisher: Nature Publishing Group, p. 3384. ISSN: 2041-1723. DOI: 10.1038/s41467-021-23609-8. URL: https://www.nature.com/articles/s41467-021-23609-8 (visited on 09/27/2022).

[23] W.-Y. Fang, R. Dahiya, H.-L. Qin, R. Mourya, and S. Maharaj. "Natural proline-rich cyclopolypeptides from marine organisms: chemistry, synthetic methodologies and biological status". In: *Marine drugs* 14.11 (Nov. 2016). Number: 11 Publisher: Multidisciplinary Digital Publishing Institute, p. 194. ISSN: 1660-3397. DOI: 10.3390/md14110194. URL: https://www.mdpi.com/1660-3397/14/11/194 (visited on 09/27/2022).

6. FUTURE DIRECTIONS

Given the work in the previous chapters, here are some additional directions the research could take in the future. The work presented here does not probe the peptide-membrane interactions that are fundamental to the activity of pHLIP. A natural extension would be to simulate pHLIP in proximity to a membrane using constant pH. The method in NAMD[1] allows the user to set the starting pK_as of each residue before calculation; the pK_a determined in chapter 2 are natural starting points for when the peptide is farther from the membrane. In simulations where the peptide starts closer to the bilayer, it may be more appropriate to use the values determined via NMR[2] as initial pK_a values. Another interesting possibility would be executing constant pH MD in tandem with umbrella sampling[3], an enhanced sampling technique that restricts the system within certain windows. Here, the reaction coordinate that would be broken into windows is the distance between the membrane and some key residues in the peptide; this should allow the determination of the pK_as of each residue as the peptide approaches the bilayer as well as the free energy of binding based on the system pH.

The first extension of the work on the lariat peptides in chapter 5 would be to increase the number of peptides sampled to more closely represent the full data set from Kelly, et al.[4]. This would aid in comparison between the data sets and increase confidence in the water-octanol method.

Besides simulating additional peptides, a natural progression of the cyclic peptide work would be utilizing a method, such as umbrella sampling, to calculate the permeability of the lariat peptides. This is the default method for determining permeability with MD, however, it is considerably more expensive than the water-octanol method proposed in chapter 5. While more expensive, this method has been very effective in the past[5, 6] and will allow for a direct comparison (via a measurement such as RMSD) between the structures found in chapter 5 and structures found from the umbrella sampling.

6.1 BIBLIOGRAPHY

[1] B. K. Radak, C. Chipot, D. Suh, S. Jo, W. Jiang, J. C. Phillips, K. Schulten, and B. Roux. "Constant-pH Molecular Dynamics Simulations for Large Biomolecular Systems". In: *Journal of chemical theory and computation* 13.12 (2017-11-22), pp. 5933–5944. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.7b00875.

[2] S. A. Otieno, S. Z. Hanz, B. Chakravorty, A. Zhang, L. M. Klees, M. An, and W. Qiang. "pH-dependent thermodynamic intermediates of pHLIP membrane insertion determined by solid-state NMR spectroscopy". In: *Proceedings of the national academy of sciences of the united states of america* 115.48 (Nov. 27, 2018), pp. 12194–12199. ISSN: 1091-6490. DOI: 10.1073/pnas.1809190115.

[3] N. F. B. Oliveira and M. Machuqueiro. "Novel US-CpHMD protocol to study the protonation-dependent mechanism of the ATP/ADP carrier". In: *Journal of chemical information and modeling* 62.10 (May 23, 2022). Publisher: American Chemical Society, pp. 2550–2560. ISSN: 1549-9596. DOI: 10.1021/acs.jcim.2c00233. URL: https://doi.org/10.1021/acs.jcim.2c00233 (visited on 09/21/2022).

[4] C. N. Kelly, C. E. Townsend, A. N. Jain, M. R. Naylor, C. R. Pye, J. Schwochert, and R. S. Lokey. "Geometrically Diverse Lariat Peptide Scaffolds Reveal an Untapped Chemical Space of High Membrane Permeability". en. In: *J. am. chem. soc.* 143.2 (Jan. 2021), pp. 705–714. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.0c06115. URL: https://pubs.acs.org/doi/10.1021/jacs.0c06115 (visited on 03/09/2022).

[5] M. Sugita, S. Sugiyama, T. Fujie, Y. Yoshikawa, K. Yanagisawa, M. Ohue, and Y. Akiyama. "Large-scale membrane permeability prediction of cyclic peptides crossing a lipid bilayer based on enhanced sampling molecular dynamics simulations". In: *Journal of chemical information and modeling* 61.7 (July 26, 2021). Publisher: American Chemical Society, pp. 3681–3695. ISSN: 1549-9596. DOI: 10.1021/acs.jcim.1c00380. URL: https://doi.org/10.1021/acs.jcim.1c00380 (visited on 09/21/2022).

[6] R. Vijayaraj, S. Van Damme, P. Bultinck, and V. Subramanian. "Molecular dynamics and umbrella sampling study of stabilizing factors in cyclic peptide-based nanotubes". In: *The journal of physical chemistry b* 116.33 (Aug. 23, 2012). Publisher: American

Chemical Society, pp. 9922–9933. ISSN: 1520-6106. DOI: 10.1021/jp303418a. URL: https://doi.org/10.1021/jp303418a (visited on 09/21/2022).