Biophysical Investigation of pHLIP-Membrane Interaction

Chitrak Gupta

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation
https://researchrepository.wvu.edu/etd/7184
Biophysical Investigation of pHLIP-Membrane Interaction

Chitrak Gupta

Dissertation submitted
to the Eberly College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in
C. Eugene Bennett Department of Chemistry

Blake Mertz, PhD, Chair
Justin Legleiter, PhD
Brian Popp, PhD
Jonathan Boyd, PhD
Visvanathan Ramamurthy, PhD

C. Eugene Bennett Department of Chemistry

Morgantown, West Virginia
2018

Keywords: Computational chemistry, Molecular simulation, Membrane-active peptide, Intrinsically disordered peptide, pH-sensitive peptide

Copyright 2018 Chitrak Gupta
Abstract

Biophysical Investigation of pHLIP-Membrane Interaction

Chitrak Gupta

Cancer is one of the leading causes of death today. Cancer tumors are very similar to healthy tissues, which cause a major challenge in detecting cancer early. Additionally, this leads to cancer drugs targeting healthy tissues, resulting in painful off-target side-effects. Additionally, cancer tumors can be extremely heterogeneous, making it difficult to find a common property for targeting. Another compounding factor is the ability of cancer tumors to undergo rapid changes which makes their detection and targeting more difficult. A potential solution is to target the acidic microenvironment of cancer tumors, which is an universal property of all cancer types, and is not subject to rapid mutation.

pH (Low) Insertion Peptide, pHLIP, is a peptide capable of sensing cellular membranes with low (acidic) pH. pHLIP binds to lipid membranes, and under acidic conditions, spontaneously forms a transmembrane helix. pHLIP can translocate cancer drugs and drug-like molecules across the membrane, and pHLIP-conjugated with cancer drugs has shown pH-dependent cell mortality. Biophysical properties of pHLIP including the kinetics and thermodynamics of its interactions with the cell membrane has been extensively studied. However, the acidic pH of most cancer tumors is not low enough for pHLIP to perform its function. The solution is to make subtle changes in the primary sequence of pHLIP to tune its properties and improve its acid-sensing to suit our needs. This requires a molecular-level understanding of the interactions between pHLIP and lipid membranes. Such information is typically inaccessible to traditional biophysical techniques, and computational studies become the technique of choice.
This dissertation focuses on using atomistic molecular dynamics simulations to study pHLIP and its interactions with lipid membranes. Advances in computer hardware, combined with the ability of simulation packages to exploit the same, has made it possible to study large biological systems for a timescale relevant to the phenomena we seek to study. Use of GPU-accelerated simulation and high-performance computing at West Virginia University (WVU) and Extreme Science and Engineering Discovery Environment (XSEDE) resources has enabled us to gain key insights into the pHLIP-membrane interactions. A variety of MD simulation techniques has been used to reveal partially folded conformations of pHLIP in solution and the interplay of multiple non-bonded forces that govern the membrane-binding of pHLIP. The work described in this dissertation will help pave the way for developing pHLIP into a cancer-targeting agent.
"I pray not for my burdens to be reduced
But for the strength to bear the burdens"

Rabindranath Tagore
Acknowledgements

I shall take this opportunity to thank the variety of people without whose help this dissertation would not have been a success.

First of all, I would like to thank my advisor, Dr. Blake Mertz, for his extremely helpful guidance throughout the course of my study in his research group. Dr. Mertz has been more of a friend-philosopher-and-guide for me over the past four years. I would also like to thank all the former and current members of the group for the wonderful experience I have had working with them. Specifically, I would thank Dr. Jun Feng for her help during my initial years in the group, Sadegh Faramarzi for being an amazing friend and colleague, and Yue Ren for starting the pHLIP project in the group that I have continued. I thank my committee members, Dr. Justin Legleiter, Dr. Brian Popp, Dr. Jonathan Boyd and Dr. Visvanathan Ramamurthy for the feedbacks they provided on my progress which has helped improve the quality of my work.

I specially thank Rebecca Secrist, Administrative Assistant, and Brenda Prentiss, Administrative Secretary, at the C. Eugene Bennett Department of Chemistry at West Virginia University, for their help with all the administrative processes.

I would thank my parents for providing their unconditional love and supporting me through my times of extreme struggle. I thank them for teaching me that life is bigger than success and failure, which has given me the strength to keep going irrespective of my circumstances.

Last but not the least, I express very special gratitude to my wife Ananya, for being the biggest and quite often the only support-system I had. This dissertation would not have been possible without her being by my side through the darkest of my times. Additionally, I would like to thank her for critical proofreading of the dissertation.
## Contents

### Abstract

Abstract

### Acknowledgements

Acknowledgements

### The bigger picture

1 The bigger picture

1.1 Challenges in cancer drug delivery

1.2 History of pHLIP

1.2.1 The focus of this work

1.3 Molecular dynamics (MD) simulation

### References

References

### Protonation Enhances inherent Helix-Forming Propensity of pHLIP

2 Protonation Enhances inherent Helix-Forming Propensity of pHLIP

2.1 Introduction

2.2 Implicit solvent model

2.2.1 Why, and why not

2.2.2 Theory of implicit solvation

2.3 Materials and Methods

2.3.1 System setup

2.3.2 Simulation

2.3.3 Analysis

2.4 Results and discussions

2.4.1 Site-Specific Interactions Prevalent in State I

2.4.2 Perturbing the C-Terminal Half of pHLIP Leads to Noticeable Increases in Helicity

2.4.3 pHLIP Transiently Samples Both Major Secondary Conformations in State I
2.4.4 Low-Energy States in Solution Are Favored by C-Terminal Modifications to pHLIP ........................................ 29

2.5 Conclusion .............................................................. 32

2.6 Future directions ....................................................... 33

2.6.1 Theory of (Gaussian) accelerated MD ......................... 34

2.6.2 Methodology .......................................................... 35

2.6.3 Preliminary results .................................................. 38

Reweighting and Gaussian behavior .................................. 38

PMF of the ϕ-ψ space ...................................................... 39

PMF of the Rg-contact space ............................................. 42

2.6.4 Final remarks ........................................................ 44

References ........................................................................ 45

3 Cooperative Non-bonded Forces and Membrane Binding of pHLIP 52

3.1 Introduction ............................................................... 52

3.2 Materials and Methods ............................................... 55

3.3 Results ................................................................. 56

3.3.1 Partitioning of pHLIP influenced by location of acidic residues 56

3.3.2 pHLIP samples secondary structural conformations in state II 61

3.3.3 pHLIP utilizes multiple types of non-bonded interactions in state II ................................................................. 65

3.3.4 Relating simulation results to theory ............................. 70

3.4 Discussion ............................................................... 71

3.5 Conclusion ............................................................... 79

References ........................................................................ 80

4 Binding of pHLIP to bilayer surfaces: Effect of ionic strength 89

4.1 Introduction ............................................................... 89

4.2 Materials and Methods ............................................... 90

4.2.1 System setup and analysis ......................................... 90

4.2.2 Simulation ........................................................... 91

Gaussian Accelerated Molecular Dynamics simulations ........... 91
Equilibrium MD simulations ............................. 92

4.3 Results and Discussions ................................. 92
  4.3.1 Gaussian Accelerated Molecular Dynamics simulations ...... 92
  4.3.2 Integrity of the lipids ................................ 94
  4.3.3 pHLIP binds transiently to POPC bilayers regardless of salt concentration ........................................ 94
  4.3.4 PC headgroups decrease diffusion of Na\(^+\) ions. ............ 96

4.4 Discussion ................................................. 98

4.5 Conclusion ................................................. 101

References ...................................................... 103

5 Influence of Bilayer Electrostatics On Membrane Binding of pHLIP110
  5.1 Introduction .............................................. 110
  5.2 Choice of MD simulation technique ........................ 111
  5.3 Methods .................................................. 113
    5.3.1 System setup ....................................... 113
    5.3.2 Replica Exchange with Solute Tempering (REST2) ....... 114
    5.3.3 Quantifying efficiency of exchange ................... 114
  5.4 Results .................................................. 119
    5.4.1 Initial attempts ..................................... 119
    5.4.2 Binding of pHLIP to 100% POPC bilayer ................. 120
    5.4.3 Presence of POPS: bilayer electrostatics ............... 122
    5.4.4 Elimination of state II fluorescence spectra .......... 125
    5.4.5 Ongoing work ....................................... 127

References ...................................................... 132
List of Figures

1.1 Crystal structure of bacteriordhopsin. Helix C, from which pHLIP is derived, is shown in blue ........................................ 2
1.2 Steps involved in pHLIP binding and insertion. pHLIP is unstructured in basic solution (pH 8) (State I). When pH is around 6.5, pHLIP binds lipid membrane surface and is still unstructured (State II). When pH reduces below 6, pHLIP forms α-helix, followed by insertion into the membrane (State III) ............................................................................. 3
1.3 Convergence properties of Lennard-Jones (red) and Coulombic (green) interaction with distance between the two interacting atoms. Energy is plotted in arbitrary units. Lennard-Jones interaction decays quickly to zero, whereas Coulombic interaction decays very slowly ....................................................... 7
2.1 pHLIP in water box (explicit solvation). System size: 69,240 atoms ................................................................. 18
2.2 pHLIP in continuum solvent (implicit solvation). System size: 560 atoms .............................................................. 18
2.3 Heating to 700 K of pHLIP does not lead to cis conformations in the peptide backbone .................................................. 21
2.4 Modifications to pHLIP lead to site-specific interactions in state I ................................................................. 23
2.5 Increased interactions do not correlate with the helicity of pHLIP ................................................................. 25
2.6 pHLIP samples multiple secondary structures in solution ................................................................. 27
2.7 pHLIP samples multiple secondary structures in solution. .............................................................................. 28
2.8 D25Aad sidechain forms salt bridge with R11 of pHLIP only in presence of D14Gla mutation. ........................................ 30
2.9 Number of contacts is loosely correlated with the radius of gyration in pHLIP. .......................................................... 31
2.10 End-to-end distance of pHLIP as a function of GaMD simulation time. ................................................................. 36
2.11 Evolution of k0 (see definition above) as a function of time for three different $\sigma_{0P}$. .................................................. 37
2.12 Distribution of boost potential in 830 ns dual-boost GaMD simulation. ................................................................. 39
2.13 Free energy profile and bin-wise anharmonicities of $\phi$ and $\psi$ dihedral angles ........................................................ 40
2.14 Two dimensional free energy profile and anharmonicity of backbone dihedral angles ............................................ 41
2.15 Free energy profile and bin-wise anharmonicities of radius of gyration and number of contacts ................................. 42
2.16 Two dimensional free energy profile and anharmonicity of contact and radius of gyration ....................................... 43

3.1 Systems investigated in this study. ........................................ 54
3.2 Quality of binding of pHLIP can depend on orientation and peptide composition. ...................................................... 57
3.3 Effective binding of pHLIP to POPC is related to specific segments of the peptide. .................................................... 59
3.4 $^{31}$P-$^{13}$C distances reveal that orientation of the pHLIP binding complex has varying degrees of influence. ................. 60
3.5 Aggregate sampling of secondary structure of wt-pHLIP and pHLIP-1 bound to POPC. ............................................. 62
3.6 Formation of helical turns occurs during binding for both wt-pHLIP and pHLIP-1. .......................................................... 63
3.7 $i+3$ salt bridge formation plays a key role in stabilization of secondary structures in pHLIP-1. ................................. 64
3.8 Dehydration of residues in pHLIP weakly correlates with partitioning into the surface of the POPC bilayer.............. 66
3.9 Non-bonded interactions can contribute to binding of pHLIP but are not necessary for complex formation. .............. 67
3.10 R11 consistently utilizes hydrogen bonds to stabilize interactions of pHLIP with a POPC bilayer surface. .............. 68
3.11 Aromatic and cationic sidechains exert distinct effects on partitioning of pHLIP into POPC. ....................... 69
3.12 Wimley-White partition free energies provide insight into optimal binding of pHLIP. .............................. 71
3.13 Sidechain orientation of tryptophans have variable effect on binding of pHLIP. ........................................ 72
3.14 Sidechain orientation of tyrosines have variable effect on binding of pHLIP. ........................................ 73
3.15 Proposed binding mechanism of pHLIP to POPC. ........... 78
4.1 GaMD acceleration obtained by NAMD vs. AMBER simulation engine ......................................................... 93
4.2 Area-per-lipid (left) and lipid order parameter of PA and OL chains of POPC (right) with and without pHLIP. ........ 94
4.3 Z-projection of distance of Ca of every residue from the P atoms of the nearest leaflet ................................. 95
4.4 Binding of sodium ions to PC headgroups. ....................... 97
4.5 Increasing salt concentration affects intramolecular interactions in pHLIP ................................................. 99
4.6 Increase in salt concentration modulates localized formation of helices in state II ................................. 100
4.7 Ionic strength reduces the area-per-lipid of POPC bilayer. 102
5.1 Exchange efficiency of REST2 simulation involving 8 replica spanning 310 K to 430 K. .......................... 115
5.2 pHLIP binds POPC that is stable over 90 ns REST2 simulation115
5.3 Exchange efficiency of REST2 simulation involving 16 replica spanning 310 K to 500 K. ................................. 116
5.4 POPC-binding of pHLIP is abolished with 16 replicas spanning 310 K to 500 K. ................................. 116
5.5 Exchange efficiency of REST2 simulation involving 16 replica spanning 310 K to 500 K and exchange attempt every 10 ps. 117
5.6 POPC-binding of pHLIP is abolished with 16 replicas spanning 310 K to 500 K with reduced frequency of exchange attempt. ................................................................. 117
5.7 Exchange efficiency of REST2 simulation involving 12 replica spanning 310 K to 500 K. ................................. 118
5.8 POPC-binding of pHLIP is abolished with 12 replicas spanning 310 K to 500 K. ................................. 118
5.9 Exchange efficiency of REST2 simulation involving 10 replica spanning 310 K to 430 K. ................................. 121
5.10 pHLIP establishes stable contact with POPC in spite of frequent exchange. ......................................................... 121
5.11 Exchange efficiency of REST2 simulation of pHLIP in a bilayer of 75% POPC and 25% POPS. ................................. 123
5.12 Exchange efficiency of REST2 simulation of pHLIP in a bilayer of 50% POPC and 50% POPS. ................................. 123
5.13 Anionic C-terminus of pHLIP is away from the surface of bilayer containing 25% POPS. ................................. 124
5.14 N-terminus of pHLIP establishes contact even with bilayer containing 50% POPS. ................................. 124
5.15 A subpopulation where W9 is buried and dehydrated is eliminated in the presence of POPS ................................. 125
5.16 W15 does not show a subpopulation where it is buried and dehydrated even in 100% POPC. ................................. 126
5.17 Exchange efficiency of REST2 simulation of pHLIP in a bilayer of 75% POPC and 25% POPE. ................................. 128
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.18</td>
<td>Exchange efficiency of REST2 simulation of pHLIP in a bilayer of 75% POPC and 25% POPE.</td>
<td>128</td>
</tr>
<tr>
<td>5.19</td>
<td>Anionic C-terminus of pHLIP is away from the surface of bilayer containing 25% POPE.</td>
<td>129</td>
</tr>
<tr>
<td>5.20</td>
<td>Anionic C-terminus of pHLIP is away from the surface of bilayer containing 25% POPE.</td>
<td>129</td>
</tr>
<tr>
<td>5.21</td>
<td>A subpopulation where W9 is buried and dehydrated is eliminated in the presence of POPE</td>
<td>130</td>
</tr>
<tr>
<td>5.22</td>
<td>A subpopulation where W9 is buried and dehydrated is eliminated in the presence of POPE</td>
<td>131</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Free energy minima for each pHLIP variant tested, based on 2D free energy landscape (number of contacts vs. radius of gyration, Figure 2.9) 32

3.1 Per-residue and cumulative free energy of partitioning of wt-pHLIP and pHLIP-1 based on the Wimley-White scale. 75

4.1 Table of the systems under study 91

4.2 Average residence time of Na\textsuperscript{+} ions near lipid phosphate oxygens 97
Dedicated to all those who keep fighting in spite of their circumstances
Chapter 1

The bigger picture

1.1 Challenges in cancer drug delivery

The American Cancer Society estimates more than a million people get cancer every year in the US alone. There are two primary challenges facing oncology today: early diagnosis\cite{34} and efficient cancer targeting. These challenges are tied to the difficulty associated with distinguishing cancer tumors from healthy tissues. Thus, conventional treatments attack healthy cells in addition to cancerous ones\cite{18}, leading to off-target side effects. Many of these side effects are extremely painful. One alternative approach to targeting cancer tumors utilize monoclonal antibodies\cite{41} and antibody drug conjugates (ADCs)\cite{24}. These rely on cancer-specific antigens to recognize cancer tumors. However, this approach has two potential pitfalls: heterogeneity of cancer tumors (different antigens for different cancer types and sometimes even two tumors of the same cancer type), and rapid mutation of the antigens being targeted. The latter gives rise to drug resistance.

These problems could be overcome by designing an approach to target unique characteristics of tumor microenvironment. One such characteristic is hypoxia or acidity of the exterior of the cell. Cancerous tumors have an extracellular environment that is more acidic (pH 6.5 - 7.0) than healthy tissues (pH 7.2 - 7.5)\cite{8}. Such tumor acidosis stems from altered and enhanced metabolism whereby cancer cells utilize glycolysis for ATP synthesis (Warburg effect\cite{39}), causing an excess of lactic acid that is pumped outside the cell. This acidosis is a universal feature of all cancer cells and possible exploitation of such cellular pH gradient for cancer treatment has been
hinted at as early as in 1996[9]. Thus, developing a therapeutic and/or diagnostic that targets this property could have a big impact on oncology.

This work focuses on answering some of the fundamental questions which are of critical importance in creating a cancer drug delivery system based on pHLIP.

1.2 History of pHLIP

Back in 1997, the group of Prof. Donald Engelman at Yale University were studying bacteriorhodopsin (Figure 1.1), a protein that has seven transmembrane (TM) helices. They asked a fundamental question: would synthetic peptides corresponding to each of these helices be able to form similar TM helix (taken out of their tertiary context)?

One of the results of their study[14, 15] was rather interesting: helix C (shown in blue in Figure 1.1) forms TM helix in a pH-dependent manner. This peptide has came to be known as pH (Low) Insertion Peptide or pHLIP, and has been the subject of extensive biophysical study in the last 20 years. The pH-sensitive behavior of pHLIP is shown in Figure 1.2.

This peptide is soluble and unstructured in solution (State I, Figure 1.2) and binds to surface of lipid membranes under neutral and alkaline pH (State II, Figure 1.2). When the pH drops to acidic level, pHLIP spontaneously forms a transmembrane helix (state III, Figure 1.2). This property makes pHLIP a great candidate for cancer diagnostic and therapeutic.

pHLIP has the ability to translocate cell impermeable cargo into cells at low pH[30, 2, 35, 23]. It has also been shown that apparent pK of insertion of pHLIP into membranes can be modulated by changing the peptide composition to either facilitate[17, 40] or hinder[25, 4] the process. A recent review summarizes the properties of pHLIP and their potential applications[6]. It is becoming increasingly clear that pHLIP peptides can be developed into diagnostic and therapeutic agents for targeted delivery into acidic tissues. Most
cancerous cells have an acidic environment, which makes pHLIP an exceptional candidate for targeting cancerous tumors[1]. However, pHLIP insertion happens at a pH (~6.2) that is much lower than the pH of most cancerous tumors. Thus, modifications to the pHLIP sequence are necessary to develop pHLIP into a novel cancer-targeting agent. An additional compounding factor is pHLIP’s propensity to aggregate, which limits the concentration in which it can be studied. There have been attempts to study the details of pHLIP aggregation[26] and modify the aggregation propensity and insertion pKa (pH where 50% of pHLIP molecules are inserted) of pHLIP[7]. Another interesting result from a study involving unnatural mutants of pHLIP[27] is that D14 and D25 residues could be independently and additively modified to tune the insertion pH and the sharpness of transition. However, the atomistic details of membrane binding and insertion of pHLIP are not known. This lack of knowledge severely hinders the development of pHLIP as a clinically useful drug delivery agent.

Binding of pHLIP to a vesicle surface has been shown to be[29] highly favorable ($\Delta G_{\text{bind}} = -7.2$ kcal/mol), while pHLIP insertion at acidic pH is moderately favorable ($\Delta G_{\text{insertion}} = -1.8$ kcal/mol).

![Figure 1.2: Steps involved in pHLIP binding and insertion.](image)

Both these free energies are large enough to compensate for the free energy penalty of translocating polar cargo molecules across the lipid bilayer. Consequently, the biophysics of pHLIP binding and insertion has been extensively studied. Fluorescence quenching studies using brominated phospholipids showed that the two tryptophans have different burial depths inside the bilayer in state II[42]. A Deep-UV Resonance
Raman (DUVRR) spectroscopy study showed that the tryptophan residues are desolvated upon binding[5]. The same study also reported partial folding of pHLIP at alkaline pH. More recently, solid-state NMR experiments have revealed the N-terminus of pHLIP to be the binding end while the C-terminus exists in an aqueous environment, possibly due to repulsion from the anionic C-terminus[32]. The authors also observed different binding conformations at pH 7.4, 6.4 and 5.3, which suggests that the mechanism of pHLIP binding might be more complex than previously assumed. A follow-up study from these authors showed that the C-terminal acidic residues are the first to be titrated (they also measured precise pKa values of each of the C-terminal aspartates)[13].

In spite of these efforts, a complete, molecular-level description of the events leading to binding and insertion of pHLIP is missing. This lack of understanding causes a severe hindrance to the development of pHLIP into a cancer-targeting agent. The next subsection describes the questions pertaining to pHLIP-membrane interaction that this work focuses on.

1.2.1 The focus of this work

Of critical importance is the folding-partitioning coupling of pHLIP, and how protonation affects the same. This requires a detailed understanding of the protonation effects on secondary structure of pHLIP and how the effect differs in pHLIP variants. In addition, there are two key factors that this work focusses on.

Until the time of the initiation of this work, influence of bilayer composition on pHLIP binding had not received much attention[42, 3]. Most of the work studied interactions of pHLIP to a simple, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer. However, it is known that cancerous cells typically contain higher levels of 1-palmitoyl-2-oleoyl-sn glycero-3-phospho-L-serine (POPS) in their outer leaflet[36], which is in contrast to healthy mammalian cells[21]. The problem becomes all the more interesting given pHLIP itself is an atypical cell-penetrating peptide in that it has an overall negative charge on its surface. Thus, a detailed investigation of the effect of bilayer electrostatics on pHLIP-binding will have far-fetching impact on the development of pHLIP into a cancer-targeting agent. In the last 3
years, experimental efforts have been targeted at understanding the influence of lipid composition\cite{16}, and specifically, the influence of anionic headgroups\cite{31, 19, 37} on lipid-binding and insertion of pHLIP.

Another factor entirely ignored at the time of initiating this work is the influence of ionic strength. Most experimental studies of pHLIP have been performed under low salt concentration (1-20 mM) to avoid aggregation of pHLIP at higher ionic strength\cite{28, 29, 17, 5, 19, 31, 32, 13, 16}.

This dissertation is organized in the following way: The following section describes the fundamentals of molecular dynamics simulation, the technique of choice for answering such questions. Chapter 2 is a detailed study of the partial secondary structure of pHLIP and the influence of site-specific mutation and protonation on the same. Most of this chapter is adapted from the implicit solvent simulations of pHLIP\cite{11}. Chapter 3 compares the membrane interactions of pHLIP and a pHLIP-variant, adapted from\cite{10}. Chapter 4 describes the influence of ionic strength on membrane-binding of pHLIP. An finally chapter 5 describes the influence of bilayer electrostatics on pHLIP-membrane interactions.

1.3 Molecular dynamics (MD) simulation

Molecular dynamics is an approach to predict the state of a system at a time t, given its state at initial time (defined t=0). State of a system is defined by the position and momentum (velocity) of each particle in the system. Thus, the state of an N-particle system can be completely defined by 6N variables. Running a molecular dynamics simulation is essentially solving a 6N-dimensional equation, subject to constraints. These constraints arise from the topology of the molecule (bond length, bond angle, dihedral angle) and the forces acting on the particles (force field) due to the interaction energies. For atomistic simulations (as is described in this proposal), every atom is treated as a separate particle. The atoms are assumed to be hard spheres following
Chapter 1. The bigger picture

classical mechanics:

\[
\begin{align*}
v_i(t + \delta t) &= v_i(t) + f_i(t)\delta t/m_i \quad (1.1a) \\
r_i(t + \delta t) &= r_i(t) + v_i(t + \delta t)\delta t \quad (1.1b)
\end{align*}
\]

where \( v \) is velocity, \( r \) is position, \( f \) is the force, and \( \delta t \) is the simulation time step. The subscript \( i \) represents the \( i \)-th particle. The initial positions of the atoms are usually taken from known X-ray or NMR structure, while initial momenta are generated for a given temperature by Maxwell-Boltzmann distribution. The interaction energies (\( V_{\text{total}} \)) are usually modeled as:

\[
V_{\text{total}} = V_{\text{bond}} + V_{\text{angle}} + V_{\text{dihedral}} + V_{\text{improper}} + V_{\text{LJ}} + V_{\text{coulomb}}
\]

(1.2)

where the terms on the right hand side mean, respectively, the energies of oscillation of two atoms about equilibrium bond length, oscillation of three atoms about equilibrium bond angle, torsional rotation of four atoms, rotation about improper dihedral angle, van der Waals interaction (modeled by Lennard-Jones potential), and coulombic interaction. The functional form of each one of them is given below.

These interactions can be classified as bonded (equation 1.3 a-d) and non bonded (1.3 e-f). The former applies to specific groups of atoms and does not change during the simulation (i.e. chemical reactions are not simulated). The latter are pairwise interactions defined for every pair of atoms in the system. Thus, there are \( N(N-1) \) such interactions in a \( N \)-particle system. For biomolecular simulations, \( N \) is typically large (~50,000), resulting in significant computational cost.
Chapter 1. The bigger picture

\[ V_{\text{bond}}(r) = \frac{1}{2} k_b (r - r_0)^2 \] (1.3a)

\[ V_{\text{angle}}(\theta) = \frac{1}{2} k_{\theta} (\theta - \theta_0)^2 \] (1.3b)

\[ V_{\text{dihedral}}(\phi) = \frac{1}{2} k_{\phi} (\phi - \phi_0)^2 \] (1.3c)

\[ V_{\text{improper}}(\psi) = \frac{1}{2} k_{\psi} (\psi - \psi_0)^2 \] (1.3d)

\[ V_{\text{LJ}}(r) = \epsilon \left( \left( \frac{R}{r} \right)^{12} - 2 \left( \frac{R}{r} \right)^{6} \right) \] (1.3e)

\[ V_{\text{coulomb}}(r) = \frac{q_1 q_2}{4 \pi \epsilon_0 r} \] (1.3f)

A first step towards reducing this cost is to recognize that van der Waals interactions (6e) are a short-range interaction, in that it decays rapidly to zero as opposed to coulombic interactions (figure 1.3). These long-range interactions are mostly treated using particle mesh ewald (PME) method wherein the interactions are separated into short and long range terms and the latter is fourier transformed resulting in much quicker convergence.

As the system evolves in time under the influence of the energies (parameterized by the force fields), we assume that, given sufficient time, the system will explore all possible conformations. In other words, at long times, a time average is equivalent to an ensemble average. This is called the ergodic hypothesis, something that is fundamental to molecular dynamics simulations. This in turn implies that a system should be allowed to evolve for sufficiently long times (ideally ten times the timescale of the

![Figure 1.3: Convergence properties of Lennard-Jones (red) and Coulombic (green) interaction with distance between the two interacting atoms. Energy is plotted in arbitrary units. Lennard-Jones interaction decays quickly to zero, whereas Coulombic interaction decays very slowly](image)
When the system has been allowed to evolve for long enough, the ensemble-average of a quantity $A$ can be calculated as:

$$
< A > = \frac{\sum A_i \exp(-\beta V)}{\sum \exp(-\beta V)}
$$

where $\beta$ is Boltzmann constant, $V$ is the potential energy and the subscript $i$ represents the individual conformations of an ensemble.

Attaining such long timescales is a big challenge for studying biophysical processes, some of which have very long characteristic times, of the order of milliseconds or even seconds. Such timescales are unrealistic as far as MD simulation is concerned. The problem is exacerbated by the fact that for all-atom simulations, we are restricted in the simulation time step that we can choose. This restriction is imposed by the most labile bond, i.e. bonds between a hydrogen and a heavy atom, as it fluctuates very rapidly. Restraining this bond allows a bigger timestep, but it is still in the femtosecond regime, requiring a huge number of simulation steps to attain biologically relevant, nanosecond timescales.

This problem can be avoided in a number of ways. One, of course, is what is commonly called “brute-force” method which implies simply running long, multiple simulations of the system so as to completely explore the phase space. While this method has been used in chapters 3 and 4, it is easy to understand that a better solution would be required in multiple situations. There are many such techniques available, commonly known as “advanced sampling method”. There also methods to speed up the simulation compared to traditional, atomistic simulations. Chapter 2 has utilized implicit solvation, a technique designed to reduce the system size and hence make the simulation faster. Some of the advanced sampling techniques used in this work is replica exchange\[33, 20, 38]\ in chapter 5 and gaussian accelerated molecular dynamics\[12, 22]\ in chapter 2.
References


REFERENCES


[34] Sarah P. Thayer et al. “Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis”. In: Nature 425.6960 (Oct. 2003), pp. 851–856. issn:
REFERENCES

0028-0836. DOI: 10.1038/nature02009. URL: http://www.nature.com/nature/journal/v425/n6960/full/nature02009.html.


Chapter 2

Protonation Enhances inherent Helix-Forming Propensity of pHLIP

2.1 Introduction

Cell-penetrating peptides (CPPs) are a class of molecules with potential applications ranging from antimicrobial agents to vehicles for drug delivery[7]. The pH Low Insertion Peptide (pHLIP) is a fairly unique CPP, in that it is highly anionic (overall charge of -5), long (AEQNPILYAWDYLFTTPLLLDLALLVDADEGT), and sensitive to changes in pH[18, 19]. In solution, pHLIP is in a coiled conformation (state I); when exposed to the cell membrane under alkaline conditions, pHLIP binds to the membrane surface, remaining in a coiled conformation (state II); upon acidification of the environment, the acidic residues in pHLIP are protonated, leading to folding into an α-helix and insertion into the membrane (state III)[34]. Although circular dichroism (CD) and fluorescence spectroscopy are commonly used to monitor the transition of pHLIP from state I → state II → state III, they cannot provide atomistic insights into the interactions that characterize each state. In particular, the behavior of pHLIP in solution (i.e., state I) is poorly understood. This lack of understanding is an issue, as most experiments with pHLIP require low peptide concentrations (<20 μM) to avoid aggregation[11]. In addition, point mutations of pHLIP have been utilized in attempts to improve the acid sensitivity of the peptide[43, 3, 41], to varying degrees of success. Often these mutations lead to loss of acid sensitivity or increased
aggregation effects. One recent study, in particular[32], in which non-natural amino acids were substituted at positions 14 and 25 in pHLIP, led to enhanced insertion properties.

In light of these recent developments, we set out to utilize long-time-scale molecular dynamics (MD) simulations of pHLIP in state I to provide a molecular-level characterization of the behavior of pHLIP in solution. By using implicit solvent in our simulations, it is possible to routinely access microsecond time scales to significantly enhance our ability to completely sample the conformational landscape of pHLIP in state I. We used a recently developed fast pairwise Generalized Born (GB) implicit solvent model that has been successfully applied to protein folding on the microsecond time scale[30, 29].

The first set of variables tested was the protonation state of aspartate residues in pHLIP. Earlier work on pHLIP had established that pHLIP undergoes approximately two protonation events during folding and insertion[35, 19]. In addition, point mutations of D14 and D25, the two interior aspartate residues, led to complete loss of acid sensitivity[27]. The combination of these results led to a conventional belief that D14 and D25 were the “protonation switches” controlling pHLIP folding and insertion[2]. However, more recent solid-state NMR studies have revealed that the C-terminal residues (D31 and D33) are titrated first under acidic conditions and thus may play a direct role in the function of pHLIP[16]. We thus examined singly protonated residues (D14, D25, D31, and D33) as well as two combinations of titrated residues (D14/D25 and “All”, D14, D25, D31, and D33).

2.2 Implicit solvent model

2.2.1 Why, and why not

As mentioned in the previous section, we need to access microsecond timescales in order to completely sample the conformational space of pHLIP in solution. Unbiased atomistic molecular dynamics simulation would make such timescales unattainable. For example, pHLIP placed within a water box (Figure 2.1) is a system comprised of 69,240 atoms. Simulating this system on NVIDIA GTX 780 GPU using pmemd.cuda
(Amber 16) had a performance of 25 ns/day. Thus, simulating 1 μs would require 40 days. Note that typically we require multiple copies of such simulations to ensure that our conclusions are actually meaningful and not an artifact of the initial conditions.

However, we do not necessarily need to simulate such a large system. Out of the 69,240 atoms, only 560 of them belong to pHLIP. Thus, > 99% of the computation time is being spent simulating water molecules that we are not interested in. This problem can be circumvented by the use of continuum solvent models or implicit solvents.

An easy way to think of this is to imagine we are simulating pHLIP in vacuum, but a special kind of vacuum whose dielectric constant matches that of water. We no longer need to keep track of all the water atoms, which reduces our system size drastically (Figure 2.2).

Using this setup improves the performance from 25 ns/day to 650 ns/day on the same NVIDIA GPU’s. This makes the aforementioned timescales easily attainable. However, it should be kept in mind that such performance improvement does come at a corresponding cost. For example, in the absence of explicit waters, salt bridge and hydrogen bonds formed between water and pHLIP can not be seen. Neither can we observe the formation of water shells around the peptide. Moreover, although the screening effect of a given salt concentration can be included, it is not possible to have explicit ions in the system (such study can be important for pHLIP, see chapter 4). This is
Chapter 2. Protonation Enhances inherent Helix-Forming Propensity of pHLIP

because implicit solvents are modeled at infinite dilution where the term “concentration” is not well-defined. For the same reason, studying interaction between multiple peptide molecules (example: aggregation) becomes challenging in implicit solvents, although some MD engines (like CHARMM and OpenMM) does allow the definition of a periodic solvation box even for implicit solvent simulations (see the work by Strodel et. al.[39])

In the following section, I provide a fundamental theoretical description of implicit solvent models. A more detailed description can be found in chapter 4 (“The Generalized Born/Surface Area Model”) of Amber 16 manual.

2.2.2 Theory of implicit solvation

The solvation energy of a solute can be estimated as: $\Delta G_{solv} = \Delta G_{el} + \Delta G_{nonel}$ where $\Delta G_{nonel}$ is the free energy of solvating the solute with the partial charges of all its atoms set to 0, and $\Delta G_{el}$ is the sum of free energy of removing these partial charges in vacuum and that of adding these charges back in the presence of the implicit solvent. In the Amber implementation, $\Delta G_{nonel}$ is assumed to be proportional to the total solvent accessible surface area of the solute. As noted by Nguyen et. al.[29], calculation of $\Delta G_{nonel}$ is usually omitted (‘gbsa = 0’ option in amber simulation) as the term is much smaller than $\Delta G_{el}$, including it significantly reduces performance, and even then the results are of questionable accuracy.

Thus, the free energy of solvation is calculated by the following steps

$$
\Delta G_{solv} = \Delta G_{el} + \Delta G_{nonel}
$$

$$
\approx \Delta G_{el} + \gamma.SASA \quad [AMBER \ implementation]
$$

$$
\approx \Delta G_{el} \quad [\gamma.SASA << \Delta G_{el}]
$$

$$
\approx \frac{1}{2} \sum_{k=1}^{N} \frac{q_{ik} q_{jk}}{f_{GB}(r_{ik}, R_{i}, R_{j})} \left[1 - \frac{\exp(-\kappa f_{GB})}{\epsilon}\right]
$$

where $\kappa$ is the Debye-Huckel screening parameter (controlling the influence of ionic strength), $\epsilon$ is the dielectric constant of water, $q_k$ and $R_k$ are, respectively, the partial
Chapter 2. Protonation Enhances inherent Helix-Forming Propensity of pHLIP

charge and the effective Born radius of the atom $k$, $r_{ij}$ is the distance between the atoms $i$ and $j$, and $f_{GB}$ is a smooth function that interpolates between the limits $r_{ij} \to 0$ and $r_{ij} \to \infty$.

Amber offers a variety of different implicit solvent models that can be chosen with the ’igb’ keyword. These models mostly vary in the way they define the effective Born radius and/or the $f_{GB}$ function. In this work, I have used the “GBNeck2” implicit solvent model\cite{29, 30} (igb=8) developed by the group of Carlos Simmerling.

2.3 Materials and Methods

2.3.1 System setup

pHLIP was taken from helix C of bacteriorhodopsin (residues 73–107 of protein data bank (PDB) 2NTU), and GLY73 was mutated to ALA. This results in pHLIP 2–36, as in Karabadzhak et al. (referred to as pHLIP-4)\cite{22}. The peptide was then solvated and ionized using visual molecular dynamics (VMD)\cite{17}, with the CHARMM36 protein force field\cite{4} used for heating simulations. The peptide was gradually heated to 700 K over 20 ps, followed by 980 ps of production to denature it from the helical conformation. A Langevin thermostat was used to maintain constant temperature, and heating simulation was carried out using NAMD2.9\cite{33}. It was verified that the heating did not lead to any cis conformations of $\omega$ in the peptide backbone (Figure 2.3).

The aspartates were protonated in VMD\cite{17} using the psfgen plugin, and these residues were renamed as ASH as per the Amber naming convention. Convpdb\cite{10} was used to convert PDB files into the Amber format, and Amber input files were generated using tleap\cite{5}. ff14SBonlysc\cite{30} and mbondi3 intrinsic radii were used. Unnatural amino acids were parameterized using the R.E.D. server\cite{9}. Briefly, a PDB file for the residue (along with a N-terminal ACE patch and a C-terminal NME patch) was constructed using Avogadro\cite{15},(cite) “$a$” and “$b$” conformers were generated by setting the $\phi/\psi$ values to -53/-47 and -119/119, respectively, optimizing each structure using Gaussian09\cite{12} and uploading the results to the R.E.D. server.
After heating pHLIP to 700 K to denature the peptide, the dihedral angle of the peptide backbone ($\omega$) was calculated to ensure that isomerization from trans to cis (a non-natural conformation for peptides) had not taken place. All residues in pHLIP remained in the trans conformation (i.e., +/-180 deg).

2.3.2 Simulation

All simulations were run in Amber16\cite{5} using the GB-Neck2 implicit solvent model\cite{29}. Each peptide system was simulated for 2 $\mu$s. Snapshots were taken every 5 ps, providing 400 data points for 2 $\mu$s simulation time (similar to Nguyen et. al\cite{30}.). This sampling was used for each analysis, except for the free energy and salt bridge analyses, which used snapshots every 100 ps (i.e., 20,000 data points for the 2 $\mu$s simulation).

2.3.3 Analysis

Clustering: Trajectories were grouped into 50 clusters using the K-means clustering algorithm in cpptraj of AmberTools\cite{5}. The Cα atoms of residues 10–33 (putative binding domain as per Hanz et. al.\cite{16}) were used for clustering. The top (i.e., most
populated) n clusters were selected for analysis, where n was chosen to include ~40% of the trajectory. Helicity: A residue was defined to be in helical conformation if it simultaneously satisfied the following two conditions: $-90 < \phi < -30$ and $-77 < \psi < -17$ (as per García and Sanbonmatsu). A stretch was defined to be helical if three or more consecutive residues satisfied the above-mentioned condition. $\phi/\psi$ values were calculated in VMD using custom-made tcl scripts. Contact maps: The distance between Cα atoms of residues was used. Distances were calculated in VMD. Free energy analysis: For the number of contacts, the contacts command in cpptraj was used to count the number of Cα atoms within 7 Å of a given Cα atom. The radius of gyration ($R_g$) was calculated in VMD. All plots were made using the matplotlib function of Python\cite{24}.

2.4 Results and discussions

2.4.1 Site-Specific Interactions Prevalent in State I

By measuring the average distances between residues, we can identify specific interactions that occur in pHLIP in state I (Figure 2.4). When pHLIP is completely deprotonated (wild-type or WT), a small portion of the N-terminus (residues 6–11) preferentially interacts with the majority of the C-terminal half of the peptide (Figure 1A).

Each individually protonated residue has localized effects, decreasing the interactions between the N-terminal segment and the C-terminal half of the peptide. However, combinations of protonations are somewhat different. When titrating both interior aspartic acid residues, there is a marked increase in interactions (Figure 2.4, D14/D25); when protonating all aspartic acid residues, the interaction between the N-terminus and the C-terminal half of pHLIP is almost completely abolished (Figure 2.4, All). The first point mutation (P20G) removes a conformational restriction from
Residue–residue interactions of Ca atoms in pHLIP. First two columns show different protonation states of pHLIP. Regardless of the protonation state, increased interactions exist between the C-terminal region (approximately residues 26–33) and the N-terminal region (approximately residues 6–11) of pHLIP. Titration of acidic residues has differing results on these interactions. WT: fully deprotonated pHLIP; D14, D25, D14/D25, D31, D33: protonated residues in pHLIP; and All: all aspartic residues protonated. The rightmost column shows the helix-forming mutant (P20G) and non-natural amino acids that improve pHLIP function\[32\]. P20G mutation leads to a sharp increase of the N-terminus interacting with most of the peptide. The non-natural amino acids individually lead to greater interactions, but when both are incorporated, these interactions are mainly lost. Aad: α-aminoadipic acid; Gla: γ-carboxyglutamic acid.
the peptide backbone (proline introduces a helical kink in the folded conformation of pHLIP) and should allow residues proximal to position 20 to interact with each other. This is the case, as the N-terminus forms a continual interaction (i.e., <10 Å) with almost all of the C-terminal residues (Figure 2.4). We hypothesized that substitution of non-natural amino acids into pHLIP would lead to increased interactions within the peptide because of the increased negative charge (γ-carboxyglutamic acid (Gla)) or extended side chain (α-amino adipic acid (Aad)) of each residue. This is only partially the case; substituting Gla at position 14 abolishes interactions with proximal residues while simultaneously increasing interactions with the C-terminal end of pHLIP (residues 24–30). A slight increase in interactions occurs proximal to the Aad substitution at position 25; however, when combining substitutions D14Gla and D25Aad, the only area of increased interactions is with residues between the two substitutions.

2.4.2 Perturbing the C-Terminal Half of pHLIP Leads to Noticeable Increases in Helicity

We next examined the effect that these variations had on the ability of pHLIP to form a helix in solution. Although pHLIP does not form a helix in state I, even at acidic pH, slight variations in the composition of pHLIP are able to form helices at alkaline pH[27, 22]. Examination of the different protonation states shows that in almost every case protonation of a single acidic residue in pHLIP (D14, D25, D31, and D33) leads to an increase in helical conformations that are sampled in state I. This effect is not localized; most often, an increase in helicity occurs in the hydrophobic stretch of leucines between P20 and D25 (Figure 2.5). In addition, multiple protonations are not necessarily cooperative because the protonation of both D14 and D25 leads to a decrease in helicity compared to that of the completely deprotonated pHLIP. (The highest residue–residue interactions also occurred with D14 and D25 both protonated. It appears that there is no direct relationship between interactions and helix-forming propensity.) It is only when all aspartic acids are protonated that a significant increase in helicity occurs through a majority of the peptide. Point mutations in pHLIP do not have as significant an effect on the increase in helicity as that of titrations of
Chapter 2. Protonation Enhances inherent Helix-Forming Propensity of pHLIP

Figure 2.5: Increased interactions do not correlate with the helicity of pHLIP.

Probability of pHLIP to form a helical segment (as defined by helical φ-ψ angles for a three-residue sequence) as a function of residue titration. First column show different protonation states of pHLIP. Single protonations in pHLIP increase helicity in state I. It is only when all aspartic acids are protonated that a significant increase in helicity occurs through a majority of the peptide. Probability of pHLIP to form a helical segment as a function of point mutations. The second column shows point-mutations. P20G leads to localized increase in helicity, due to the removal of proline kink. Non-natural amino acids enhance helicity only at position 25 (D25Aad, D14Gla/D25Aad). Thick black line: helix-forming propensity for the fully deprotonated, wild-type pHLIP.

acidic residues (Figure 2.5). P20G leads to an increase in helicity in the interior of the peptide, consistent with observations from CD spectra[3]. D25Aad has the most noticeable increase in helicity near the hydrophobic stretch, possibly due to the charged side chain moving farther away from the peptide backbone. However,
combining both point mutations (D14Gla and D25Aad) leads to a much smaller increase in helicity compared to that of the wild-type pHLIP.

2.4.3 pHLIP Transiently Samples Both Major Secondary Conformations in State I

It is only when we consider the entire secondary structure conformational landscape that the true behavior of pHLIP in state I emerges. When none of the acidic residues in pHLIP are titrated (WT), the peptide samples α-helical and β-sheet conformations almost equally (Figures 2.6 and 2.7). Titration of the interior acidic residue, D14, leads to an increase in sampling of β-strand-like backbone conformations while maintaining the same level of sampling of helical conformations (Figure 2.6). This trend is most striking when both D14 and D25 are protonated (D14/D25); in this titration state, pHLIP is twice as likely to sample a β-strand than an α-helix (Figures 2.6 and 2.7). The other singly protonated states, corresponding to the C-terminal aspartate residues (D25, D31, and D33), lead to an increase in sampling of an α-helix compared to that of a β-sheet (Figure 2.6). Finally, when all aspartic residues are titrated (All), pHLIP is twice as likely to sample an α-helix than a β-sheet (Figure 2.6).

Variable effects on secondary structure formation are observed with respect to point mutations in pHLIP. Even though P20G has a localized increase in helicity, the overall helicity is lower than that of WT pHLIP (Figures 2.6 and 2.7). Combined with an increased sampling of β-strand conformations, P20G is more likely to sample strands than helices. Increased sampling of β-strands also occurs with the D14Gla mutant. However, both cases where the non-natural amino acid was introduced at position 25 lead to an increase in helicity and the propensity to sample helical space versus sheet conformations (Figures 2.6 and 2.7).

Regardless of the variables tested, pHLIP is capable of sampling conformational space corresponding to β-strands, in over a third of the cases more often than α-helices. There is precedent for these phenomena experimentally: although it was originally suggested that pHLIP adopts a random coil in state I[19], the corresponding CD spectrum has a negative mean residue ellipticity from 210 to 230 nm. This would indicate that a mixture of secondary structures (helices and sheets) exists[38] and is in
agreement with the theory that single-spanning transmembrane helices are in a coil–
helix equilibrium in solution\cite{42}. Sheets and helices are often key secondary structural
components in protein aggregation associated with neurodegenerative diseases\cite{37, 1},
and a recent study showed that amyloid fibrils can form from $\alpha$-helical transmembrane
proteins\cite{40}. The fact that pHLIP aggregates at $>10$ $\mu$M into a tetrameric unit with
exciton CD spectra representative of $\beta$-sheet formation\cite{34} is similar in nature to the
aggregation of peptides associated with amyloid formation.

![Figure 2.6: pHLIP samples multiple secondary structures in solution.](image)

\begin{enumerate}
  \item[(A)] Distributions of the $\psi$ backbone dihedral angle for different protonation states of
  pHLIP. The fully deprotonated state (WT, black line) samples the helical region of phase
  space (-50 to 0$^\circ$) more often than the $\beta$-sheet region of phase space (120 to 170$^\circ$). D31: blue
  line; D33: blue dash; D14: black dash; D25: red line; D14/D25: red dash; All: green line.
  \item[(B)] Distributions of the $\psi$ backbone dihedral angle for mutants of pHLIP. All mutants favor
  helical over sheetlike conformations. Point mutations at position 25 (D25Aad: blue line;
  D14Gla/D25Aad: green line) increase the likelihood of pHLIP to sample helical instead of
  sheetlike conformations. WT: black line; P20G: black dash; D14Gla: red line.
  \item[(C)] Ratio of $\alpha$-helical to $\beta$-strand sampling, as determined by the areas under the respective curves in (A)
  and (B).
\end{enumerate}
Chapter 2. Protonation Enhances inherent Helix-Forming Propensity of pHLIP

**Figure 2.7:** pHLIP samples multiple secondary structures in solution.

(A) Titration of interior acidic residues in pHLIP leads to an increase in sampling of $\beta$-strand-like backbone conformations. This is most striking when both D14 and D25 are protonated (D14/D25), leading to increased populations of $\beta$-strand conformations. An overlay of the allowed regions in the Ramachandran plot is shown for the deprotonated wild-type pHLIP (WT). (B) Point mutations can also favor $\beta$-sheet-like conformations. Interestingly, the P20G mutation, which leads to increased helical formation of pHLIP in state II[3], has noticeable sampling of $\beta$-sheet conformations.
2.4.4 Low-Energy States in Solution Are Favored by C-Terminal Modifications to pHLIP

Finally, we characterized the energy landscape of pHLIP as a function of the radius of gyration and intramolecular interactions (i.e., contacts). Generally speaking, the most energetically favorable states of pHLIP are when it is the most compact. Nearly all of the variants tested sample their lowest-energy conformations when the radius of gyration is 12 Å or less (Table 2.1). The one noticeable exception is when D14 and D25 are protonated (Figure 2.9). It appears that neutralizing the two interior aspartic acid residues leads to less favorable intramolecular interactions, as this configuration also has the lowest number of contacts. To compensate for this decrease in interactions, pHLIP will rearrange into a β-sheet conformation. In contrast, when all of the aspartic acids are protonated, pHLIP predominantly lies in a very narrow distribution of the radii of gyration (<10 Å). A common fold for the lowest-energy population had a kinked α-helix from residues 21 to 31 as well as a helix from residues 8 to 13.

Each of the point mutations studied has a slightly different effect on the conformation of pHLIP. D25Aad is the only mutation that results in a slightly smaller radius of gyration than that of WT pHLIP (Figure 2.9). Although P20G introduces greater helicity at position 20, this local ordering is offset by poor folding in the rest of the peptide, manifesting in a much broader distribution of contacts and radii of gyration. Incorporation of the bulkier and more negatively charged Gla side chain at position 14 led to a broader distribution of the radii of gyration. Examination of the simulated structures shows that the D14Gla variant has two predominant populations: (1) the Gla side chain is within the vicinity of the positively charged R11 side chain, and (2) the Gla side chain is >15 Å from R11 (Figure 2.8). The D25Aad variant, which has a longer side chain, has no interactions with the R11 side chain. However, when incorporating both non-natural amino acids (D14Gla and D25Aad), the D25Aad side chain forms a salt bridge with the side chain of R11 for nearly half of the simulation. This drastic shift to favor salt bridge formation does not necessarily result in more compact conformations of pHLIP (Figure 2.9); rather, the energetic gain from the salt bridge likely offsets the less energetic conformations with fewer contacts and larger radii of gyration. The behavior of pHLIP with respect to compactness is also
Figure 2.8: D25Aad sidechain forms salt bridge with R11 of pHLIP only in presence of D14Gla mutation.

(A) The negatively-charged sidechain at position 14 is conformationally restricted from forming a salt bridge with the positively-charged sidechain of R11. Even when an extra negative charge is introduced by the Gla mutation (-2 instead of -1), the distribution of distances from R11 remains > 15 Å. Black: distances for wt-pHLIP; red: corresponding distance for D14Gla; blue: corresponding distance for D14Gla/D25Aad. (B) The negatively-charged sidechain at position 25 in pHLIP does not form a salt bridge with the positively-charged sidechain of R11, except in the D14Gla/D25Aad double mutant. The major distribution of R11-D25Aad interactions for the double-mutant are for a salt bridge. Black: distances for wt-pHLIP; red: corresponding distance for D25Aad; blue: corresponding distance for D14Gla/D25Aad.
Figure 2.9: Number of contacts is loosely correlated with the radius of gyration in pHLIP.

Singly protonated residues in pHLIP (D14, D31, and D33) in general do not have a significant effect on the compactness of pHLIP (first two columns). The lone exception is when D25 is protonated (D25), which leads to a decrease in the radius of gyration. Protonation of both interior aspartates leads to a significantly less compact conformation (D14/D25). When all aspartates are protonated, pHLIP is the most compact (All). Far left: representative snapshots of WT, D14/D25, and All protonation states of pHLIP. Sticks: deprotonated aspartic acid residues; spheres: protonated aspartic acids. Note the presence of helices for WT and All as well as the formation of β-sheet when D14 and D25 are protonated. Point mutations (rightmost column) have little effect on the compactness of pHLIP. The P20G mutation leads to a less compact conformation than that of the wild-type pHLIP (WT). Notably, the double mutant (D14Gla/D25Aad) samples a broader distribution of radii of gyration with an almost equal probability of number of contacts.

consistent with that of intrinsically disordered proteins. The two-dimensional free energy landscape of pHLIP can be related to recent work studying the effect of topology on the propensity of polyampholytes to aggregate in solution[36, 6, 23]. The Pappu and Ghosh groups developed a framework whereby the primary peptide sequence could be used in combination with the fraction of charged residues to relate to the compactness of a monomer in terms of the radius of gyration[36, 6]. More recently, Lin and Chan were able to experimentally show that the radius of gyration for a single peptide chain is correlated to the phase behavior of multiple peptide chains[23].
Although pHLIP is an anionic peptide with a highly asymmetric charge distribution, the protonation states that we have tested follow the same relationship between the radius of gyration and sequence charge decoration (i.e., the radius decreases with increasing protonation up to a fully deprotonated pHLIP) that was observed both theoretically and experimentally (Table 2.1).

<table>
<thead>
<tr>
<th>Variant</th>
<th>ΔG (kcal/mol)</th>
<th>Rg (Å)</th>
<th>Rg error (10⁻⁵)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt-pHLIP</td>
<td>13.57</td>
<td>220</td>
<td>9.78</td>
</tr>
<tr>
<td>D14</td>
<td>13.21</td>
<td>208</td>
<td>9.85</td>
</tr>
<tr>
<td>D25</td>
<td>13.39</td>
<td>220</td>
<td>10.05</td>
</tr>
<tr>
<td>D31</td>
<td>14.18</td>
<td>224</td>
<td>9.82</td>
</tr>
<tr>
<td>D33</td>
<td>13.64</td>
<td>214</td>
<td>9.90</td>
</tr>
<tr>
<td>D14/D25</td>
<td>15.12</td>
<td>212</td>
<td>10.46</td>
</tr>
<tr>
<td>All</td>
<td>12.73</td>
<td>222</td>
<td>9.47</td>
</tr>
<tr>
<td>P20G</td>
<td>14.66</td>
<td>230</td>
<td>9.58</td>
</tr>
<tr>
<td>D14Gla</td>
<td>13.23</td>
<td>210</td>
<td>9.94</td>
</tr>
<tr>
<td>D25Aad</td>
<td>13.60</td>
<td>224</td>
<td>9.67</td>
</tr>
<tr>
<td>D14Gla/D25Aad</td>
<td>14.77</td>
<td>214</td>
<td>10.18</td>
</tr>
</tbody>
</table>

Table 2.1: Free energy minima for each pHLIP variant tested, based on 2D free energy landscape (number of contacts vs. radius of gyration, Figure 2.9)

2.5 Conclusion

We have used MD simulations of pHLIP in implicit solvent to characterize the behavior of pHLIP in state I. pHLIP can sample multiple conformations in solution, as observed from previous CD studies. Significantly, we determined that pHLIP is capable of folding into both helices and strands when acidic residues are titrated. These conformational effects appear to be cooperative, as the most noticeable folding occurs when either the interior or all aspartic acid residues become protonated. This phenomenon is relevant to our continued improved understanding of pHLIP function, as it is necessary for determining optimal conditions for soluble administration of pHLIP as a diagnostic imaging or drug-delivery agent as well as its response to fluctuations in environmental pH while in solution.
2.6 Future directions

While it is informative to explore the protein folding events of pHLIP and pHLIP variants, the use of implicit solvent model means we miss out on some details like formation of water-shell, peptide-water hydrogen bond, etc (see section 2.2.1). Careful control of ionic strength is also not possible in implicit solvents. For these reasons, a good next step would be simulating pHLIP in explicit waters. As pointed out earlier, this would cause a significant increase in system size with a corresponding decrease in performance. For example, pHLIP solvated in a cubic box of TIP3P waters would have 70000 atoms (compared to 560 atoms of the implicit solvent simulation reported here). This hinders our possibility of achieving the relevant folding/unfolding timescales (70000 atom system has a performance of 25 ns/day using pmemd.cuda on NVIDIA GTX 780 GPU while the 560 atom system gave 650 ns/day). To overcome this difficulty, the following steps were taken:

1. Use of Gaussian accelerated Molecular Dynamics (GaMD) to accelerate the binding/unbinding process.
2. Use of hexagonal box instead of square box causes a slight reduction in the number of waters.
3. Following the protocol of Doshi et. al[8], we performed a short GaMD simulation which caused pHLIP to shrink, which could be solvated in an even smaller solvation box.
4. Additionally, we used OPC water[21] as TIP3P is known to bias secondary structure formation.

We chose to use GaMD for the following reasons:

1. Unlike many other biasing methods (like Umbrella Sampling), aMD methods do not require the selection of a reaction coordinate. For the problem at hand, this is a big advantage given we are studying an unstructured peptide in solution for which experimental data is scarce if at all.
2. Along with enhanced sampling, aMD simulations can be reweighted to calculate the free energy profile (potential of mean force, PMF) of the system under study.
3. GaMD provides more accurate reweighting compared to aMD (see section 2.6.1).
2.6.1 Theory of (Gaussian) accelerated MD

Accelerated molecular dynamics aims to accelerate transition between energy basins by reducing the energy barrier between them. First proposed by Hamelberg et. al[13], its idea is to add a boost potential $\Delta V$ to the potential to the potential energy surface when the true potential is below a certain cutoff. In the following equation, $V(r)$ is the original potential energy surface, $E$ is the user-defined cutoff, and $V^*(r)$ is the modified potential.

$$V^* (r) = \begin{cases} V(r), & \text{if } V(r) \geq E. \\ V(r) + \Delta V(r), & \text{if } V(r) < E. \end{cases}$$

Here, $\Delta V$ is called the 'boost potential', whose form is given by

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))}$$

where $\alpha$ is a tuning parameter (low $\alpha$ resembles higher acceleration).

While this paper focused on boosting the dihedral potential, later work from the group introduced boosting the overall potential[31] as well as performing “dual-boost” where the overall and dihedral potential can be boosted separately[14].

An important aspect of aMD method is recovering the unbiased free energy surface of the system from the biased (accelerated) simulation. This is done by reweighting[13, 26]. However, this step introduces statistical inaccuracies which increases with increasing acceleration (lower $\alpha$).

To overcome this problem, Miao et. al[25]. proposed Gaussian aMD (GaMD) where the boost potential takes a gaussian form:
\[ \Delta V(r) = \frac{1}{2} k (E - V(r))^2 \]

where \( k \) is the (user-defined) harmonic force constant. This was based on the observation that when the boost potential follows gaussian (or near-gaussian) distribution, improved reweighting could be achieved\[26\].

GaMD simulation requires its own “equilibration” (beyond the regular minimization and equilibration) whereby it sets the GaMD parameters. This is explained in the figure on GaMD’s manual page. The simulation begins with a conventional MD (cMD) period (“ntcmd”) of which the first few (“ntcmdprep”) steps are purely for system equilibration. During (ntcmd-ntcmdprep) period, initial GaMD parameters are collected, and GaMD equilibration begins (“nteb”). Once again, the first few (“ntebprep”) steps are for equilibration, beyond which GaMD parameters are updated every “ntave” steps.

### 2.6.2 Methodology

The same initial structure was placed in an octahedral box of OPC waters, neutralized by adding 5 Na+ ions, and minimized by steepest descent restraining the protein atoms with a force constant of 200 cal/molÅ². System was then equilibrated in for 50 ps where restrain was 100 kcal/molÅ² and then for 100 ps with restrain of 25 kcal/molÅ². This was followed by 100 ps equilibration without restraint.

The system was prepared following a procedure similar to Doshi et. al\[8\], except that we used GaMD instead of RaMD. The idea is to start with a large water box to prevent pHLIP from interacting with its periodic neighbor (this will be termed ’Part I’ in this text), run a short a GaMD simulation whereby pHLIP shrinks significantly (see figure 2.10). This structure was then resolvated in a smaller water box (termed ’Part II’ here), allowing for faster simulation. This new system is again subjected to GaMD. GaMD equilibration of parts I and II were performed separately. All simulations were performed at 300 K using dual-boost GaMD. For Part I, 2 ns preparation cMD
pHLIP shrinks during the 'Part I' GaMD simulation, as reflected by the reduction of its end-to-end distance. This is because pHLIP has a propensity to fold in solution: something we already know from the implicit solvent simulations and are exploring further here. This works to our advantage as it allows for a smaller water box (with fewer waters) to solvate pHLIP completely, after it has shrunk. A smaller system with fewer atoms can be simulated faster. The original system had 42261 atoms (10424 waters) which gave ~30 ns/day performance on NVIDIA GTX 980 GPU. After 100 ns GaMD, pHLIP had shrunk to almost ~66% of its original end-to-end distance. This molecule, after solvation, had 23409 atoms (5711 atoms), and gave ~55 ns/day performance on the same GPU.

(“ntcmd”) was used where first 0.4 ns was equilibration (“ntcmdprep”). Then, 20 ns (“nteb”) GaMD was performed and the potential statistics was updated every 0.4 ns (“ntave”) after the first 1.6 ns (“ntebp”). $\sigma_{0P}$ and $\sigma_{0D}$ were both set to 6.0 kCal/mol (default). GaMD production was performed for 100 ns, by which pHLIP end-to-end distance had decreased by ~30%. This structure was resolvated to create the system for Part II. For Part II, the optimal value of sigma0P and sigma0D was optimized so as to obtain the maximum acceleration with minimum boost which ensures accurate reweighting. To do this, $\sigma_0$ values of 6.0, 3.0 and 1.0 kCal/mol ($\sigma_{0P}$...
In GaMD simulations, there is a trade-off between the acceleration and the accuracy of reweighting. Higher boost (denoted by higher $\sigma_0$) results in more acceleration (denoted by higher $k_0$), but reduces the reweighting accuracy. Three parallel GaMD equilibration was performed with the ‘Part II’ system where everything was same except $\sigma_0$ ($\sigma_{0P}$ was always kept equal to $\sigma_{0D}$). While $\sigma_{0P} = 1.0$ kCal/mol (blue) did not provide any acceleration ($k_0 \sim 0$), $\sigma_{0P} = 3.0$ kCal/mol (red) and $\sigma_{0P} = 5.0$ kCal/mol (black) eventually achieved maximum possible acceleration ($k_0 = 1$). Similar behavior was also obtained for $\sigma_{0D}$ (not shown). Thus, $\sigma_{0P} = \sigma_{0D} = 3.0$ kCal/mol provided the best trade-off, and was used in the rest of the simulation.

and $\sigma_{0D}$ was kept equal in all simulations) were tried. Figure 2.11 shows the acceleration achieved in the overall potential boost for these three simulations. The value $k_0$ varies between 0 and 1 where $k_0 = 0$ implies no acceleration and $k_0 = 1$ implies the maximum possible acceleration. $\sigma_{0P} = 3.0$ kCal/mol provided as good acceleration as $\sigma_{0P} = 6.0$ kCal/mol. Similar results were obtained for the dihedral boost (data not shown). Thus, $\sigma_{0P} = \sigma_{0D} = 3.0$ kCal/mol was the optimal choice and was used to extend the simulation to 830 ns.
2.6.3 Preliminary results

Reweighting and Gaussian behavior

As mentioned before, reweighting of aMD simulations introduces statistical inaccuracies, which increases with increasing acceleration. There are more than one ways of approximating the reweighting factor, which has been summarized by Miao et. al [26]. The study noted that accurate reweighting was possible when the distribution of the boost (ΔV) was gaussian or near-gaussian. In such cases, the reweighting factor, \( <e^{βΔV}> \) could be approximated by cumulant expansion to the second order:

\[
<e^{βΔV}> = exp\left(\sum_{k=1}^{∞} \frac{β^k}{k!}C_k\right)
\]

\[
≈ exp\left(\sum_{k=1}^{2} \frac{β^k}{k!}C_k\right)
\]

where the cumulant terms \( C_k \) are given by

\[
C_1 = <ΔV>
\]

\[
C_2 = <ΔV^2> - <ΔV>^2 = σ_{ΔV}^2
\]

The study also introduced a quantity “anharmonicity”, denoted by \( γ \), to quantify the degree of deviation from gaussian behavior. As a first check on our simulation, we looked at the distribution of ΔV and its anharmonicity. Figure 2.12 shows the distribution of the total boost potentials (sum of dihedral and potential boost of each
Chapter 2. Protonation Enhances inherent Helix-Forming Propensity of pHLIP

Figure 2.12: Distribution of boost potential in 830 ns dual-boost GaMD simulation.

Accurate reweighting of GaMD simulation is possible when the distribution of boost potential ∆V follows gaussian or near-gaussian distribution. Deviation from gaussian behavior is quantified by anharmonicity (γ). In our 830 ns dual-boost GaMD simulation of pHLIP in octahedral OPC water box using σ₀P = σ₀D = 3.0 kCal/mol, ∆V distribution was near-gaussian, with an anharmonicity of 0.0286.

Frame) for a 830 ns GaMD simulation of pHLIP in octahedral OPC water box. The distribution is near-gaussian, justifying the use of the cumulant expansion to the second order for reweighting. The anharmonicity, shown in the top-right corner, is 0.0286. This is of the same order of magnitude (0.00922) reported by Miao et. al[25] for Chignolin. Note that pHLIP is significantly longer (35 residues) than chignolin (10 residues) and is unstructured.

PMF of the φ-ψ space

In order to calculate the dihedral PMF, we first optimized the bin width for the reweighting. While a smaller bin width provides better resolution of data, it could lead to some bins with too few data points, resulting in an increase in anharmonicity
Chapter 2. Protonation Enhances inherent Helix-Forming Propensity of pHLIP

Figure 2.13: Free energy profile and bin-wise anharmonicities of $\phi$ and $\psi$ dihedral angles

Free energy profiles of backbone dihedrals $\phi$ and $\psi$ of all residues of pHLIP, calculated by cumulant expansion to the second order, from 830 ns dual-boost GaMD simulation. (A) PMF of $\phi$ dihedral angle. The region near 0 degrees is poorly sampled, giving rise to high uncertainties in the PMF calculated from bin width of 3 degrees. PMF from bin width of 6 and 9 degrees have much lower uncertainty. (B) PMF of $\psi$ dihedral angles. In general, uncertainties are much less than $\phi$ PMF. PMF with bin width of 6 and 9 degrees are similar. (C) Bin-wise anharmonicity of $\phi$ and (D) $\psi$ dihedral angles. Anharmonicity consistently reduces when increasing bin width from 3 to 6 to 9. Anharmonicity of $\psi$ is, in general, lower than that of $\phi$.
Chapter 2. Protonation Enhances inherent Helix-Forming Propensity of pHLIP

Figure 2.14: Two dimensional free energy profile and anharmonicity of backbone dihedral angles

(A) Two dimensional PMF of the backbone dihedral ($\phi$-$\psi$ space). Regions near $\alpha$-helix and $\beta$-sheet show regions of low free energy, as expected from the implicit solvent simulations reported earlier. Additionally, we see a shallow potential well near the left-handed helix region that was not observed earlier. (B) Bin-wise anharmonicity of the $\phi$-$\psi$ space. Regions pertaining to $\alpha$-helix, $\beta$-sheet as well as left-handed helix are well-sampled, resulting in lower anharmonicities.

lower uncertainty. The uncertainty of PMF is similar between bin size of 6 and 9 degrees. Similar behavior was observed by Miao et. al[26], wherein increasing bin size monotonically decreased anharmonicity while PMF uncertainty reached a plateau. The $\psi$ dihedral angle showed much lower anharmonicities and PMF uncertainties (Figure 2.13 B-D). For both $\phi$ and $\psi$, bin size of 6 was determined as optimal.

We then looked at the 2D free energy landscape of the Ramachandran space of pHLIP in solution (Figure 2.14). Similar to the implicit solvent simulations, we see two potential wells, one near the $\alpha$-helix region, and the other near the beta sheet region. We also observe a new, unexpected potential well near the left-handed helix region. Thus, 830 ns GaMD simulations allowed us to sample regions of conformational space that was inaccessible in multiple 2 $\mu$s implicit solvent simulations.
PMF of the R$_g$-contact space

![Figures 2.15: Free energy profile and bin-wise anharmonicities of radius of gyration and number of contacts](image)

Free energy profiles of radius of gyration and number of contacts, calculated by cumulant expansion to the second order, from 830 ns dual-boost GaMD simulation. (A) PMF of radius of gyration. Uncertainties reduce when increasing bin width from 0.1 to 0.3, then stabilize at 0.6. (B) PMF of number of contacts. PMF is slightly rugged with bin size of 2, then smoothen out at bin size of 4, and is unchanged at bin size of 6. (C) Bin-wise anharmonicity of radius of gyration. Anharmonicities reduce significantly when increasing bin size form 0.1 to 0.3 and essentially flatten out at bin size of 0.6. (D) Bin-wise anharmonicity of contacts. Anharmonicities are low in general. High anharmonicity near the ends reduce significantly when increasing the bin size from 2 to 4, and is unchanged at 6.

Finally, we looked at the radius of gyration (R$_g$) and number of contacts, similar to section 2.4.4. Again, we started with optimizing the bin-width for reweighting. Figure 2.15 shows the PMF and bin-wise anharmonicity of the R$_g$ contacts. The radius of gyration (Figure 2.15 A-C) PMF is rugged at bin size of 0.1 but smoothens out at bin size of 0.3, while anharmonicities continue to reduce when increasing bin size further to 0.6. Contacts only take even values and hence bin only even-numbered bin sizes were used (Figure 2.15 B-D). The PMF is slightly rugged with bin size of
2 but becomes smooth with bin size of 4, and is unchanged when increasing bin size to 6. The anharmonicities are generally low, except the extremeties (Contacts < 140 and contacts > 200). However, when increasing bin size to 4, there is a huge reduction in these anharmonicities. Thus, the optimal bin size for the \((R_g, \text{contact})\) space was determined to be \((0.6,4)\).

Finally, we looked at the 2D free energy landscape of radius of gyration and number of contacts \ref{fig:2d_free_energy}). While the minima of radius of gyration (around 12 ) corresponds to what we observed in implicit solvent simulations (Figure \ref{fig:pmf khoản} A), the minima of the number of contacts has shifted to lower values, of \(\approx 180\). While we are currently unaware of the reason, one possibility could be that the implicit solvent model we used\cite{29, 30} promoted “folded-over” conformations of pHLIP. The OPC water model\cite{21} used in this work might also have an influence: in future we might try to compare GaMD simulations involving OPC and more traditional TIP3P water model.

![Figure 2.16](image_url)

**Figure 2.16:** Two dimensional free energy profile and anharmonicity of contact and radius of gyration

(A) Two dimensional PMF of radius of gyration and number of contacts. Region near \(\approx 180\) contacts and \(R_g\) of \(\approx 12\) shows a potential minima. (B) Bin-wise anharmonicity of the \(R_g\)-contact space. Most of the region is well-sampled resulting in low anharmonicities.
2.6.4 Final remarks

In this section, we have shown a possible way forward for studying state I of pHLIP using MD simulation. While implicit solvent models allow achieving long (µs) timescales, such approach has its inherent shortcomings as has been listed earlier. Use of explicit waters, on the other hand, causes a significant increase of system size and a corresponding reduction in simulation speed. To avoid the insufficient sampling that might arise out of that, we explored the possibility of using Gaussian Accelerated Molecular Dynamics (GaMD)[25] in this section. This technique not only speeds up conformational sampling, but also enables us to calculate the free energy landscape of our system. A big advantage of accelerated MD simulation is that it does not necessitate a good knowledge of the system.

Using a explicit solvent water box, it will be straightforward to accurately control the ionic strength of the solution. The current study used 0 mM salt concentration (plus 5 neutralizing Na⁺ ions). As we show in chapter 4, lipid binding of pHLIP is influenced by the ionic strength of the solution. This raises the question whether conformation of pHLIP itself is affected by the presence of ions. Such a study is not possible in implicit solvents as the latter is modeled under infinite dilution condition. The method described in this section might be useful for the systematic study of the influence of ionic strength on the conformation of pHLIP in state I.

Another potential use might be studying the aggregation of pHLIP. pHLIP has a propensity to aggregate[20], even more so at higher ionic strength[11]. It would be of significant interest to be able to study the mechanism of pHLIP aggregation. Recently, Narayanan et. al[28]. has used small angle xray scattering (SAXS) spectra to indicate pHLIP aggregates might be a tetramer. Simulating multiple monomers in implicit solvent models, while not impossible, is rare. This would become much simpler by using periodic box of explicit waters. GaMD might be the technique of choice for studying such systems.
References


REFERENCES


REFERENCES


[40] Karen Stroobants et al. “Amyloid-like Fibrils from an α-Helical Transmembrane Protein”. In: Biochemistry 56.25 (June 2017), pp. 3225–3233. ISSN: 0006-2960. DOI: 10.1021/acs.biochem.7b00157. URL: https://doi.org/10.1021/acs.biochem.7b00157.


Chapter 3

Cooperative Non-bonded Forces
and Membrane Binding of
pHLIP

3.1 Introduction

Oncology is difficult from the perspective of both diagnosis and treatment: early
detection is often impossible to achieve (e.g., pancreatic cancer)[45] and conven-
tional treatments suffer from off-target side effects in which chemotherapeutics attack
healthy as well as cancerous cells[28]. New approaches to treatment aim to increase
the effectiveness of targeting, as evidenced by monoclonal antibodies[55] and antibody
drug conjugates (ADCs)[34], which distinguish between cancerous and healthy cells
by binding cancer-specific antigens. However, this approach has potential shortcom-
ings, due to heterogeneity of tumor cells and rapid mutations of targeted antigens,
leading to resistance. An alternative targeting method is to utilize unique charac-
teristics of the tumor microenvironment for exploitation (e.g., hypoxia or acidity in
the exterior of the cell). The extracellular environment of tumor cells (pH 6.5-7.0)
is more acidic than normal tissue (pH 7.2-7.5)[13]. Tumor acidosis stems from al-
terred and enhanced metabolism within the cancer cell. This is manifested through
the tendency of cancer cells to utilize glycolysis for ATP synthesis (i.e., the Warburg
effect[50]), leading to an excess of lactic acid that is pumped outside the cell. One
facet of acidosis that makes it an appealing targeting factor is the fact that it is a
universal feature of all cancer cells. Developing a therapeutic approach targeting this
property would potentially have far-reaching effects in oncology.

One approach is the pH (Low) insertion peptide (pHLIP)[18, 19]. pHLIP resides in three distinct states that are dependent upon the surrounding environment: unstructured and soluble in water at alkaline pH (state I); unstructured and bound to the cell membrane surface at alkaline pH (state II); and inserted across the membrane as an α-helix at acidic pH (state III) (Figure 3.1). Peptide insertion is induced by a shift from alkaline to acidic pH, with an apparent pK of insertion in 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membranes of ~6.1[38]. This shift in pH spurs the protonation of two acidic residues (D14 and D25) in the membrane-spanning region of the peptide, leading to an increase in hydrophobicity and triggering the folding and insertion of the peptide across the lipid bilayer[38, 1]. The presence of two tryptophan residues (W9 and W15) allow intrinsic fluorescence spectroscopy to be used to detect the transition of pHLIP from state I → II → III.

Our understanding of the state I → state II transition – binding of pHLIP to the membrane surface – has evolved since pHLIP was first discovered, due to advances in the techniques used to characterize pHLIP as well as the lipid systems being studied. Isothermal titration calorimetry experiments on pHLIP and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) vesicles determined that the most important thermodynamic step in pHLIP function is binding to the membrane surface (ΔGbind = −7.2 kcal-mol⁻¹, whereas ΔGinsert = −1.8 kcal-mol⁻¹)[38]. This binding is accompanied by burial of W9 and W15 due to desolvation of pHLIP, a necessary prerequisite for the state II → state III transition[56, 6]. Solid-state NMR studies have also shown that 1) residue-specific differences in pHLIP-membrane interactions occur at each of three pH values (pH 7.4, 6.4, and 5.3)[43] and 2) each aspartic acid residue in pHLIP has a distinct pKₐ, with the C-terminal residues being titrated first[16]. When adding anionic phospholipids (i.e., phosphatidylserine) to the vesicles, the free energy of binding is essentially the same[42]. However, Ladokhin and coworkers determined that state II is spectroscopically silent in the presence of non-PC lipids – pHLIP will bind to lipid vesicles but does not partition below the headgroup region, leading to a lack of change in tryptophan fluorescence from state I to state II[30, 47]. These more recent studies strongly suggest that a more complex mechanism of
Figure 3.1: **Systems investigated in this study.**

(A) Primary sequence of wt-pHLIP and pHLIP-1. **Boldface:** acidic residues in the sequence; **underline:** putative region that undergoes folding into a transmembrane helix. (B) Schematic of how pHLIP binds to a plasma membrane surface. As pHLIP diffuses in solution, it can take on many conformations (state I, left). pHLIP will then encounter the membrane surface and partition into the headgroup region (state II, right). **Ribbons:** pHLIP; **spheres:** lipids in a POPC bilayer.

binding exists for pHLIP. Despite these efforts to characterize state II, knowledge of the detailed interactions between pHLIP and the cell membrane that eventually lead to folding and insertion are sorely lacking. Without this knowledge, it will be extremely difficult to intelligently design pHLIP variants with the ability to modulate apparent $pK'$s of insertion to closely match the pH of acidic microenvironments such as cancer cells.

Molecular dynamics (MD) simulations are capable of providing atomistic details on the biophysical interactions between pHLIP and a lipid bilayer that are unattainable through experimental methods. In this study we have carried out equilibrium MD simulations to characterize the biophysical interactions that govern the behavior of pHLIP (wt-pHLIP) and a fast-folding variant of pHLIP (pHLIP-1) in state II. Our goal was to address the following unresolved issues with respect to state II: 1) Is there a favored binding complex? 2) What is the mechanism by which this binding occurs? 3) Does partitioning-folding coupling occur during binding? 4) If so, is folding
conserved in a specific region of pHLIP? We chose to study wt-pHLIP and pHLIP-1[24] in order to compare the canonical version of pHLIP with a variant that possesses helical character in state II. Our simulations capture the non-bonded interactions that lead to binding of both peptides, showing good agreement with recent experiments and identifying common motifs in the binding process among different orientations of pHLIP with respect to the bilayer surface. We observe stable formation of α-helical folds in pHLIP-1 centered around the D14 and D25 protonation switches; surprisingly, wt-pHLIP also shows an ability to adopt stable helical turns. Most importantly, a clear difference exists in the effectiveness of binding of wt-pHLIP and pHLIP-1 to POPC, indicating that peptide orientation has a direct effect on promoting formation of pHLIP-bilayer complexes.

3.2 Materials and Methods

Previously we carried out MD simulations of wt-pHLIP in implicit solvent to determine the most likely conformation of pHLIP in state I[15]. We selected a representative snapshot from our most populated cluster in our K-means clustering analysis (i.e., with the lowest root-mean squared deviation to the average structure in the cluster) as the starting structure of wt-pHLIP for our state II simulations. pHLIP-1 (based on the sequence from Karabadjak et al.[24]) was modified from helix C of the crystal structure of bacteriorhodopsin (PDB ID: 2NTU). To drive pHLIP-1 into a coiled conformation, an MD simulation in the NVT ensemble (T = 700K) was conducted in vacuo for 2 ns using a 2.0 fs timestep using NAMD 2.9[36]. This structure was then solvated in implicit solvent as in[15] and simulated for 2 μs. K-means clustering was then used to select the most probable conformation to form the peptide-bilayer complex.

A POPC bilayer of 200 lipids was prepared using the membrane builder in CHARMM-GUI[22, 5, 32, 53, 23, 21] and equilibrated in the NPT ensemble (T = 310 K, P = 1 atm) for 50 ns using NAMD 2.10 and the c36 lipid force field[27]. A 2.0 fs timestep, with a force-based switching function for Lennard-Jones interactions from 10 to 12 Å, Langevin thermostat, Nosé-Hoover barostat, and a flexible cell with constant ratio
were used. wt-pHLIP and pHLIP-1 structures were then merged with the equilibrated POPC bilayer to generate five independent peptide-bilayer complexes for each type of peptide, referred to as 0°, 72°, 144°, 216° and 288°, with degrees measured as the rotation of each respective peptide around their first principal component axis. Note that these angles are arbitrary and do not correlate between wt-pHLIP and pHLIP-1 simulations. Overlapping waters and lipids were removed. The final systems had a lipid:peptide ratio ~200:1 (exact ratio varied depending on how many lipids had to be eliminated for each system), with a slight degree of bilayer asymmetry.

Unbiased molecular dynamics simulations in the tensionless NPT ensemble were performed with the GPU-accelerated version of pmemd in Amber 16\cite{7, 40} using the CHARMM 36 force field including a modification to better account for cation-pi interactions\cite{26, 27, 3}. A hard cutoff of 8 Å was applied to non-bonded forces, as recommended by the CHARMM community\cite{32}. The particle mesh Ewald (PME) approach was used for computing electrostatic forces\cite{8, 12}, and hydrogens were restrained with the SHAKE algorithm\cite{39}. Simulations were run with a 2.0 fs timestep at 310 K with a Langevin thermostat. The Berendsen barostat was used to maintain pressure at 1 atm with semi-isotropic pressure scaling. Visualization was performed using VMD\cite{17}, data analysis was carried out using VMD and in-house python scripts, and matplotlib and gnuplot were used for making plots.

3.3 Results

3.3.1 Partitioning of pHLIP influenced by location of acidic residues

The recent NMR work of An and Qiang examined partitioning of pHLIP into POPC vesicles during the state II → state III transition over incremental pH jumps (pH 7.4 → 6.4 → 5.3) using rotational-echo double-resonance (REDOR) NMR spectroscopy\cite{43}. The C-terminal residues that were isotopically labeled (L21, L22, L26, and A27) displayed a high degree of dynamics, leading to their hypothesis that the polar C-terminus of pHLIP (D31, D33, and E34) plays a major role in preventing partitioning of that end of pHLIP into the bilayer surface. A subsequent follow-up
Chapter 3. Cooperative Non-bonded Forces and Membrane Binding of pHLIP

Figure 3.2: Quality of binding of pHLIP can depend on orientation and peptide composition.

(A) Per-residue distance of wt-pHLIP from the P atoms of the upper leaflet as a function of time. wt-pHLIP can bind to the surface of a POPC bilayer through electrostatic or nonpolar interactions. In the case of 0° and 72° orientations, the N-terminal portion of pHLIP binds stably (over hundreds of ns), mainly due to a combination of salt bridge formation with R11 and partitioning of aromatic sidechains (tyrosine and tryptophan). The 144° orientation is transiently bound to the bilayer surface and completely dissociates around 300 ns. In contrast, the 216° and 288° orientations remain stably bound to the bilayer surface, this time through partitioning of the nonpolar “sinker stretch” of the C-terminal half of the peptide.

(B) Per-residue distance of pHLIP-1 from the P atoms of the upper leaflet as a function of time. In general, the opposite behavior is observed for pHLIP-1. C-terminal residues predominantly bind to the POPC surface for extended periods of time (hundreds of ns). For the most part, binding of pHLIP-1 is independent of orientation; see Figure 3.3 for details on the distribution of binding.

study from An and Qiang showed that the acidic residues in the C-terminus of pHLIP not only affect the ability for pHLIP to partition into the bilayer surface, but they are also the first residues to be titrated upon acidification of the surrounding environment[16]. What was unclear was the specific relationship between binding of pHLIP to the bilayer surface and titration of acidic residues – is titration facilitated by a change in hydrophobicity of the surrounding environment through burial in the headgroup region, or does exposure to solvent, as most likely with the C-terminus of pHLIP, allow for faster transitions from the acidic to the neutral state? Our simulations on wt-pHLIP and pHLIP-1 consistently remain bound to the bilayer surface, but with significant distinctions between the two peptides and between orientations. For wt-pHLIP, the 0° and 72° orientations utilize a combination of salt bridge formation with R11 and partitioning of aromatic sidechains within the N-terminal segment of
the peptide to effectively bind to the bilayer surface (Figure 3.2 A). However, the 144° orientation completely dissociates from the bilayer around 300 ns, while the 216° and 288° orientations establish stable binding via the sinker stretch of nonpolar residues from position 21-30. In addition, the hydrophobic sinker stretch from T18 to L24 is able to remain embedded within the bilayer for all of the bound orientations of wt-pHLIP, indicating that it plays a significant role in the quick cooperative response in state II that was observed by An and Qiang when shifting from pH 7.4 to pH 6.4. Unlike wt-pHLIP, pHLIP-1 remains stably bound to the bilayer surface regardless of orientation. Binding is dominant from position 10 to the C-terminus and is a significant extension of the binding region of the sinker stretch (Figure 3.2 B). The most likely explanation for this stability in binding is due to the lack of acidic residues on the C-terminus of pHLIP-1.

Our results qualitatively agree with the aforementioned NMR studies; each respective half of the peptide with terminal polar residues (N-terminus for pHLIP-1, C-terminus for wt-pHLIP) lies furthest from the bilayer surface. If we group wt-pHLIP and pHLIP-1 by region (N-terminus: residues < 10; middle: residues 10-27; C-terminus: residues > 27), more general patterns begin to emerge. The middle segment in wt-pHLIP most consistently binds to POPC. In addition, the N- and C-terminal segments of wt-pHLIP have opposite binding affinities to POPC: the N-terminus, with only one charged residue, binds more effectively, while the C-terminus, with three acidic residues, binds poorly or not at all (Figure 3.3 A). Unlike wt-pHLIP, binding of pHLIP-1 is fairly consistent between all five orientations for both the N-terminus and middle segments — no binding of the N-terminus (higher concentration of acidic residues) and effective binding of the middle segment. For the C-terminal segment of pHLIP-1, binding also occurred in the majority of the simulations, with a larger distribution (Figure 3.3 B). Each orientation of the C-terminus has equivalent or improved binding effectiveness in pHLIP-1 compared to wt-pHLIP. The solid-state NMR study of An and coworkers determined the distances of isotopically-labeled alanine residues (A10, A13, and A27) in wt-pHLIP from the upper leaflet of POPC vesicles in order to characterize the quality of binding of these residues in state II. In most cases of our wt-pHLIP simulations, A10 and A13 do not fully reproduce these
Figure 3.3: Effective binding of pHLIP to POPC is related to specific segments of the peptide.

(A) Distance distribution of segments of wt-pHLIP with respect to the center of mass of the POPC bilayer. The middle segment in wt-pHLIP has the most consistent level of binding to POPC. In addition, the N- and C-terminal segments of wt-pHLIP have opposite binding affinities to POPC: the N-terminus, with only one charged residue, binds more effectively, while the C-terminus, with three acidic residues, binds poorly or not at all. This result is consistent with the solid-state NMR studies of Qiang, An, et al., who hypothesized that the negative residues on the C-terminus lead to charge-charge repulsion between pHLIP and the headgroup region of a POPC bilayer.\[43\] N-terminus: residues < 10, middle: residues 10 to 27, C-terminus: residues > 27. Solid black lines indicate the z-component of the center of mass of the phosphorous atoms in the leaflet in contact with wt-pHLIP. B) Distance distribution of segments of pHLIP-1 with respect to the center of mass of the POPC bilayer. Unlike wt-pHLIP, binding of pHLIP-1 is fairly consistent between all five orientations for both the N-terminal and middle segments — no binding of the N-terminus (higher concentration of acidic residues) and effective binding for the middle segment. For the C-terminal segment of pHLIP-1, binding also occurred in the majority of the simulations, with a larger distribution. Each orientation of the C-terminus binds as well or more effectively in pHLIP-1 than in wt-pHLIP.

Equilibrium distances (Figure 3.4 A). However, it is of note that the lipid to peptide ratio used in the NMR experiments was 75:1, whereas our simulations were conducted at a ratio of 200:1. Previous studies have shown that binding of pHLIP is affected by lower lipid:peptide ratios, leading to a “parking problem” for pHLIP on the bilayer surface\[38\]. For the orientations that remained bound to POPC, the $0^\circ$, $72^\circ$, and $288^\circ$ orientations all had distance distributions for A27 that were $>10$ Å in agreement
Figure 3.4: $^{31}$P-$^{13}$C distances reveal that orientation of the pHLIP binding complex has varying degrees of influence.

(A) Distribution of distances between the nearest P atom in the upper leaflet of the POPC bilayer and A10 (black), A13 (red), and A27 (blue) in wt-pHLIP (all orientations except 144° pooled together). A10 and A13 do not exactly reproduce the equilibrium distances determined from solid-state NMR experiments at pH 7.4\cite{43} (dashed lines). However, it is of note that the lipid to peptide ratio used in the NMR experiments was 75:1, whereas our simulations were conducted at a ratio of 200:1. Distributions for A27 is > 10 Å in agreement with the NMR experiments. Dashed lines represent $^{31}$P-$^{13}$C distances for A10 and A13 at pH 7.4 as reported in Shu et al.  

(B) Distribution of distances between the nearest P atom in the upper leaflet of the POPC bilayer and A10 (black), A13 (red), and A27 (blue) in pHLIP-1. In general, distances for pHLIP-1 are much closer than wt-pHLIP for all three alanine residues. This is in part because A27 is not proximal to the three C-terminal acidic residues present in wt-pHLIP, which has the potential for electrostatic repulsion with the negatively-charged phosphate region of the PC headgroups. In addition, the distances for A10 and A13 are slightly reversed; A13 is closer to the P atoms than A10. This reversal could be due in part to the presence of negatively-charged residues at positions 2, 3, and 6 in pHLIP-1, which could have a more marked effect on the closer alanine residue (i.e., A10).

with the NMR experiments. For pHLIP-1, distances are much closer to the POPC surface than wt-pHLIP for all three alanine residues (Figure 3.4 B). This is in part because A27 is not proximal to the three C-terminal acidic residues present in wt-pHLIP, which has the potential for electrostatic repulsion with the negatively-charged phosphate region of the PC headgroups. In addition, the distances for A10 and A13 are slightly reversed; A13 is closer to the P atoms than A10. This reversal could be due in part to the presence of negatively-charged residues at positions 2, 3, and 6 in pHLIP-1, which could have a more marked repulsive effect on the closer alanine residue (i.e., A10).
3.3.2 pHLIP samples secondary structural conformations in state II

The conventional model of state II proposes that the peptide remains in a coiled conformation when bound to a lipid bilayer at neutral or alkaline pH. For our simulations of wt-pHLIP, this is predominantly the case. However, we also observed that wt-pHLIP can form stable α-helices for hundreds of ns (Figure 3.6 A). In particular, the 0° and 288° orientations of wt-pHLIP undergo helical folding in multiple regions of the N-terminal half of the peptide, with turns centered around W9 and F17. There is precedence for this phenomenon, as deep-UV resonance Raman (DUVR) spectroscopy experiments showed that pHLIP can adopt helical structure in state II (Cooley and Jiji BC 2014). In contrast, pHLIP-1 undergoes helical folding regardless of orientation (Figure 3.6 B). This is consistent with CD spectra that showed pHLIP-1 possesses a noticeable degree of helicity in state II (Karabadzhak BJ 2012). Folding occurs in an extended segment from R11 to T19, and for three of the orientations (0°, 72°, and 216°), the C-terminus folds into a helix. In general, any occurrence of secondary structure is helical in nature (Figure 3.5). Visual inspection of our simulations shows that for both wt-pHLIP and pHLIP-1, partitioning via sidechain-headgroup interactions is most prevalent with the more non-polar residues (i.e., leucine and isoleucine). The role of aromatic and charged sidechains in binding of pHLIP to a POPC bilayer surface can be noticeably different between the two peptides, as will be discussed below.

This stark contrast in the onset of folding in state II is due to several factors. One of the most significant that we observed is formation of an i+3 salt bridge between R11 and D14 in pHLIP-1. All simulations of pHLIP-1 show stable formation of this salt bridge with little to no fluctuations over hundreds of ns (Figure 3.7). The R11-D14 interaction also directly coincides with the region of pHLIP-1 that folds into an α-helix. In the wt-pHLIP simulations, fewer salt bridges are formed and are very transient; no i+3 salt bridges occur (Figure 3.7). There appears to be no dominant intramolecular interaction, as R11 in wt-pHLIP interacts transiently with all of the negatively-charged residues, depending on the orientation of the peptide in the bound complex.
Figure 3.5: Aggregate sampling of secondary structure of wt-pHLIP and pHLIP-1 bound to POPC.

(A) Ramachandran map of the different orientations of wt-pHLIP (0°, 72°, 144°, 216°, and 288°) when bound to POPC. Data points represent cumulative phi/psi angles sampled for all residues of wt-pHLIP for each respective orientation. wt-pHLIP has a strong preference for sampling right-handed α-helical regions (i.e., lower left quadrant). Specifically, 0° and 288° orientations have negligible sampling of the β-sheet region (upper left-hand quadrant) and even in the 72° and 216° orientations, sampling of α-helix region is much stronger compared to β-sheet. The one notable exception is the 144° orientation, which dissociates from the bilayer. (B) Ramachandran map of the different orientations of pHLIP-1 (0°, 72°, 144°, 216°, and 288°) when bound to POPC. In all orientations, pHLIP-1 has a stronger preference for sampling α-helical regions compared to β-sheet regions. This is consistent with experiments which has shown that pHLIP-1 forms partial helices in state II.
Another key finding from the DUVRR study was that pHLIP undergoes desolvation in state II[6]. We observe similar behavior through a general decrease in the solvent-accessible surface area (SASA) upon binding to the bilayer surface. Although wt-pHLIP has variable partitioning into the POPC bilayer based on the orientation of the bound complex (Figure 3.2 A), large portions of the peptide undergo dehydration, illustrated by a decrease in SASA. In the 0° and 72° orientations, the N-terminal segment of wt-pHLIP is most dehydrated, while in the 216° orientation the sinker...
Figure 3.7: \(i+3\) salt bridge formation plays a key role in stabilization of secondary structures in pHLIP-1.

(A) Time-dependent salt bridge formation between R11 and sidechains of acidic residues in wt-pHLIP for \(0^\circ\), \(72^\circ\), \(144^\circ\), \(216^\circ\), and \(288^\circ\) orientations. For several orientations, formation of salt bridges between R11 and other residues is transient (\(0^\circ\), \(72^\circ\), and \(288^\circ\)) or switches between multiple acidic residues (\(144^\circ\)). Only the \(216^\circ\) orientation forms a stable salt bridge for hundreds of ns, but it is least likely to form secondary structures (Figure 3.6). \(\text{Black: E3; red: D14; grey: D25; magenta: D31; blue: D33; green: E34; grey background: minimum distance for salt bridge formation.}\)

(B) Time-dependent salt bridge formation between R11 and sidechains of acidic residues in pHLIP-1 for \(0^\circ\), \(72^\circ\), \(144^\circ\), \(216^\circ\), and \(288^\circ\) orientations. Even though there are more acidic residues on the N-terminus of pHLIP-1 that could be salt bridge partners, the only salt bridge that is formed is between R11 and D14. Not surprisingly, this also corresponds to the region in pHLIP-1 that forms a helical turn when bound to POPC. \(i+3\) salt bridges like the R11-D14 salt bridge have been shown to stabilize \(\alpha\)-helical peptides in solution[4, 33, 14], either through energetic stabilization of the backbone hydrogen bond needed to form a helical turn[4] or exclusion of solvent from the backbone hydrogen bond by the proximity of the sidechains[48]. Color scheme is the same as in A, except that E3 is replaced by D3 in pHLIP-1.
stretch is most dehydrated, and in the 288° orientation, the majority of the peptide is dehydrated (Figure 3.8 A). Dehydration can be a hallmark of partitioning or folding: for example, the N-terminus of the 0° orientation is partitioned into the headgroup region, dehydrated, and folds into an α-helix. However, the N-terminal portion of the 288° orientation is not well-partitioned but is dehydrated and folded into a helical turn. In contrast, regardless of orientation, regions of dehydration of pHLIP-1 are consistent with areas that 1) partition into the POPC bilayer and 2) fold into a helical turn (i.e., the middle portion of the peptide) (Figure 3.8 B). Finally, the SASA of these peptides in comparison to their behavior is solution may also explain why pHLIP-1 undergoes folding more readily than wt-pHLIP: pHLIP-1 has a markedly lower SASA in state I, requiring less than a 30% reduction in SASA from state I to state II, whereas wt-pHLIP must undergo almost 50% reduction in SASA (Figure 3.8).

3.3.3 pHLIP utilizes multiple types of non-bonded interactions in state II

Further inspection shows that R11 plays multiple roles in both wt-pHLIP and pHLIP-1 yet is not a requirement for formation of a bound complex. It is possible for R11 to form hydrogen bonds or salt bridges with the phosphatidylcholine headgroup in POPC. For wt-pHLIP, R11 can form from three to six hydrogen bonds with the PC headgroups (Figures 3.9 A and 3.10 A). These hydrogen-bonded interactions may be enhanced by stable salt bridge formation; in particular, the 0° and 72° orientations have salt bridges that are stable for hundreds of ns. Not surprisingly, these orientations show the most stable binding to POPC (Figure 3.2). Similarly, in pHLIP-1 R11 can form multiple hydrogen bonds as well as stable salt bridges over hundreds of ns (Figures 3.9 B and 3.10 B). This behavior correlates with stable binding of the surrounding residues in pHLIP-1 of each of the three orientations, again showing that R11 can play an important role in the formation of pHLIP-bilayer complexes (i.e., state II). Finally, other residues also contribute to hydrogen-bonded interactions of pHLIP with the bilayer; for example, pHLIP-1 forms multiple stable hydrogen bonds in the 0° and 72° orientations, none of which involve R11. The transmembrane
Dehydration of residues in pHLIP weakly correlates with partitioning into the surface of the POPC bilayer.

**Figure 3.8:** Dehydration of residues in pHLIP weakly correlates with partitioning into the surface of the POPC bilayer

(A) Upper plots: time-dependent change in per-residue solvent-accessible surface area (SASA) of wt-pHLIP bound to POPC. Although wt-pHLIP has variable partitioning into the POPC bilayer based on the orientation of the bound complex (Figure 3.2 A), large portions of the peptide undergo dehydration, shown by a decrease in SASA. In the 0° and 72° orientations, the N-terminal segment of wt-pHLIP is most dehydrated, while in the 216° orientation the sinker stretch is most dehydrated, and in the 288° orientation, almost the entire peptide is dehydrated. Dehydration can be a hallmark of either partitioning or folding: for example, the N-terminus of the 0 deg orientation is partitioned into the headgroup region, dehydrated, and folds into an α-helix. However, the N-terminal portion of the 288 deg orientation is not well-partitioned but is dehydrated and folded into a helical turn. Bottom plot: distribution of solvent-accessible surface area (SASA) of wt-pHLIP in solution (blue) and for the different orientations of wt-pHLIP bound to the bilayer surface (gray). In solution, wt-pHLIP is in a loosely-packed coil, leading to high solvent accessibility. In contrast, wt-pHLIP becomes less hydrated when bound to POPC. (B) Upper plots: time-dependent change in per-residue SASA of pHLIP-1 bound to POPC. In contrast, regardless of orientation, regions of dehydration of pHLIP-1 are consistent with areas that 1) partition into the POPC bilayer and 2) fold into a helical turn (i.e., the middle portion of the peptide). Bottom plot: SASA of pHLIP-1 in solution (blue) and bound to POPC (gray). Like wt-pHLIP, pHLIP-1 is less hydrated when bound to POPC. However, in solution, pHLIP-1 adopts a more compact coiled conformation, leading to less solvent accessibility.
Figure 3.9: Non-bonded interactions can contribute to binding of pHLIP but are not necessary for complex formation.

(A) Time-dependent hydrogen bond (gray) and salt bridge (red) formation between the phosphate of the PC headgroup and sidechain of R11 of wt-pHLIP. Salt bridge formation is considered only if it lasted for more than 8 ns. For both wt-pHLIP and pHLIP-1, two of the five orientations did not form any hydrogen bonds or salt bridges with POPC. It is clear that R11 is capable of playing a significant role in the binding of wt-pHLIP to POPC — the maximum number of hydrogen bonds formed by wt-pHLIP is six, and R11 often forms up to three hydrogen bonds with the PC head groups. These hydrogen-bonded interactions may also be enhanced by stable salt bridge formation; in particular, the 0° and 72° orientations have salt bridges that are stable for hundreds of ns. Not surprisingly, these orientations also show the most stable binding to POPC (Figure 3.2). (B) Time-dependent hydrogen bond (gray) and salt bridge (red) formation between the phosphate of the PC headgroup and sidechain of R11 of pHLIP-1. Similar to wt-pHLIP, R11 in pHLIP-1 can form multiple hydrogen bonds as well as stable salt bridges over hundreds of ns. This behavior correlates with stable binding of the surrounding residues in pHLIP-1 of each of the three orientations, again showing that R11 can play an important role in the formation of pHLIP-bilayer complexes (i.e., state II).

Segment of wt-pHLIP and pHLIP-1 (residues 8-30 according to solid-state NMR experiments[16]) has a dual topology. In particular, the N-terminal portion has several aromatic sidechains (Y8, W9, Y12, and W15) that possess highly favorable partitioning free energies[52]. Based on the depth of partitioning of W9 and W15, it appears that there is no discernible behavior with respect to the tryptophan sidechains that enhance binding to the POPC bilayer (Figure 3.11). However, a closer examination
Figure 3.10: R11 consistently utilizes hydrogen bonds to stabilize interactions of pHLIP with a POPC bilayer surface.

(A) Distribution of the number of R11-choline hydrogen bonds in the different orientations of wt-pHLIP (0°, 72°, 144°, 216°, and 288°) when bound to POPC. Even if R11 has not formed a stable hydrogen-bonded complex with POPC (i.e., the 288° orientation, Figure 3.9), R11 can still form a hydrogen bond with the PC headgroups over 40% of the total simulation time.

(B) R11-choline hydrogen bond in the different orientations of pHLIP-1 (0°, 72°, 144°, 216°, and 288°) when bound to POPC. pHLIP-1 has a significantly higher frequency of hydrogen bond formation involving R11 compared to wt-pHLIP.

tells a different story. If we characterize the relationship between orientation of the tryptophan sidechain and the depth of partitioning into the POPC bilayer, W9 in wt-pHLIP adopts two distinct conformations (∼40° and ∼120°, slightly oriented towards and away from the bilayer normal) at the level of the phosphorus atoms in the PC headgroups (Figure 3.13 A). W15 in wt-pHLIP also adopts two major orientations, with the dominant population at 60° leading to favorable partitioning into the headgroups (∼5 Å below the P atoms). In contrast, there is no apparent correlation between the tryptophan sidechain orientation and depth of binding for pHLIP-1 (Figure 3.13 B). Extending this analysis to Y8 and Y12 yields similar results: both Y8 and Y12 in wt-pHLIP have sidechain orientations that adopt a bimodal distribution. The populations that have sidechain orientations pointed inward and nearly parallel to the membrane normal (i.e., < 45°) also possess the most effective binding (Figure 3.14 A). Y8 in pHLIP-1 is similar to W9 and W15, with no preferred orientation and lying above the headgroup region (Figure 3.14 B). Interestingly, the dominant population of Y12 in pHLIP-1 occurs when the tyrosine sidechain adopts an orientation that is parallel to the bilayer surface; this orientation does not allow the sidechain
Figure 3.11: Aromatic and cationic sidechains exert distinct effects on partitioning of pHLIP into POPC.

(A) Time-dependent distance of the center of mass (COM) of the sidechains of W9 (black), R11 (green), and W15 (grey) of wt-pHLIP from the COM of the POPC bilayer. No conserved behavior appears to exist for wt-pHLIP with respect to orientation of the peptide to the bilayer surface. The most stable and effective binding is for the 288° orientation, in which R11 is below the choline level of the PC headgroup and can form either salt bridges or hydrogen bonds with the phosphatidyl moiety (see Figure 3.9 for more details). (B) Time-dependent distance of the COM of the sidechains of W9 (black), R11 (green), and W15 (grey) of pHLIP-1 from the COM of the POPC bilayer. Regardless of orientation, R11 binds more effectively than W9 or W15.

to intercalate between lipids, unlike the other aromatic sidechains in either pHLIP peptide. This preferred orientation may be introducing a localized defect in the membrane surface, a potential insertion mechanism for several types of cell-penetrating peptides[25].
3.3.4 Relating simulation results to theory

Wimley and White developed a scale for determining the free energy of partitioning from the aqueous phase into a POPC bilayer for amino acid residues. Aromatic and nonpolar amino acids have the most favorable partitioning free energies, while polar (and in particular, charged) residues have the most unfavorable partitioning free energies. Using this scale, we can predict the segment of wt-pHLIP and pHLIP-1 that will most favorably bind to the bilayer (Figure 3.12). The most favorable stretch of wt-pHLIP that can partition into the bilayer is residues 8–31 ($\Delta G_{\text{partition}} = -6.6$ kcal/mol, Table 3.1). This stretch corresponds remarkably well with the residues that partition most effectively into the POPC surface (see $0^\circ$ orientation, Figure 3.2). Likewise, for pHLIP-1, the most favorable stretch that can partition into the bilayer is residues 8-30. Note that in pHLIP-1, the favorable stretch includes the C-terminus, most likely due to the absence of charged residues that are present in wt-pHLIP.

Comparing the per-residue depth of partitioning for wt-pHLIP shows that in general, the C-terminal end of wt-pHLIP remains distal from the P atoms in the PC headgroups, with slight partitioning of the peptide from residues $\sim$17-23 for the $0^\circ$, $216^\circ$, and $288^\circ$ orientations (Figures 3.2 and 3.3). The residues in wt-pHLIP with the most negative partitioning free energies (i.e., the aromatic residues Y8, W9, Y12, and W15) do not exert a localized partitioning effect. However, stretches of residues favorable to partitioning (e.g., the leucine residues from position 21 to 29) have a more proximal effect on the ability of wt-pHLIP to bind and partition into the POPC surface. Similar to the C-terminus of wt-pHLIP, the N-terminal half of pHLIP-1 predominantly lies away from the PC headgroups, due to the presence of acidic residues. However, this repulsion is compensated by partitioning of the middle and C-terminal portions of pHLIP-1 into the bilayer (Figure 3.3 B). It appears that partitioning of the C-terminal “sinker stretch” is more pronounced in pHLIP-1 compared to wt-pHLIP, most likely due to the absence of acidic residues (D31, D33, and E34) on the C-terminus of pHLIP-1.
Figure 3.12: Wimley-White partition free energies provide insight into optimal binding of pHLIP.

(A) Theoretical per-residue cumulative partition free energy of wt-pHLIP. Values are based on the scale for free-energy partitioning of amino acid residues into membranes developed by Wimley and White[52]. \textit{x-axis:} starting point in wt-pHLIP for inclusion of per-residue partitioning free energies. \textit{y-axis:} ending point in wt-pHLIP for inclusion of per-residue partitioning free energies. (B) Theoretical per-residue cumulative partition free energy of pHLIP-1. Dashes indicate the segment of each peptide that possess the most favorable partitioning free energies.

3.4 Discussion

In this study we have computationally modeled state II of wt-pHLIP and the variant, pHLIP-1. This is the first computational study of its kind, as previous MD studies have focused on insertion of \(\alpha\)-helical pHLIP[9, 49], conformational sampling of gold nanoparticle-pHLIP conjugates[29], and pHLIP in solution[16, 15]. In contrast, the field of antimicrobial peptides and cell-penetrating peptides has been extensively studied by MD simulations[46, 35, 44]. This distinction of pHLIP from other CPPs is mainly tied to the anionic versus cationic nature of these two classes of peptides: the ability of acidic residues to quickly undergo titration to protonated states allows pHLIP to make rapid transitions from state II (coiled, bound) to state III (helical, inserted) when exposed to acidic conditions[24, 37]. However, explicitly changing protonation states is not easily accomplished with conventional MD simulation techniques, since classical mechanics do not allow for formation and scission of the O–H bond in acidic amino acid residues. This shortcoming is also compounded by the long
Figure 3.13: Sidechain orientation of tryptophans have variable effect on binding of pHLIP.

A) Cumulative 2-D distribution of orientations of the indole sidechain with respect to depth in the POPC bilayer in W9 (top) and W15 (bottom) of wt-pHLIP. Depth is defined as the projection on the membrane normal of the distance between the center of mass of each respective tryptophan sidechain and the average position in the z axis of the center of mass of the phosphorus atoms in the upper leaflet of the POPC bilayer. For W9, two major populations exist, both of which contribute to partial burial of W9 into the bilayer (i.e., 40° and 130° orientations). In contrast, a single population is dominant for W15 which leads to burial of the sidechain: 60°. The majority of the other orientations of W15 reside at the bilayer surface. (B) Cumulative 2-D distribution of orientations of the indole sidechain with respect to depth in the POPC bilayer in W9 (top) and W15 (bottom) of pHLIP-1. Unlike wt-pHLIP, pHLIP-1 does not have a preferred orientation of either tryptophan sidechain. W9 resides at the surface of the bilayer, whereas W15 mainly lies at the same level as the P atoms in the PC headgroups. This indicates that the tryptophan residues in pHLIP-1 are not essential to effective binding.

Timescales associated with pHLIP insertion (ms–s), as even state-of-the-art membrane protein simulations are only capable of reaching tens to hundreds of μs[10, 11]. We have chosen to focus on the binding step in pHLIP function, as it avoids these two issues. We can glean important details about the binding mechanism of pHLIP based on several structural studies. The first study, using brominated phospholipids, determined that the tryptophan residues of pHLIP (W9 and W15) are embedded
Figure 3.14: Sidechain orientation of tyrosines have variable effect on binding of pHLIP.

(A) Cumulative 2-D distribution of orientations of the indole sidechain with respect to depth in the POPC bilayer in Y8 (top) and Y12 (bottom) of wt-pHLIP. Depth is defined as the projection on the membrane normal of the distance between the center of mass of each respective tryptophan sidechain and the average position in the z axis of the center of mass of the phosphorus atoms in the upper leaflet of the POPC bilayer. For Y8, the largest population of orientations are when the tyrosine sidechain is parallel to the membrane normal (180°) and above the phosphorus atoms. For Y12, there are two preferred orientations (45° and 135°), with slightly different effectiveness in binding. (B) Cumulative 2-D distribution of orientations of the indole sidechain with respect to depth in the POPC bilayer in Y8 (top) and Y12 (bottom) of pHLIP-1. As with the N-terminal tryptophan residue (W9), the sidechain of Y8 is completely above the headgroup region and samples almost all possible orientations, indicating it is in free solution. However, Y12 most often lies parallel to the bilayer surface and partitions into the headgroup region (about 3 Å below the level of the P atoms), which may be causing local perturbation of the bilayer as well as facilitating other non-bonded interactions such as formation of hydrogen bonds and salt bridges of R11 with the PC headgroups.

at different depths within the bilayer[56]. The DUVRR study of Cooley, Jiji, and coworkers determined that pHLIP undergoes partial helical formation in state II, most likely in conjunction with dehydration of the peptide backbone[6]. In addition, the REDOR NMR study of An, Qiang, and coworkers focused on interactions of 13C-labeled nonpolar residues (A10, A13, L21, L22, L26, A27) with POPC[43]. They
found that the N-terminal residues (A10, A13) were in close proximity to the phosphatidylcholine headgroups in both state II and state III, whereas the C-terminus most likely remained unbound to the bilayer surface until the pH reached slightly acidic levels (pH 6.4). Although they observed two populations of conformations at pH 6.4, the C-terminus (i.e., the “sinker stretch”) was helical for both, indicating that helix formation most likely initially occurs around the D25 protonation switch. In their latest study, $^{13}$C solid-state NMR was used to determine that the C-terminal aspartic acid residues undergo titration first (i.e., highest $pK_a$’s), leading An, Qiang, and coworkers to postulate that the solvent-exposed residues on the C-terminus underwent protonation under acidic conditions more quickly than residues in the middle of the peptide (D14 and D25)[16]. By combining the insights gained from our simulations with these experimental observations of pHLIP, we can begin to formulate a general mechanism for binding of pHLIP in state II.

The first principle in understanding state II of pHLIP based on our simulations is that location of negatively-charged residues matters. In wt-pHLIP, the three negatively-charged residues on the C-terminus (D31, D33, and E34) undergo repulsive interactions with the PC headgroups; this relationship is the opposite for pHLIP-1, since there are negatively-charged residues on the N-terminus (E2, D3, and D6) (Figure 3.2). These unfavorable interactions are conserved in all of our simulations. A second factor that contributes to effective partitioning of pHLIP is hydrophobic interactions. The nonpolar residues from L21 to V30 form the “sinker stretch” posited by An and coworkers[43] and is bisected by D25. Using the hydrophobicity scale proposed by Wimley and White[52], the sinker stretch has a favorable cumulative free energy of partitioning from water to the membrane of $-2.45$ kcal-mol$^{-1}$ (Table 3.1). If we include the C-terminal acidic residues present in wt-pHLIP, the free energy of partitioning becomes unfavorable ($+2.03$ kcal-mol$^{-1}$). Clearly, the ability of the sinker stretch to partition into the bilayer can be influenced by the number of anionic residues present in the C-terminus of pHLIP. However, the presence of these residues do not lead to complete dissociation of wt-pHLIP, as the sinker stretch remains bound to the POPC surface in the majority of our simulations (Figure 3.2 A). Other non-bonded interactions play a critical role in partitioning of pHLIP in state
II, specifically

Table 3.1: Per-residue and cumulative free energy of partitioning of wt-pHLIP and pHLIP-1 based on the Wimley-White scale.

<table>
<thead>
<tr>
<th>pHLIP-1</th>
<th>Cumulative</th>
<th>wt-pHLIP</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 0.17</td>
<td></td>
<td>A2 0.17</td>
<td></td>
</tr>
<tr>
<td>E2 2.02</td>
<td>2.19</td>
<td>E3 2.02</td>
<td>2.19</td>
</tr>
<tr>
<td>D3 1.23</td>
<td>3.42</td>
<td>Q4 0.58</td>
<td>2.77</td>
</tr>
<tr>
<td>Q4 0.58</td>
<td>4</td>
<td>N5 0.42</td>
<td>3.19</td>
</tr>
<tr>
<td>N5 0.42</td>
<td>4.42</td>
<td>P6 0.45</td>
<td>3.64</td>
</tr>
<tr>
<td>D6 1.23</td>
<td>5.65</td>
<td>I7 -0.31</td>
<td>3.33</td>
</tr>
<tr>
<td>P7 0.45</td>
<td>6.1</td>
<td>Y8 -0.94</td>
<td>2.39</td>
</tr>
<tr>
<td>Y8 -0.94</td>
<td>5.16</td>
<td>W9 -1.85</td>
<td>0.54</td>
</tr>
<tr>
<td>W9 -1.85</td>
<td>3.31</td>
<td>A10 0.17</td>
<td>0.71</td>
</tr>
<tr>
<td>A10 0.17</td>
<td>3.48</td>
<td>R11 0.81</td>
<td>1.52</td>
</tr>
<tr>
<td>R11 0.81</td>
<td>4.29</td>
<td>Y12 -0.94</td>
<td>0.58</td>
</tr>
<tr>
<td>Y12 -0.94</td>
<td>3.35</td>
<td>A13 0.17</td>
<td>0.75</td>
</tr>
<tr>
<td>A13 0.17</td>
<td>3.52</td>
<td>D14 1.23</td>
<td>1.98</td>
</tr>
<tr>
<td>D14 1.23</td>
<td>4.75</td>
<td>W15 -1.85</td>
<td>0.13</td>
</tr>
<tr>
<td>W15 -1.85</td>
<td>2.9</td>
<td>L16 -0.56</td>
<td>-0.43</td>
</tr>
<tr>
<td>L16 -0.56</td>
<td>2.34</td>
<td>F17 -1.13</td>
<td>-1.56</td>
</tr>
<tr>
<td>F17 -1.13</td>
<td>1.21</td>
<td>T18 0.14</td>
<td>-1.28</td>
</tr>
<tr>
<td>T18 0.14</td>
<td>1.35</td>
<td>T19 0.14</td>
<td>-1.28</td>
</tr>
<tr>
<td>T19 0.14</td>
<td>1.49</td>
<td>P20 0.45</td>
<td>-0.83</td>
</tr>
<tr>
<td>P20 0.45</td>
<td>1.94</td>
<td>L21 -0.56</td>
<td>-1.39</td>
</tr>
<tr>
<td>L21 -0.56</td>
<td>1.38</td>
<td>L22 -0.56</td>
<td>-1.95</td>
</tr>
<tr>
<td>L22 -0.56</td>
<td>0.82</td>
<td>L23 -0.56</td>
<td>-2.51</td>
</tr>
<tr>
<td>L23 -0.56</td>
<td>0.26</td>
<td>L24 -0.56</td>
<td>-3.07</td>
</tr>
<tr>
<td>L24 -0.56</td>
<td>0.3</td>
<td>D25 1.23</td>
<td>-1.84</td>
</tr>
<tr>
<td>D25 1.23</td>
<td>0.93</td>
<td>L26 -0.56</td>
<td>-2.4</td>
</tr>
<tr>
<td>L26 -0.56</td>
<td>0.37</td>
<td>A27 0.17</td>
<td>-2.23</td>
</tr>
<tr>
<td>A27 0.17</td>
<td>0.54</td>
<td>L28 -0.56</td>
<td>-2.79</td>
</tr>
<tr>
<td>L28 -0.56</td>
<td>-0.02</td>
<td>L29 -0.56</td>
<td>-3.35</td>
</tr>
<tr>
<td>L29 -0.56</td>
<td>-0.58</td>
<td>V30 0.07</td>
<td>-3.28</td>
</tr>
<tr>
<td>V30 0.07</td>
<td>-0.51</td>
<td>D31 1.23</td>
<td>-2.05</td>
</tr>
<tr>
<td>D31 1.23</td>
<td>-0.36</td>
<td>D32 1.23</td>
<td>-0.65</td>
</tr>
<tr>
<td>D32 1.23</td>
<td>0.14</td>
<td>E34 2.02</td>
<td>1.37</td>
</tr>
<tr>
<td>E34 2.02</td>
<td></td>
<td>G35 0.01</td>
<td>1.38</td>
</tr>
<tr>
<td>G35 0.01</td>
<td></td>
<td>T36 0.14</td>
<td>1.52</td>
</tr>
<tr>
<td>T36 0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8 to 31 -6.6
8 to 30 -6.92
the N-terminal half of the transmembrane segment. The presence of several aromatic residues (Y8, W9, Y12, and W15) and the lone cationic residue (R11) near the D14 protonation switch lead to a noticeably different binding mechanism from the sinker stretch in pHLIP. (Interestingly, the N-terminal half of the transmembrane segment in wt-pHLIP and pHLIP-1 has a more favorable free energy of partitioning (−3.67 kcal-mol$^{-1}$) than the sinker stretch (Table 3.1).) In addition to the aforementioned DUVRR study showing significant desolvation of wt-pHLIP in state II, the fluorescence spectra of pHLIP-1 in state II were more intense than wt-pHLIP, indicating that the tryptophans of pHLIP-1 become more deeply embedded within the bilayer upon binding[24, 6]. In our simulations, tryptophans play different roles for wt-pHLIP and pHLIP-1: in wt-pHLIP, one or both tryptophan residues partition below the surface and discarding solvent molecules only when the sidechain is parallel to the membrane normal, while in pHLIP-1 the orientations of tryptophans appear to have little effect on binding. Similar behavior is seen with the tyrosine residues, with Y12 in pHLIP-1 playing a potential role in creating a local defect in the bilayer surface. Binding of aromatic sidechains is favored due to a combination of factors (the quadrupolar moment of the indole ring, the hydrophobic effect, and the complex electrostatic environment of the headgroup region[54]). In lieu of binding through interactions of aromatic sidechains with the bilayer, pHLIP-1 appears to utilize R11 to form a salt bridge with D14 to facilitate partitioning. R11 may be further stabilizing partitioning by forming salt bridges with the PC headgroups; this is a common mechanism for stabilizing small peptide-membrane interactions (e.g., the Kras protein[20]). The combination of nonbonded interactions with the bilayer illustrates how a peptide like pHLIP can partition into the membrane surface, even in the presence of charged residues.

Finally, binding of pHLIP to the membrane is enhanced by formation of helical turns. Partitioning-folding coupling for cationic peptides is well-characterized, with a reduction of -0.4 kcal-mol$^{-1}$ per residue for a peptide to undergo folding into a helix while bound to a bilayer surface[31]. In a similar way, the folding that occurs in both pHLIP peptides contributes to favorable partitioning into the bilayer. Our observation of stable helical turns in pHLIP-1 is consistent with the original investigation of
pHLIP-1, which detected the presence of α-helical folds in CD spectra of state II[24]. What was unexpected is that we observed formation of stable helical turns in wt-pHLIP in several of our simulations as well (Figure 3.6 A). The fact that wt-pHLIP and pHLIP-1 undergo folding into helical turns in different segments of the peptide is consistent with proposed models of the reversible state II to state III transition[16, 2, 41, 49]. In wt-pHLIP, folding mainly occurs from I7-Y12 and D14-T18 but also in the sinker stretch from L21-V30. In pHLIP-1, folding is almost completely conserved from Y12-T18 and can extend from L22-V30. Based on solid-state NMR and fluorescence spectroscopy studies, the C-terminal half of the transmembrane segment of pHLIP is most sensitive to changes in pH ([16, 41] and can form a peripheral helix prior to insertion[2]. In addition, recent constant-pH MD studies that were initiated in state III showed that at pH 7, the N-terminal half of the transmembrane segment (i.e., including D14) diffuses towards the surface of the bilayer but remains in an α-helical fold[49]. Regardless of the location of folding (i.e., centered around D14 or D25), it appears that wt-pHLIP and pHLIP-1 are undergoing a priming action for the transition to state III: helical formation can occur while either proton switch is deprotonated and the hydrophilic side chain is solvent-exposed. In the case of D25, upon acidification and protonation of the residue (a transition from +1.23 to −0.07 kcal-mol$^{-1}$ in partitioning free energy), the C-terminus of pHLIP becomes continuously nonpolar, facilitating the more extensive helical formation that is the hallmark of the transition from state II to state III. Protonation of D14 would lead to a similar outcome, as that portion of the transmembrane segment also possesses a favorable partitioning free energy. The propensity for each of these regions to undergo partial helical folding, even in state II, is not surprising; our recent study showed that pHLIP spends a noticeable portion of time sampling helical conformations in solution[15]. It is conceivable that pHLIP could be partially folded when binding to a membrane surface. Multiple pathways exist for pHLIP to bind to the cell membrane surface, but preferential binding is governed by specific biophysical characteristics of the peptide (Figure 3.15). Before the transition to acidic pH and state III, the sinker stretch of the C-terminus of the peptide begins to bury itself beneath the headgroup region and into the hydrophobic interior of the bilayer. In our simulations, this effect can be modulated by the presence (wt-pHLIP) or absence (pHLIP-1) of acidic residues.
pHLIP utilizes a combination of non-bonded forces to achieve binding to the cell membrane surface. Blue: the C-terminal half of pHLIP uses its “sinker stretch” of hydrophobic residues to embed below the level of the lipid headgroups, sequestering itself from aqueous solvent. This interaction is responsible for the helical priming action observed around the D25 protonation switch. Inset: colored spheres represent sinker stretch residues that interact with the headgroups and acyl chains of the POPC bilayer upper leaflet, grey surface. Green: burial of the aromatic sidechains of Y8, W9, Y12, and W15 within the PC headgroups helps anchor the middle portion of pHLIP to the bilayer surface. These interactions stabilize pHLIP around the D14 protonation switch and allow for helical formation. Inset: interaction of W9/W15 (spheres) with the headgroup region of the POPC bilayer. Red: hydrogen bonding or salt bridge formation of R11 with the PC headgroup can work in conjunction with aromatic burial to stabilize binding of pHLIP. Inset: R11 (sticks) forming a hydrogen bond with the phosphate of a PC headgroup (spheres).

Furthermore, it is clear that binding via the N-terminal half of pHLIP can also occur. This interaction may require more extensive sampling of potential peptide-bilayer configurations, since binding takes place through a combination of hydrophobic effects (aromatic side chains) and electrostatic ones (R11). Our results indicate that
state II is the result of a two-step binding process, one in which both halves of the
transmembrane helical domain (bisected by P20) undergo burial into the bilayer in-
terface. A more general issue we have begun to address is the complexity of how
pHLIP interacts with a cell membrane. The field has progressed from the conven-
tional three-state model to begin to account for differences in membrane composition
from POPC[30, 42, 47] as well as utilizing advanced spectroscopic techniques that
allow for a more detailed characterization of pHLIP conformational sampling in the
bound and inserted states[43, 16, 2, 41]. A combination of experimental and compu-
tational approaches will be required to fully understand how pHLIP functions, which
will subsequently enable design of more specific and effective variants of pHLIP.

3.5 Conclusion

The computational studies carried out here provide strong support for the experimen-
tal biophysical studies characterizing state II of pHLIP. Intimate knowledge at the
molecular level of detail is fundamental to identifying specific interactions that govern
the free energy release upon binding of pHLIP to the cell membrane surface. This
free energy release forms one component of the thermodynamic cycle associated with
the “partitioning-folding coupling” process first proposed by Wimley and White[52].
They postulated that with this fundamental understanding, it would be possible to
formulate a governing set of rules for the rational design of peptides with specific
partition coefficients and secondary structures[51]. Efforts are currently underway
in our lab to extend computational investigations of this thermodynamic cycle and
bring Wimley and White’s goal to fruition.
References


REFERENCES

1474-175X, 1474-1768. DOI: 10.1038/nrc1478. URL: http://www.nature.com/doifinder/10.1038/nrc1478.


[20] Lorant Janosi and Alemayehu A. Gorfe. “Segregation of Negatively Charged Phospholipids by the Polycationic and Farnesylated Membrane Anchor of Kras”.

[1474-175X, 1474-1768]
[21] Sunhwan Jo, Taehoon Kim, and Wonpil Im. “Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations”. In: PLOS ONE 2.9 (Sept. 2007), e880. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0000880. URL: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000880.


Chapter 4

Binding of pHLIP to bilayer surfaces: Effect of ionic strength

4.1 Introduction

As mentioned in chapter 1, most experimental studies of pHLIP have been performed under low salt concentration (1-20 mM) to avoid aggregation of pHLIP at higher ionic strength\cite{29, 30, 14, 1, 17, 33, 34, 8, 13}. However, development of pHLIP into a cancer-targeting agent requires a fundamental understanding of pHLIP behavior at biological ionic strength (~150 mM). In principle, addition of ions could modify pHLIP-membrane bilayer in two ways: (1) interactions of ions with pHLIP and (2) interaction of ions with the membrane. The latter has been extensively studied by experiments as well as simulations (see below). A recent review summarizes the significance of membrane-ion interactions\cite{7}.

Interaction of Na\(^+\) ions with lipid membranes was first demonstrated using NMR quadrupolar coupling back in 1972\cite{18}. Later, similar behavior was shown for other alkali metal cations\cite{19}. Competition between different cations for lipid headgroups has also been reported\cite{16}. More recently, Petrarche et. al\cite{27}. has reported that in addition to electrostatic effects, salts also cause a weakening of van der Waals interactions in membranes. Klasczyk et. al\cite{15}. has measured zeta-potential of membrane surfaces in the presence of ions, showing that ions increase the zeta potential before saturating at ~100 mM. The same work also showed that ion-binding to membranes was entropy-driven, related to the release of waters from their hydration shells of the ion and the membrane headgroups. AFM studies have shown that ions lead to
rigidity of membranes, and the effect is related to the nature and size of the cation as well as the nature of lipid headgroups\[28\].

Molecular dynamics simulations have been extensively used to study the ion-membrane interaction in detail. Bockmann et. al\[2\] used a combination of fluorescence correlation spectroscopy and MD simulation to reveal the binding of Na\(^+\) ions to the carbonyl oxygens of a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer, and demonstrate that such binding reduces the diffusivity and increases the area-per-lipid of the POPC lipids. Kagawa et. al\[11\] has suggested that the decrease of membrane fluidity might be an indirect effect of the ions: mediated by lipid dehydration caused by the presence of cations. Yang et. al\[38\] used long-timescale MD simulations to compare the effects of Na\(^+\), K\(^+\), and Ca\(^{2+}\) ions on the anionic, 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1’-rac-glycerol) (POPG) bilayer. Metadynamics MD simulation has also been used to calculate free energy profiles (potential of mean force or PMF) of each of these ions to dimyristoylphosphatidylcholine (DMPC) bilayer\[39\].

Here, we employ atomistic molecular dynamics simulations to gain insights into the effects of ionic strength on the interactions of pHLIP with lipid membranes. We intended to use Gaussian Accelerated Molecular Dynamics (GaMD) method outlined in section 2.6.1. The rationale was that GaMD would allow us to watch pHLIP bind to the membrane as well as enable us to calculate the associated potential of mean force (PMF). GaMD have been successfully used for lipid-containing systems\[23, 24\]. However, preliminary analysis showed that the pHLIP-membrane bound complex was too unstable for GaMD acceleration. We eventually used what we call “brute-force method”, similar to that used in chapter 3, wherein we ran unbiased MD simulation which were repeated 5 times to obtain good statistics.

4.2 Materials and Methods

4.2.1 System setup and analysis

pHLIP was prepared as described in chapter 2 and merging of pHLIP into the bilayer was performed in a manner described in chapter 3 with two key differences. First,
instead of a folded-over conformation taken from the implicit solvent simulation, we chose to start with an extended conformation of pHLIP. Second, instead of merging pHLIP into the bilayer, we placed it close to the bilayer surface (pHLIP COM ~20 from the surface of phosphorus atoms). This allows us to observe the binding event as well as maintain bilayer symmetry (same number of lipid molecules in both the leaflets). Note that we need 5 Na\(^+\) ions to neutralize the system as (completely deprotonated) pHLIP has a surface charge of -5. The ionic strength and number of NaCl molecules mentioned in the rest of this chapter does not include these 5 sodiums.

Five different systems were generated which were identical except for the number of NaCl molecules (plus a slight variation in the number of water molecules which happens as Na\(^+\) and Cl\(^-\) ions were added by replacing water molecules). The number of water and NaCl molecules along with the final ionic strength is provided in the table below. The density of water was taken as 31.05 Å\(^3\) per molecule. Additionally, two lipid-only systems were also generated for comparison. All analysis was performed using VMD\(^{[10]}\), LOOS\(^{[31]}\), CPPTRAJ\(^{[4]}\), and home-built tcl and python scripts.

<table>
<thead>
<tr>
<th>Ionic strength</th>
<th>Number of water molecules</th>
<th>Number of NaCl molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 mM</td>
<td>20395</td>
<td>0</td>
</tr>
<tr>
<td>20 mM</td>
<td>20393</td>
<td>8</td>
</tr>
<tr>
<td>50 mM</td>
<td>20345</td>
<td>19</td>
</tr>
<tr>
<td>100 mM</td>
<td>20345</td>
<td>38</td>
</tr>
<tr>
<td>150 mM</td>
<td>20345</td>
<td>57</td>
</tr>
<tr>
<td>25 mM (lipid only)</td>
<td>6216</td>
<td>3</td>
</tr>
<tr>
<td>154 mM (lipid only)</td>
<td>6186</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 4.1: Table of the systems under study

4.2.2 Simulation

Gaussian Accelerated Molecular Dynamics simulations

Initially, we attempted to model the lipid binding of pHLIP using dual-boost Gaussian Accelerated Molecular Dynamics (GaMD)\(^{[22]}\). We have used it successfully to model state I of pHLIP (section 2.6) using AMBER simulation. However, when simulating state II, we initially preferred to use NAMD simulation engine, primarily because of the ease with which we can define a collective variable (colvar). In order to keep
Chapter 4. Binding of pHLIP to bilayer surfaces: Effect of ionic strength

pHLIP from interacting with the lipids in the periodic image, we define a “upper-WallConstant” colvar that defines the distance of different parts of pHLIP from the lipid bilayer. Putting a large energetic cost when this variable crosses a cutoff makes it impossible for pHLIP to interact with the periodic image, without affecting the dynamics of the system. GaMD has been implemented in NAMD[26].

For these simulations, we used CHARMM 36 force field and modeled non-bonded interactions by applying force-switching after 10 Å before cutting-off at 12 Å. We also performed did simulations without force-switching as well as without the colvars described above to see its effect on GaMD acceleration (Section 4.3.1). For comparison, we also performed GaMD simulations in AMBER 16. We used $\sigma_{0P} = \sigma_{0D} = 6.0$ kCal/mol for all our simulations.

Equilibrium MD simulations

After minimizing and equilibrating the structure for 3 ns, it was converted to Amber parameters (lipid 14, ff14SB, OPC water model). We chose to use OPC water as TIP3P is known to bias secondary structure formation. Five replicates of all but the lipid-only systems were simulated. Each simulation was for 400 ns. Overall, results from a total of 10.8 $\mu$s simulation is presented here. All simulations were performed using Amber16[3] in the NPT ensemble. Temperature was kept at 310 K using Langevin thermostat. Berendsen (default) barostat was used with anisotropic pressure coupling. Non-bonded interactions were cutoff at 10. Hydrogens were constrained by SHAKE algorithm.

4.3 Results and Discussions

4.3.1 Gaussian Accelerated Molecular Dynamics simulations

Surprisingly, we noticed a large difference in the acceleration obtained from NAMD and AMBER simulations, under similar settings (Figure 4.1). To verify whether this was due to the colvars and/or the force-switching (AMBER simulations used hard cutoff of 8 Å), we repeated the simulations without colvars and without force-switching. This had very little effect on the GaMD acceleration using NAMD. At this point we are unsure of the reason behind this difference.
We then performed GaMD simulations in AMBER. However, we failed to observe stable binding even after 200 ns simulation (data not shown). We speculate that GaMD boost potentials might be destabilizing the pHLIP-POPC complex. Rest of the work is based on unbiased MD simulation performed in AMBER simulation engine.

**Figure 4.1:** GaMD acceleration obtained by NAMD vs. AMBER simulation engine
4.3.2 Integrity of the lipids

Integrity of the lipids was verified by calculating the area-per-lipid and the lipid order parameter of the palmitoyl (PA) and oleoyl (OL) chains (figure 4.2). Figure shows results for 150 mM (with pHLIP, averaged over 5 simulations) and 154 mM (without pHLIP) simulations, but similar results were obtained for all systems (data not shown).

![Figure 4.2: Area-per-lipid (left) and lipid order parameter of PA and OL chains of POPC (right) with and without pHLIP.](image)

4.3.3 pHLIP binds transiently to POPC bilayers regardless of salt concentration

We characterized the binding of pHLIP to POPC bilayer by measuring the Z-projection of the distance of Ca of every residue from the P atoms of the nearest leaflet. Analysis of the per-residue distance between pHLIP and the P atoms in the headgroup region does not show any trends as a function of salt concentration (Figure 4.3). During the timescale of our simulations (400 ns), we observed binding that could be either transient or long-lived (> 100 ns). For all concentrations, multiple association and dissociation events were observed. When stable binding does occur, it is often associated with an increase in the number of hydrogen bonds between pHLIP and the bilayer surface. Based on these results, it appears that salt concentration does not influence membrane-binding of pHLIP.
Chapter 4. Binding of pHLIP to bilayer surfaces: Effect of ionic strength

Figure 4.3: Z-projection of distance of Ca of every residue from the P atoms of the nearest leaflet

Each row is an independent simulation. Red: bound to the surface; blue: dissociated from POPC. Color-coded distances are represented by heat bar on right. Bottom: number of pHLIP-lipid hydrogen bonds as function of time. Several observations can be made from these simulations. Binding of pHLIP to POPC is transient, but it can also be somewhat stable, lasting for hundreds of ns in many cases. In addition, it is possible for pHLIP to undergo more than one binding/unbinding event over the course of a 400 ns simulation. Finally, an increase in the number of hydrogen bonds is correlated with binding of pHLIP in general (i.e., red regions of heatmap).
4.3.4 PC headgroups decrease diffusion of Na$^+$ ions.

Next, we looked at the interactions of Na$^+$ ions with pHLIP and the O atoms of the phosphate moiety of POPC. We see that increased salt concentration does not affect the coordination of Na$^+$ ions with the PC headgroups. The first coordination shell is at ~2.2 \text{ Å} from the oxygen atoms (Figure 4.4 A). When salt concentration increases, there is a slight decrease in the RDF at the second coordination shell. The residence times of Na$^+$ within 2.5 Å of the phosphate oxygens show that is not affected much by ionic strength, and is in the range of 1.0 to 1.1 ns (Figure 4.4 B). This is long enough for the Na$^+$ to bind to up to four headgroup oxygens (Figure 4.4 C). Our residence times are consistent with previous studies of Na$^+$ binding to headgroup regions of the CHARMM c36-lipid force field[36]. The coordination of Na$^+$ ions with PC headgroups is also consistent with previous MD simulation studies[25, 2]. Interestingly, we see a broadening of the RDF of the Na$^+$ ions from the C-terminal acidic residues of pHLIP when ionic strength increases, especially above 20 mM (Figure 4.4 D). The residence time of Na$^+$ ions near the C-terminal acidic residues is ~300 - 400 ps. This is also invariant across all ionic strengths (Figure 4.4 E). This time is long enough to allow the ions to coordinate with the negatively-charged acidic residues. For example, at 150 mM NaCl, a Na$^+$ ion can coordinate between two carboxylate groups on the sidechains of D31 and D33, screening the charge-charge repulsion of these two neighboring residues (Figure 4.4 F). The increase in salt concentration has specific but different effects on both pHLIP and the POPC bilayer.

Increase in ionic strength has the most effect on the anionic C-terminus of pHLIP. Increased ionic strength causes the coordination of Na$^+$ ions around the acidic residues of pHLIP to decrease, as shown by the RDF (Figure 4.5 A). However, ionic strength has no effect on the overall shape of pHLIP as it interacts with the POPC bilayer (Figure 4.5 B). This is consistent with our observation that salt concentration has no effect on the membrane-binding of pHLIP (Figure 4.3). Ionic strength does, however, have an effect on the interactions of the Na$^+$ ions with specific residues of pHLIP. The most noticeable effect is near the C-terminus of pHLIP. There is a significant increase in the interactions of Na$^+$ ions with the C-terminal acidic residues. Interestingly, the protonation switches (D14 and D25) has much less interactions with Na$^+$ ions. The
Chapter 4.  Binding of pHLIP to bilayer surfaces: Effect of ionic strength

Figure 4.4: Binding of sodium ions to PC headgroups.

(A) Radial distribution function (RDF) of Na\(^+\) ions around phosphate O atoms. (The RDF measures the probability of finding a Na\(^+\) ion within a given distance from the phosphate O atoms.) For all salt concentrations, sodium cations form the first coordination shell at ~2.2 Å. A very slight decrease in the RDF is observed with an increase in salt concentration, indicating that as concentration increases, it is less likely to observe Na\(^+\) ions near PC headgroups than in bulk solution. (B) Residence time of Na\(^+\) ions within 2.5 Å of phosphate oxygen. Color scheme is same as in A. Inset: residence times for Na\(^+\) atoms. We observe very little change in residence time as a function of concentration. (C) Representative snapshot of a Na\(^+\) ion complex with four PC headgroups in the POPC bilayer. (D) RDF of Na\(^+\) ions near the C-terminal acidic residues of pHLIP. Notice that the distance of the first shell is consistent with the RDF for lipids (A), but that there is a clear separation between 20mM and 50 mM (and higher) salt concentrations. We might be observing a transition to more general electrostatic screening effects from 20mM to 50mM (and above) (E) Residence times of Na\(^+\) near C-terminal acidic residues. Although the trend is similar to the residence times of sodium with lipids, the residence times are three times smaller (see table 4.2). (F) Distribution of distances between the carboxylate groups on the sidechains of D31 and D33 in pHLIP. Distances consistently stay between 6-10Å except at 150 mM NaCl, at which point coordination of a Na\(^+\) cation between the sidechains screens the charge-charge repulsion between the negatively-charged sidechains.

<table>
<thead>
<tr>
<th>Ionic strength</th>
<th>Average residence time (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 mM</td>
<td>1.01489</td>
</tr>
<tr>
<td>20 mM</td>
<td>1.10820</td>
</tr>
<tr>
<td>50 mM</td>
<td>1.11089</td>
</tr>
<tr>
<td>100 mM</td>
<td>1.01844</td>
</tr>
<tr>
<td>150 mM</td>
<td>1.13288</td>
</tr>
<tr>
<td>25 mM (lipid only)</td>
<td>1.21938</td>
</tr>
<tr>
<td>154 mM (lipid only)</td>
<td>0.99126</td>
</tr>
</tbody>
</table>

Table 4.2: Average residence time of Na\(^+\) ions near lipid phosphate oxygens.
same is true for the N-terminal acidic residue E3 (Figure 4.5 C). The coordination of Na+ ions with the C-terminal residues facilitates an increase in i + 3 salt bridge formation between R11 and D14, a hallmark of peptide folding[20] (Figure 4.5 D). Taken together with our analysis of the radius of gyration of pHLIP, we see that increased ionic strength screens the C-terminus of pHLIP, which leads to an elongation of pHLIP.

There is an increase in the helicity of pHLIP with increasing ionic strength. At low salt concentration, the N-terminal of pHLIP shows higher helix-forming propensity (Figure 4.6 A). As the salt concentration increases, the C-terminal acidic residues are screened by the Na+ ions. This causes an increase in the helicity in the sinker stretch (L21 to V30). It is to be noted here that in spite of the increased helicity, pHLIP, for the most part, is still in an overall coiled conformation, which is consistent with CD experiments.

4.4 Discussion

Understanding the effect of ionic strength on state II of pHLIP requires an understanding of the influence of ions on both pHLIP as well as the bilayer. Na+ binds well to carboxylate side chains and backbone carbonyls[37, 35, 21, 9]. Na+ has preferential binding to Asp over Glu[6]. Hence, we expect to see Na+ bind more often to D14, D25, D31, and D33 than to E3 and E34. Stronger binding of Na+ would make it difficult to protonate these residues. Thus, a higher [H+] would be needed in order to drive the protonation of acidic residues in pHLIP that lead to folding and insertion. We have shown that increase in salt concentration does not affect the binding of pHLIP to the POPC bilayer surface. We have observed multiple binding/unbinding events at each ionic strength, and stable binding with lifetimes > 100 ns is usually associated with increased hydrogen bond formation between pHLIP and POPC. Increased ionic strength causes a decrease in the specific coordination of Na+ around pHLIP. However, ionic strength has no effect on the overall shape of pHLIP, as shown by the radius of gyration (Figure 4.5 B). We see significant effects near the acidic C-terminus. There is a large increase in the interaction of the Na+ ions with the C-terminal acidic residues of pHLIP, which is 3-5 times higher than that with
Figure 4.5: Increasing salt concentration affects intramolecular interactions in pHLIP

(A) Radial distribution function (RDF) of Na\(^{+}\) ions around pHLIP (0 mM has 5 neutralizing ions). As salt concentration increases, there is less organization of coordination of cations around the negatively-charged residues in pHLIP (i.e., there is a continuous distribution of salt in the system). We speculate that this is due to the screening of negative sidechains of pHLIP at higher salt concentration. (B) Distribution of asymmetry parameter $\alpha$ calculated as $\alpha = (2I_1 - I_2 - I_3)/(I_1 + I_2 + I_3)$ where $I_j$ are the three moments of inertia. pHLIP is most compact (lower $\alpha$) at 0 and 100 mM. We do not see any definitive trends in the data, except that for all salt concentrations, the asymmetry parameter is $> 0.3$. (C) Percent of time when a Na\(^{+}\) ion is within 5 Å of each residue. As ionic strength increases, Na\(^{+}\) ions spend a significant amount of time near the C-terminus of pHLIP, which has the majority of the acidic residues. We see no discernible difference between 50 and 100 mM NaCl, but then it increases again with concentration at 150 mM. This result is also consistent with the smoothed out RDF in part A; since we are significantly increasing the number of Na\(^{+}\) ions, it is more likely for them to spend time exchanging between bulk solution and the negatively-charged amino acid residues on the C-terminus. (D) Percent of time R11 forms salt bridge with each of the acidic residues in pHLIP. At 0 mM, the R11-D33 salt bridge results in a globular conformation with lower $\alpha$ (B). With increasing ionic strength, the overall negative charge of the C-terminus is screened by Na\(^{+}\) ions resulting in an increase of the R11-D14 salt bridge and an elongation in the conformation of the polypeptide chain.
Per-residue helix-forming propensity of pHLIP as a function of salt concentration. Helix-forming propensity is defined using the STRIDE software, which identifies residues in pHLIP that undergo folding and satisfy the $\phi/\psi$ requirements for classification as an $\alpha$-helix. At zero salt content, pHLIP has a much greater probability of folding into a helix in the N-terminal half. Introduction of salts (starting at 20 mM and higher) leads to a screening effect on the C-terminal half of pHLIP. The highly polar nature of the acidic residues in the C-terminus of pHLIP (D25, D31, D33, and E34) is effectively screened by positively-charged sodium ions, acting in a similar manner to decreasing pH. This drives folding of the peptide into an $\alpha$-helix. However, overall helicity is not higher than ~20%, consistent with state II CD spectra that shows no $\alpha$-helical characteristics.

D14 and D25. Screening of the negatively charged residues in the C-terminal can facilitate $i+3$ salt bridge formation between R11 and D14[20]. Increased ionic strength also increases the helix-forming propensity of the sinker stretch of pHLIP, ranging from L21 to V30, which becomes possible after the Na$^+$ ions screen the repulsion between the acidic residues in the C-terminus.

Ions are known to decrease area per lipid and increase bilayer thickness, which could prevent pHLIP from partitioning into the bilayer[2, 25, 12]. Indeed, we see the
POPC area-per-lipid shift to lower values in our simulations involving POPC in 25 mM and 154 mM [NaCl] without pHLIP (Figure 4.7). Increase in NaCl concentration will also create a positive electrostatic potential at the surface of the bilayer. This would cause the surface to become more productive for the binding of the negatively-charge peptide. Na\(^+\) binding can also screen the electrostatic repulsion between anionic sidechains\(^{[5, 9]}\). This screening of charge-charge repulsion by ions is known to facilitate the interactions between those residues that are in close proximity when Na\(^+\) coordinates them\(^{[6]}\). In our case, that would be D25, D31, D33, and E34. The case of D31 and D33 is shown in Figure 4.4 F. Our simulations have shown that screening of near-neighbor acidic residues by increased salt concentration can facilitate folding into native state (i.e., helical turn)\(^{[40]}\). Baldwin et. al\(^{[32]}\). also saw this at low salt concentrations (<= .15 M), with an uncharged polypeptide.

4.5 Conclusion

Our simulations shed light on the complex interplay of interactions between pHLIP, POPC and ions that affect the binding of pHLIP to the surface of a POPC bilayer. We have found that increased salt concentration does not have a detrimental effect on pHLIP-binding. Presence of ions seem to have opposing effects on the membrane-binding of pHLIP - while coordination of Na\(^+\) ions to the acidic residues in the C-terminus of pHLIP screens their repulsion allowing them to form partial helices and interact better with the bilayer, Na\(^+\) ions also create a positive electrostatic potential near the surface of the bilayer which hinders the binding of pHLIP. Additionally, higher salt concentration can increase bilayer rigidity, affecting pHLIP-binding. Such knowledge is crucial for assessing the applicability of pHLIP as a cancer targeting agent, as it requires extending the results obtained from biophysical studies of pHLIP at low ( < 20 mM) ionic strength to understanding its behavior at biological ionic strength (~150 mM).
Figure 4.7: Ionic strength reduces the area-per-lipid of POPC bilayer

(A) Area-per-lipid as a function of simulation time in a 200-POPC system with 25 mM (red) and 154 mM (blue) salt concentration, and (B) Distribution of the area-per-lipid. Increased salt concentration reduces the area-per-lipid, indicating a more rigid bilayer. This rigidity might make it harder for pHLIP to bind to the bilayer surface at higher salt concentration, offsetting the advantage of the screening of C-terminal acidic residues of pHLIP. (C) Lipid order parameters of the palmitoyl (red) and oleyl (blue) chains of POPC in 25 mM (solid lines) and 154 mM (dashed lines) salt concentration. In spite of the difference in area-per lipid, the lipid order parameters are almost invariant to changes in ionic strength.
References


REFERENCES


REFERENCES


[34] Nicolas S. Shu et al. “Residue-specific structures and membrane locations of pH-low insertion peptide by solid-state nuclear magnetic resonance”. In: Nature Communications 6 (July 2015), p. 7787. issn: 2041-1723. DOI: 10.1038/ncomms8787. URL: http://www.nature.com/doifinder/10.1038/ncomms8787.


Chapter 5

Influence of Bilayer Electrostatics On Membrane Binding of pHLIP

5.1 Introduction

Ever since its discovery, pH-low insertion peptides\cite{16, 17} or pHLIPs have evolved as a family of pH-sensitive cell-penetrating peptides with properties uniquely suited to target the acidic microenvironment of cancer cells. pHLIP, derived from the helix C of bacteriorhodopsin, is soluble and unstructured in solution (state I), binds to surface of lipid membranes at alkaline to neutral pH (state II), and spontaneously forms a transmembrane helix under acidic conditions (state III), with an apparent pK of insertion in 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membranes at \( \sim 6.1 \)\cite{35}. Protonation of two acidic residues (D14 and D25) at low pH causes the membrane-spanning region to become more hydrophobic, which is followed by folding and membrane-insertion of pHLIP\cite{35, 1}. Two native tryptophan residues (W9 and W15) are usually used as a fluorescence probe to study the membrane interactions of pHLIP, as tryptophan fluorescence is known to be sensitive to the hydrophobicity of its environment.

A significant number of experimental\cite{2, 24, 38, 14} as well as computational\cite{9, 44, 10} studies have been targeted at studying the biophysical properties of pHLIP. Most of these have used a zwitterionic homogeneous bilayer composed of POPC. However, eukaryotic cell membranes are much more complex\cite{18}. The first exploration of the
influence of bilayer properties on pHLIP insertion was published in 2008\cite{barrera2008}. Barrera et. al\cite{barrera2008} has studied the effect of membrane fluidity on pHLIP insertion. More recently, Karabadzhak et. al\cite{karabadzhak2015} has determined the influence of membrane curvature, thickness, and fluidity on the pHLIP-binding and the kinetics of pHLIP-insertion.

One aspect that has been largely ignored by the pHLIP community is the effect of bilayer electrostatics. It is known that cancerous cells typically contain higher levels of 1-palmitoyl-2-oleoyl-sn glycero-3-phospho-L-serine (POPS) in their outer leaflet\cite{yang2010}, which is in contrast to healthy mammalian cells\cite{garcia2003}. The problem becomes all the more interesting given pHLIP itself is an atypical cell-penetrating peptide in that it has an overall negative charge on its surface. Thus, a detailed investigation of the effect of bilayer electrostatics on pHLIP-binding will have far-fetching impact on the development of pHLIP into a cancer-targeting agent.

In 2015, Scott et. al\cite{scott2015} and Kyrychenko et. al\cite{kyrychenko2015} independently studied the influence of bilayer electrostatics and arrived at seemingly contradicting results. While the former work noted negligible difference in the free energy of binding upon addition of POPS lipids, the latter concluded an elimination of state II based on the environment of the tryptophans. Recently, Vasquez-Montes et. al\cite{vasquez2015} has reported a shallower binding of pHLIP in the presence of anionic headgroups, something that was hinted at by Scott et. al. The tryptophans in this shallow bound state (referred to as State II$^a$ by the authors) exist in a aqueous environment, whose spectra resemble that of state I of pHLIP. To confirm this, we have employed atomistic molecular dynamics (MD) simulation to study state II of pHLIP in the presence and absence of POPS. Our results show behavior consistent with experiments and provides a possible explanation for the elimination of characteristic state II$^a$ spectra in state II$^a$.

### 5.2 Choice of MD simulation technique

A major challenge in atomistic MD simulation is achieving biological timescales. Even with the latest advances in computational resources, simulating a pHLIP-lipid system for milliseconds timescale would be prohibitive. Earlier we have performed microsecond and high nanosecond timescale simulations of pHLIP in implicit solvent\cite{goll2008} and
in presence of lipids\cite{10}. However, the latter “brute-force” method does not always ensure good sampling. Advanced sampling methods in MD simulation allow for faster convergence of such simulations. Many common techniques like umbrella sampling\cite{41,26}, adaptive biasing\cite{7,8} and metadynamics\cite{28,5} requires the definition of a reaction coordinate. Unfortunately, defining such a variable is challenging for pHLIP for two primary reasons:

1. pHLIP is mostly unstructured both before and after binding
2. Experimental data on pHLIP state II is extremely limited.

One common technique that bypasses the requirement of defining a reaction coordinate is accelerated molecular dynamics (aMD)\cite{12} and its variant gaussian accelerated molecular dynamics (GaMD)\cite{33}. As mentioned in chapter 4, accelerated MD is not suitable for studying a weakly bound complex like pHLIP state II.

Another popular advanced sampling technique is replica exchange molecular dynamics (REMD)\cite{40}. In this technique, multiple replicates of the system is simulated at different temperatures while coordinates are exchanged between neighboring replica at regular intervals. High-temperature replicates, having higher kinetic energy, overcomes energy barriers more efficiently than the target (room-temperature) replicate. The coordinate exchange means that the target replica is seeded at regular intervals by structures that have visited high-energy points on the energy landscape, ensuring faster convergence. One drawback of this method is the huge amount of computational resources required to simulated all the replica.

Liu et. al\cite{30}, proposed replica exchange with solute tempering (REST), wherein only a part of the system (“solute”) experiences higher temperature. In practice, all REST simulations are performed at room temperature; the solute-solvent and solvent-solvent interaction energies are scaled so as to mimic higher temperature and enable barrier crossing. An improved version of REST, named REST2, has been proposed\cite{45}, and has been implemented\cite{19} in NAMD 2.10. This NAMD implementation has been used to study peptide-membrane interactions\cite{39} and demonstrated that it could successfully reproduce previous results obtained with REMD\cite{31}. In this
work, we have used this as well as a pre-release version of the NAMD 2.12 implemen-
tation of REST2.

5.3 Methods

5.3.1 System setup

pHLIP was taken from helix C of bacteriorhodopsin (residues 73–107 of protein data
bank (PDB) 2NTU), and GLY73 was mutated to ALA. This results in pHLIP 2–36,
as in Karabadzhak et al. (referred to as pHLIP-4)[24]. The peptide was then solvated
and ionized using visual molecular dynamics (VMD)[15], with the CHARMM36 pro-
tein force field[4] used for heating simulations. The peptide was gradually heated to
700 K over 20 ps, followed by 980 ps of production to denature it from the helical
conformation. A Langevin thermostat was used to maintain constant temperature,
and heating simulation was carried out using NAMD2.9[34]. It was verified that the
heating did not lead to any cis conformations of ω in the peptide backbone.

APOPC bilayer of 200 lipids was prepared using the membrane builder in CHARMM-
GUI[21, 6, 29, 46, 22, 20] and equilibrated in the NPT ensemble (T = 310 K, P = 1
atm) for 50 ns using NAMD 2.10 and the c36 lipid force field[25]. A 2.0 fs timestep,
with a force-based switching function for Lennard-Jones interactions from 10 to 12
Å, Langevin thermostat, Langevin barostat, and a flexible cell with constant ratio
were used. Similar procedure was followed to generate bilayers comprising of 25% and
50% POPS lipid.

pHLIP was then placed close to the bilayer surface (pHLIP COM ~20 from the
surface of phosphorus atoms). This allows us to observe the binding event as well
as maintain bilayer symmetry (same number of lipid molecules in both the leaflets).
For POPC-only system, we added 5 Na+ ions to neutralize the system as (completely
deprotonated) pHLIP has a surface charge of -5. For POPS-containing systems,
additional Na+ ions were added to make the system electrically neutral. All analysis
was performed using VMD[15], LOOS[36], and home-built tcl and python scripts.
5.3.2 Replica Exchange with Solute Tempering (REST2)

We initially followed the protocol of Smith et. al[39], who simulated Aβ peptides which were of similar size as pHLIP. We used 8 replicas to scan a temperature range of 310 K to 430 K (note that Smith et. al[39] used 330 K was their target temperature as they were using DMPC bilayer), with exchange attempted every 2 ps. Simulation was performed using NAMD 2.10 for 90 ns, and we observed stable pHLIP binding (Figure 5.2). However, the m(T) values (see below) indicated poor exchange rate of the target replica (Figure 5.1). We increased the highest temperature to 500 K, increased number of replica up to 16, and varied the frequency of attempting exchange from every 2 ps to every 10 ps (see Figures 5.3, 5.4, 5.5, 5.6, 5.7, 5.8). Each gave better exchange rates of the target replica, however, stable pHLIP binding was completely abolished. We speculate that higher temperatures ( > 430 K) might be detrimental to the stability of pHLIP-membrane complex. We eventually used 10 replicas to scan the temperature range of 310 K to 430 K with exchange attempted every 2 ps. These simulations were performed in NAMD 2.12 with a pre-release version of the REST2 implementation. This gave us good exchange rates as well as stable pHLIP binding.

5.3.3 Quantifying efficiency of exchange

We quantified efficiency of exchange of our simulation by two factors. One, we looked at the random walk of replicas over temperatures, as done by Smith et. al[39]. Second, we looked at the mixing parameter m(T) defined by[13] and used by Smith et. al[39]. m(T) is given by:

\[
m(T) = 1 - \sqrt{\frac{\sum_{r=1}^{R} t_r^2}{\sum_{r=1}^{R} t_r}}
\]

(5.1)

where R is the number of replicas and t_r is the temperature of the r-th replica.
Chapter 5. Influence of Bilayer Electrostatics On Membrane Binding of pHLIP

Figure 5.1: Exchange efficiency of REST2 simulation involving 8 replica spanning 310 K to 430 K.

(A) Random walk of replicas over temperatures. Each row represents a temperature, while each color represents a replica (x-axis represents simulation time in ns). Target replica (310 K) rarely exchanges, and is mostly restricted to the first three replica. (B) \( m(T) \), as calculated by equation 5.1. The green line denotes the theoretical value of \( m(T) \) if all replicas are equally distributed across all temperatures (given by \( 1 - 1/\sqrt{R} \)). While all replicas had lower values than this value, the target replica showed exceptionally poor mixing, characterized by significantly lower \( m(T) \).

Figure 5.2: pHLIP binds POPC that is stable over 90 ns REST2 simulation

Per-residue distance of pHLIP from the surface of the POPC bilayer as a function of time. Red regions denote close proximity of a given residue of pHLIP to the bilayer surface. pHLIP binds POPC primarily through its N-terminus, consistent with experiment\[38\]. Contact is established within the first 10 ns and the bound complex is stable across 90 ns of REST2 simulation.
Chapter 5. Influence of Bilayer Electrostatics On Membrane Binding of pHLIP

Figure 5.3: Exchange efficiency of REST2 simulation involving 16 replica spanning 310 K to 500 K.

(A) Random walk of replicas over temperatures. Each row represents a temperature, while each color represents a replica (x-axis represents simulation time in ns). Replicas are evenly distributed over all temperatures. (B) $m(T)$, as calculated by equation 5.1. The green line denotes the theoretical value of $m(T)$ if all replicas are equally distributed across all temperatures (given by $1 - 1/\sqrt{R}$). While still deviating from ideal value, the general mixing is much better than with 8 replica, spanning 310 K to 430 K, as seen in Figure 5.1 B.

Figure 5.4: POPC-binding of pHLIP is abolished with 16 replicas spanning 310 K to 500 K.

Per-residue distance of pHLIP from the surface of the POPC bilayer as a function of time. Red regions denote close proximity of a given residue of pHLIP to the bilayer surface. pHLIP binding is abolished under these simulation conditions. pHLIP does not establish stable contact with the lipid surface in 40 ns REST2 simulation.
Chapter 5. Influence of Bilayer Electrostatics On Membrane Binding of pHLIP

Figure 5.5: Exchange efficiency of REST2 simulation involving 16 replica spanning 310 K to 500 K and exchange attempt every 10 ps.

(A) Random walk of replicas over temperatures. Each row represents a temperature, while each color represents a replica (x-axis represents simulation time in ns). Replicas are evenly distributed over all temperatures. (B) $m(T)$, as calculated by equation 5.1. The green line denotes the theoretical value of $m(T)$ if all replicas are equally distributed across all temperatures (given by $1 - 1/\sqrt{R}$). While still deviating from ideal value, the general mixing is much better than with 8 replica, spanning 310 K to 430 K (Figure 5.1 B) and 16 replica spanning 310 K to 500 K with exchange attempt every 2 ps (Figure 5.3 B).

Figure 5.6: POPC-binding of pHLIP is abolished with 16 replicas spanning 310 K to 500 K with reduced frequency of exchange attempt.

Per-residue distance of pHLIP from the surface of the POPC bilayer as a function of time. Red regions denote close proximity of a given residue of pHLIP to the bilayer surface. pHLIP binding is abolished under these simulation conditions. pHLIP does not establish stable contact with the lipid surface in 40 ns REST2 simulation.
Figure 5.7: Exchange efficiency of REST2 simulation involving 12 replica spanning 310 K to 500 K.

(A) Random walk of replicas over temperatures. Each row represents a temperature, while each color represents a replica (x-axis represents simulation time in ns). Replicas are evenly distributed over all temperatures. (B) m(T), as calculated by equation 5.1. The green line denotes the theoretical value of m(T) if all replicas are equally distributed across all temperatures (given by 1 - 1/R). Mixing is much better than with 8 replicas, spanning 310 K to 430 K (Figure 5.1 B). However, m(T) of the target replica deviates from the ideal value much more than with 16 replica spanning 310 K to 500 K with exchange attempt every 2 ps (Figure 5.5 B).

Figure 5.8: POPC-binding of pHLIP is abolished with 12 replicas spanning 310 K to 500 K.

Per-residue distance of pHLIP from the surface of the POPC bilayer as a function of time. Red regions denote close proximity of a given residue of pHLIP to the bilayer surface. pHLIP binding is abolished under these simulation conditions. pHLIP briefly establishes contact with the lipid in the first 20 ns but the complex rapidly dissociates.
Chapter 5. Influence of Bilayer Electrostatics On Membrane Binding of pHLIP 119

5.4 Results

5.4.1 Initial attempts

We initially performed REST2 simulations with 8 replica scanning 310 K to 430 K. We obtained stable binding of pHLIP to the POPC bilayer (Figure 5.2). pHLIP establishes contact with the bilayer within the first 10 ns, and the bound complex stays stable across 90 ns of REST2 simulation. Additionally, we find that pHLIP binds primarily via its N-terminus, consistent with experiment[38]. This simulation had exchange rates of ~20%. However, as can be seen, m(T) of the target replica is significantly lower than all other replicates, and also lower than the theoretical limit. The random walk also shows that the target replica has rarely exchanged with the others. This is an example of poor exchange. To improve sampling, we increased the highest temperature to 500 K and the number of replica to 16. This gave much better m(T) values more even random walk (Figure 5.3). On the other hand, the binding of pHLIP to the POPC bilayer was completely abolished (Figure 5.4), and pHLIP did not establish contact with the bilayer surface in 40 ns REST2 simulation.

One possibility was that our exchange attempt (every 2 ps) was too frequent, and the peptide did not get sufficient time to interact with the bilayer before being exchanged. To confirm this, we repeated the simulation with exchange attempt every 10 ps. This gave an even random walk and m(T) behavior was even better than with 2 ps exchange rage (Figure 5.5). Lipid-binding of pHLIP, though, was still absent (Figure 5.6).

Next, we reduced the number of replicas from 16 to 12. The rationale was similar: this would reduce the number of exchanges, effectively allowing more time to the replica to find the bound complex. Expectedly, the m(T) behavior was slightly worse, but random walk was still even (Figure 5.7). No improvement in lipid-binding of pHLIP was observed. We speculate that temperature as high as 500 K was causing the pHLIP-POPC complex to dissociate. Thus, we used the temperature range of 310 K to 430 K, and used 10 replicas to improve exchange (see below).
5.4.2 Binding of pHLIP to 100% POPC bilayer

As stated above, we used 10 replicas to scan a temperature range of 310 K to 430 K, for 50 ns. This setting gave ~28-36% exchange rates. The m(T) values (Figure 5.9 B) are in fact little better than that obtained scanning 310 K to 500 K using 12 replicas (Figure 5.7 B). The temperature random walk of replicas (Figure 5.9 A) is also reasonably uniform. All of this indicate good mixing of replicas in the 50 ns REST2 simulation.

We then looked at the binding efficiency of pHLIP to the POPC bilayer (Figure 5.10). While the binding was not as stable as obtained with 8 replica (Figure 5.2), pHLIP does establish contact with the bilayer. The efficient exchange meant that this replica regularly exchanged coordinates with high-temperature replica in which pHLIP was far away from the bilayer, giving the intermediate blue regions. However, each such exchange was quickly followed by re-establishment of contact. Thus, we have observed unstructured pHLIP in solution (state I) bind to a POPC bilayer (state II) in a 50 ns REST2 simulation.

Additionally, we see that pHLIP interacts with the bilayer primarily through its N-terminus. This is consistent with experiments\cite{38} which showed that in state II, N-terminus of pHLIP is less flexible and the C-terminus. Intuitively, this makes sense as the anionic C-terminus of pHLIP is expected to have unfavorable interactions with the lipid headgroup. We, however, do see some interactions of the C-terminus with the bilayer as well. This small discrepancy might be due to the difference in lipid:peptide ratio used in the experiments\cite{38}. While they had 75 lipids per peptide, the simulation had 200 POPC lipids and one pHLIP molecule. Thus, the experimental set up might have seen some “parking problem” described by Reshetnyak et. al\cite{35}. Here, since more lipid surface area was available to pHLIP, it might have caused increased interactions with the C-terminus in addition to the primary N-terminal mediated binding.
Figure 5.9: \textbf{Exchange efficiency of REST2 simulation involving 10 replica spanning 310 K to 430 K.}

(A) Random walk of replicas over temperatures. Each row represents a temperature, while each color represents a replica (x-axis represents simulation time in ns). Replicas are evenly distributed over all temperatures. (B) $m(T)$, as calculated by equation 5.1. The green line denotes the theoretical value of $m(T)$ if all replicas are equally distributed across all temperatures (given by $1 - 1/\sqrt{R}$). $m(T)$ values are close to their ideal value, denoting good mixing.

Figure 5.10: \textbf{pHLIP establishes stable contact with POPC in spite of frequent exchange.}

Per-residue distance of pHLIP from the surface of the POPC bilayer as a function of time. Red regions denote close proximity of a given residue of pHLIP to the bilayer surface. pHLIP establishes contact with the POPC bilayer through its N-terminus, consistent with experiments\cite{38}. Blue regions denote exchange with high-temperature replica where pHLIP was away from the lipid surface. Each such exchange is followed by re-establishment of contact.
Chapter 5. Influence of Bilayer Electrostatics On Membrane Binding of pHLIP

5.4.3 Presence of POPS: bilayer electrostatics

Next, we looked at the influence of bilayer electrostatics on membrane-binding of pHLIP. As mentioned earlier, cancer cells often contain elevated levels of POPS in their outer leaflet\[42\]. To develop pHLIP into a cancer-targeting agent, it is important to understand how its membrane-binding is modified in the presence of POPS. POPS contains anionic headgroups, resulting in altered electrostatics at the bilayer surface. Being an anionic peptide, pHLIP is expected to have less favorable interaction with bilayers containing negatively charged headgroups. Effect of bilayer electrostatics on membrane-binding of pHLIP has been experimentally studied\[37, 27, 43\]. Interestingly, free energy of binding was found to be essentially unchanged upon addition of POPS\[37, 43\]. What is more, fluorescence spectra characteristic of state II is eliminated in the presence of POPS, leading the authors to indicate at a shallow-binding state II* state\[27, 43\].

We performed 50 ns REST2 simulation of pHLIP with bilayer containing 25% and 50% POPS. POPS content was chosen to match that used by Kyrychenko et. al\[27\]. and Vasquez-Montes et. al\[43\]. Note that Scott et. al\[37\]. used lower POPS content (\(<= 30\%\)) in their study. As a first check, we again looked at the mixing efficiencies of these two simulations (Figures 5.11 and 5.12). In both the cases, we find even random walk of replicas over the temperatures, and m(T) close to the ideal value.

Then, we looked at the efficiency of membrane-binding of pHLIP. Figure 5.13 shows that in a membrane containing 25% POPS, C-terminus of pHLIP is now farther away from the bilayer surface compared to 100% POPC. This is expected as 4 of the 6 anionic residues are near its C-terminus, making it the anionic end. Interactions of this end with the bilayer surface is hindered by the presence of the negatively charged phosphoserine headgroups. However, the N-terminus of pHLIP still establishes contact with the bilayer. This is consistent with experiments which show binding free energy of pHLIP is unchanged in the presence of POPS\[37, 43\]. Overall, we see a shallow binding as was originally hinted at by Scott et. al\[37\]. and later confirmed by Vasquez-Montes et. al\[43\]. When the POPS content is increased to 50% (Figure 5.14), C-terminus still maintains distance from the bilayer surface as expected. Surprisingly, the N-terminus seems to bind the membrane a bit more
Figure 5.11: Exchange efficiency of REST2 simulation of pHLIP in a bilayer of 75% POPC and 25% POPS.

(A) Random walk of replicas over the temperatures. Each row represents a temperature, while each color represents a replica (x-axis represents simulation time in ns). Replicas are evenly distributed over all temperatures. (B) \( m(T) \), as calculated by equation 5.1. The green line denotes the theoretical value of \( m(T) \) if all replicas are equally distributed across all temperatures (given by \( 1 - 1/\sqrt{R} \)). \( m(T) \) values are close to their ideal value, denoting good mixing.

Figure 5.12: Exchange efficiency of REST2 simulation of pHLIP in a bilayer of 50% POPC and 50% POPS.

(A) Random walk of replicas over temperatures. Each row represents a temperature, while each color represents a replica (x-axis represents simulation time in ns). Replicas are evenly distributed over all temperatures. (B) \( m(T) \), as calculated by equation 5.1. The green line denotes the theoretical value of \( m(T) \) if all replicas are equally distributed across all temperatures (given by \( 1 - 1/\sqrt{R} \)). \( m(T) \) values are close to their ideal value, denoting good mixing.
Chapter 5. Influence of Bilayer Electrostatics On Membrane Binding of pHLIP 124

Figure 5.13: Anionic C-terminus of pHLIP is away from the surface of bilayer containing 25% POPS.

Per-residue distance of pHLIP from the surface of the 75% POPC, 25% POPS bilayer as a function of time. The anionic C-terminus of pHLIP is now away from the bilayer surface. However, pHLIP still establishes contact with the bilayer, consistent with experiments[37, 43].

Figure 5.14: N-terminus of pHLIP establishes contact even with bilayer containing 50% POPS.

Per-residue distance of pHLIP from the surface of the 50% POPC, 50% POPS bilayer as a function of time. Even when PS-content increases to 50%, pHLIP still maintains contact with the bilayer. Interestingly, binding seems to be stronger than with 25% POPS (Figure 5.13). While this could be due to the limited timescale of our simulation, we speculate the stronger binding might be due to the favorable interaction between the cationic R11 with the bilayer containing higher proportion of anionic headgroups.
stably than with 25% POPS. While this might be due to the limited timescale (50 ns) of our simulations, we speculate this is due to more favorable interactions of the positively charged N-terminus with an anionic bilayer.

5.4.4 Elimination of state II fluorescence spectra

![Figure 5.15: A subpopulation where W9 is buried and dehydrated is eliminated in the presence of POPS](image)

2D probability distribution of W9 burial depth and SASA of W9 in (A) 100% POPC, (B) 25% POPS, and (C) 50% POPS. In 100% POPC, there is a heavily populated region where W9 is buried and dehydrated (low SASA). This subpopulation is eliminated in the POPS-containing systems. We speculate this subpopulation is the origin of the characteristic Trp-fluorescence spectra of pHLIP in state II. Such spectra is has not been observed in state II of pHLIP in bilayer containing POPS[27, 43].
We then asked the question, why is characteristic Trp fluorescence of state II eliminated\cite{27,43} in the presence of POPS? The recent work by Vasquez-Montes et. al\cite{43}, has hinted at a shallow binding where the tryptophans of pHLIP are in an essentially aqueous environment, leading to state I-like fluorescence. To verify this, we looked at two quantities: (1) burial of Trp aromatic rings into the bilayer, and (2) solvent accessible surface area (SASA) of the Trp residues. SASA reports on the degree of dehydration of residues (lower SASA means higher dehydration). Since the change from state I to state II spectra of pHLIP is related to the Trp aromatic sidechain experiencing a more hydrophobic environment, we expect to see Trp burial accompanied by Trp-dehydration (low SASA).

![Figure 5.16: W15 does not show a subpopulation where it is buried and dehydrated even in 100% POPC.](image)

2D probability distribution of W15 burial depth and SASA of W9 in (A) 100% POPC, (B) 25% POPS, and (C) 50% POPS. Unlike W9 (Figure 5.15), W15 does not show a subpopulation where it is buried and dehydrated. We speculate this is due to the limited time scale (50 ns) of our simulation.
Figure 5.15 shows the 2D probability of the depth of burial of W9 aromatic sidechain into the bilayer and the SASA of W9. In 100% POPC, there is a heavily populated region where Trp is buried into the membrane and is dehydrated. This subpopulation is entirely missing in the POPS-containing systems. We speculate that this population is responsible for the state II Trp fluorescence spectra, where the aromatic ring of Trp is buried into the bilayer and is desolvated.

Interestingly, we did not observe a similar behavior for W15 (Figure 5.16). This might be due to the limited timescale of our simulations, and we expect that at longer timescales, W15 would also explore this subpopulation in 100% POPC bilayer.

5.4.5 Ongoing work

We are currently exploring the influence of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) lipid on the membrane-binding of pHLIP. Kyrychenko et. al[27], and Vasquez-Montes et. al[43], have studied POPC-POPE systems in terms of pHLIP binding, and observed elimination of characteristic state II fluorescence spectra, similar to POPS-containing systems. They also observed, from quenching studies, that the environment of the Trp residues of pHLIP was even more aqueous in 25% POPE system than 25% POPS system. Since POPE is a zwitterionic lipid, repulsion from anionic headgroup is not a factor here. The authors ascribe the shallower binding to differences in bilayer properties upon addition of POPE.

To investigate this, we performed 50 ns REST2 simulations of pHLIP with a bilayer containing 25% and 50% POPE. System was prepared similar to the POPS-containing systems, as described in section 5.3. 5 neutralizing Na\(^+\) ions were added to neutralize the system. Similar to the previous systems we studied, POPE-containing systems also showed a even random walk of the replicas and m(T) close to the theoretical value (Figures 5.17 and 5.18). However, contrary to expectation, pHLIP seemed to bind better to the POPE-containing membranes (Figures 5.19 and 5.20) than the POPS-containing membranes (Figures 5.13 and 5.14). There is also a discrepancy in terms of the shallower binding of pHLIP in POPE-containing bilayers that has been observed in experiments[27, 43]. Although W9 showed expected behavior - removal
Figure 5.17: Exchange efficiency of REST2 simulation of pHLIP in a bilayer of 75% POPC and 25% POPE.

(A) Random walk of replicas over the temperatures. Each row represents a temperature, while each color represents a replica (x-axis represents simulation time in ns). Replicas are evenly distributed over all temperatures. (B) $m(T)$, as calculated by equation 5.1. The green line denotes the theoretical value of $m(T)$ if all replicas are equally distributed across all temperatures (given by $1 - 1/\sqrt{R}$). $m(T)$ values are close to their ideal value, denoting good mixing.

Figure 5.18: Exchange efficiency of REST2 simulation of pHLIP in a bilayer of 75% POPC and 25% POPE.

(A) Random walk of replicas over the temperatures. Each row represents a temperature, while each color represents a replica (x-axis represents simulation time in ns). Replicas are evenly distributed over all temperatures. (B) $m(T)$, as calculated by equation 5.1. The green line denotes the theoretical value of $m(T)$ if all replicas are equally distributed across all temperatures (given by $1 - 1/\sqrt{R}$). $m(T)$ values are generally close to their ideal value, while the target replica has a slightly reduced $m(T)$.
Per-residue distance of pHLIP from the surface of the 75% POPC, 25% POPE bilayer as a function of time. Red regions denote close proximity of a given residue of pHLIP to the bilayer surface. Contrary to expectation, we found that pHLIP bound better to a bilayer containing 25% POPE compared to 25% POPS. We are currently unsure of the reason behind this observation. Interestingly, pHLIP seems to establish contact with the bilayer through its middle segment as opposed to the N-terminus.

Per-residue distance of pHLIP from the surface of the 50% POPC, 50% POPE bilayer as a function of time. The results were even more unexpected, wherein we see increasing POPE-content improving pHLIP binding. We are currently unsure of the reasons behind this observation.
Figure 5.21: A subpopulation where W9 is buried and dehydrated is eliminated in the presence of POPE

2D probability distribution of W9 burial depth and SASA of W9 in (A) 100% POPC, (B) 25% POPE, and (C) 50% POPE. In 100% POPC, there is a heavily populated region where W9 is buried and dehydrated (low SASA). This subpopulation is eliminated in the POPE-containing systems. We speculate this subpopulation is the origin of the characteristic Trp-fluorescence spectra of pHLIP in state II. Such spectra has not been observed in state II of pHLIP in bilayer containing POPE[27, 43].

of the subpopulation putatively responsible for state II characteristic fluorescence (Figure 5.21) - W15 now shows a strong presence of a similar subpopulation where it is buried and dehydrated (Figure 5.22). At this point we are not sure of the possible reasons for this, and are currently looking into it. While it could be due to the limited (50 ns) timescale of our simulations, we here note that the experiments[27, 43] used a higher lipid:peptide ratio (500:1) compared to our simulations, which had 200 lipids and a single pHLIP molecule.
Figure 5.22: A subpopulation where W9 is buried and dehydrated is eliminated in the presence of POPE.

2D probability distribution of W9 burial depth and SASA of W15 in (A) 100% POPC, (B) 25% POPE, and (C) 50% POPE. Interestingly, W15 in 50% POPE bilayer shows a subpopulation putatively responsible for the characteristic state II Trp fluorescence - one where it is buried and dehydrated. This is in disagreement with experiments that reported elimination of state II Trp fluorescence spectra in POPE-containing bilayers[27, 43]. While this might be due to the limited timescale (50 ns) of our simulations, we speculate this is possibly due to the difference in the lipid:peptide ratio. While Kryrychenko et. al[27]. and Vasques-Montes et. al[43]. used 500 lipids per peptide, our simulation system had 200 lipids and a single pHLIP peptide.
References


REFERENCES


REFERENCES


REFERENCES


[38] Nicolas S. Shu et al. “Residue-specific structures and membrane locations of pH-low insertion peptide by solid-state nuclear magnetic resonance”. In: *Nature Communications* 6 (July 2015), p. 7787. ISSN: 2041-1723. DOI: 10.1038/ncomms8787. URL: http://www.nature.com/doifinder/10.1038/ncomms8787.


