Using enhanced sampling molecular dynamics to probe the binding process of the membrane insertion peptide pHLIP

Yue Ren

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

Recommended Citation

Ren, Yue, "Using enhanced sampling molecular dynamics to probe the binding process of the membrane insertion peptide pHLIP" (2015). Graduate Theses, Dissertations, and Problem Reports. 6492.
https://researchrepository.wvu.edu/etd/6492

This Thesis is brought to you for free and open access by The Research Repository @ WVU. It has been accepted for inclusion in Graduate Theses, Dissertations, and Problem Reports by an authorized administrator of The Research Repository @ WVU. For more information, please contact ian.harmon@mail.wvu.edu.
Using enhanced sampling molecular dynamics to probe the binding process of the membrane insertion peptide pHLIP

Yue Ren

Thesis submitted
to the Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Master of Science in
Chemistry/Physical Chemistry

Blake Mertz, Ph.D., Chair
Justin Legleiter, Ph.D.
Stephen Valentine, Ph.D.
Department of Chemistry

Morgantown, West Virginia
2015

Keywords: pHLIP peptide, membrane binding, pH-dependent insertion, drug delivery, molecular dynamics simulation, umbrella sampling, free energy, lipids distortion

Copyright 2015 Yue Ren
ABSTRACT

Using enhanced sampling molecular dynamics to probe the binding process of the membrane insertion peptide pHLIP

Yue Ren

Peptides with the ability to bind and insert into the cell membrane are an ever-growing field of research due to their potential biomedical applications. pH (Low) Insertion Peptide (pHLIP), which is a water-soluble polypeptide derived from helix C of bacteriorhodopsin, has the ability to insert into a membrane at acidic pH to form a stable transmembrane α-helix. The insertion process takes place in three stages: pHLIP is unstructured and soluble in water at neutral pH (state I), unstructured and bound to the surface of a membrane at neutral pH (state II), and inserted into the membrane as an α-helix at low pH (state III). It has been shown that pHLIP binding and insertion occurs over large timescales and multiple kinetic steps from state I to state III. Our study focuses on the initial step, uncoiled pHLIP binding to a lipid bilayer surface, using enhanced sampling molecular dynamics simulation techniques. We have quantified the thermodynamics of this process by computing the free energy change upon binding of a pHLIP variant at several orientations to model lipid bilayers. In addition, our studies provide atomistic details about the binding behaviors of pHLIP to a lipid bilayer that provide a fundamental understanding of the biophysical underpinnings of the pHLIP mechanism.
INTRODUCTION

Peptides with the ability to bind to and insert into the cell membrane (both prokaryotes and eukaryotes) have recently attracted increased interest due to their potential biomedical applications. Most constitutive membrane proteins fold and insert into the cell membrane with the assistance of complex molecular machines. However, some non-constitutive membrane proteins, such as toxins, antimicrobial peptides, and WALP peptides can spontaneously fold and insert across the cell membrane (London and Shahidullah, 2009). The insertion process of these peptides is accompanied by a release of energy, which can be utilized to translocate cargo molecules across the membrane. Therefore these polypeptides have the potential to be developed as a novel class of delivery agents (Andreev et al., 2010).

pH low insertion peptide (pHLIP) is a water-soluble peptide derived from transmembrane helix C of bacteriorhodopsin (Fig. 1). Peptide insertion is triggered by a pH shift from an alkaline to an acidic environment (Hunt et al., 1997), and can be tracked using several spectroscopic techniques. Circular dichroism (CD) spectroscopy can determine the secondary structure of proteins, due to the fact that the electronic transitions of peptide bonds in different conformations (e.g., α-helical or β-sheet) produce distinct absorption spectra for left- and right-handed circularly polarized light. Tryptophan fluorescence is another spectroscopic technique that can estimate the local structure and dynamics of a polypeptide, taking advantage of the characteristic that tryptophan is sensitive to the polarity of its local environment. For example, tryptophan fluorescence can be used to investigate the translocation of a peptide from a completely solvated environment to a transmembrane environment in the bilayer. In the initial study by Hunt et al., these two spectroscopic techniques were used to investigate pHLIP, finding that spontaneous insertion of pHLIP involves a coupling between the peptide structure and the
protonation state of two aspartic acid residues in the membrane-spanning region of the peptide. This can be explained by the three states of the membrane insertion of pHLIP: (1) unstructured and soluble in water at neutral pH; (2) unstructured and bound to the surface of the membrane at neutral pH; (3) inserted across the membrane as an α-helix at lower pH (Fig.2). Based on a pH titration analysis, they identified that the protonation of only one of the interior aspartic acid residues induces the change from an uncoiled to an α-helical conformation. In addition, polarized FTIR of macroscopically-ordered peptide/phospholipid multilayers was used to investigate the orientation of pHLIP in membranes, finding that it possesses a helical tilt of 15° with respect to the membrane normal.

**Figure 1. Primary and tertiary structure of pHLIP membrane insertion peptide.** (A) The primary sequence of pHLIP and pHLIP-1. (B) Three-dimensional representation of pHLIP-1. *Green ribbon:* peptide backbone; *blue:* acidic residues; *orange:* tryptophan residues.

pHLIP has demonstrated significant differences from other conventional vehicles used as drug delivery agents. In a study using rat antigen-induced arthritis as a model, it was shown that pHLIP can preferentially accumulate in inflammatory foci, and that it also has the ability to image the renal cortical interstitium of the kidney (Andreev et al., 2007). More recently pHLIP was shown to localize to breast cancer tumors in mice using a self-quenching fluorophore that
dequenches upon insertion into the membrane and disulfide cleavage in the cytoplasm (Karabadzhak et al., 2014). Furthermore, pHLIP can selectively translocate therapeutic polar cargoes into the cell cytoplasm (Wijesinghe et al., 2011), a distinct ability from other peptide-based drug delivery methods. Thus pHLIP technology can be widely used as a novel method in detecting, targeting, and treating acidic diseased tissue by translocating typically membrane-impermeable cargoes such as bulky and/or polar molecules across the cell membrane (Deacon 2015).

The insertion process is associated with the protonation of internal Asp residues (Asp14 and Asp25), which leads to an increase of hydrophobicity and triggers the folding and insertion of the peptide across the lipid bilayer (Andreev et al., 2007). A follow-up study characterized the

**Figure 2. pHLIP has a pH-dependent insertion mechanism.** In solution and at alkaline pH, pHLIP exists as an uncoiled peptide (State I). When lipid vesicles are added to solution, pHLIP binds to the outer surface of the vesicle (State II). A drop in pH to acidic values leads to folding of pHLIP into an α-helix and unidirectional insertion into the cell membrane (State III). Yellow: unperturbed lipids; light blue: perturbed lipids; dark blue: lipids in direct contact with pHLIP. Figure modified from Reshetnyak et al. (2008) PNAS.
thermodynamics of binding and insertion of pHLIP using florescence spectroscopy and isothermal titration calorimetry (Reshetnyak et al., 2008). When pHLIP binds to the membrane surface (State I → State II), it is accompanied by an energy release of 6–7 kcal-mol\(^{-1}\), while insertion is accompanied by an additional energy release of ~1.8–2.0 kcal-mol\(^{-1}\). Based on their binding equilibrium model of pH titration data, three distinct classes of lipids that interact with pHLIP were identified: (1) lipids interacting directly with the pHLIP peptide; (2) lipids not interacting directly with the pHLIP peptide, but perturbed by the interaction; (3) lipids not significantly perturbed by the interaction. In addition, measurement of the change in heat capacity determined that the number of protons involved in the process of pHLIP folding and insertion is 1.5–1.8, a significant deviation from what was originally proposed (Hunt et al 1997).

The kinetics of pHLIP binding, insertion, and exit are highly dependent on peptide composition. When variants of pHLIP were designed with truncated C-termini (Karabadzhak et al., 2012), it was found that an increase in protonatable groups at the C-terminus led to increased times for peptide translocation across the bilayer. For example, pHLIP-1, which has all C-terminal acidic groups removed, inserted and exited on a uniform timescale (80-400 ms) that was two orders of magnitude faster than wild-type pHLIP (30-50 s), regardless of the range of the pH jump. Also, pHLIP-2, which had two protonatable C-terminus residues, had a slower rate (3-8 s) than pHLIP-1, but faster than wild-type pHLIP. This behavior means that the presence of protonatable groups at the inserting (C-terminal) end of pHLIP have a direct effect on the timescale of the insertion and exit mechanisms. We chose pHLIP-1 for our computational studies of the State I → State II transition due to the fact that 1) it possesses faster timescales and 2) lacks titratable residues on the C-terminus of the peptide.
Molecular dynamics (MD) simulations are capable of providing atomistic details on the interactions between pHLIP and a lipid bilayer that are unattainable through experimental methods. However, MD simulations have two shortcomings with respect to modeling pHLIP: (1) titratable residues (i.e., Asp14 and Asp25) must possess a fixed ionization state, since bonds cannot be made or broken in classical MD, and (2) pH ranges cannot be explicitly modeled (see (1)). These characteristics have limited the scope of how MD has been applied to modeling pHLIP. Previously, coarse-grained MD simulations were used to characterize the potential of mean force (PMF) for the transfer of pHLIP across a lipid bilayer (Gkeka and Sarkisov, 2010). In this study it was found that the distinct shape of the PMF of pHLIP indicated a strong preference for transmembrane insertion perpendicular to the surface of the bilayer. Another study investigated the interaction of pHLIP with a POPC bilayer using two different amino acid titration states (negatively-charged and neutral acidic residues) with all-atom MD simulations (Deng et al., 2013). Those simulations were unbiased, but purely phenomenological, since they couldn’t measure quantities such as the free energy of binding as determined through calculation of a PMF. In addition, a pre-formed alpha-helix was used, so it was impossible to investigate binding of the uncoiled peptide to the bilayer surface or the process of peptide folding upon a change from alkaline to acidic pH.

Our objective is to investigate the atomistic details and thermodynamics that are implicated in pHLIP binding to a bilayer surface. Umbrella sampling (US), a biased MD simulation technique, is widely used to explore the thermodynamic properties of biophysical processes that occur infrequently under equilibrium conditions. Numerous studies have used this technique to investigate the free energy changes of binding, insertion, or permeation of small peptides similar to pHLIP (Lin and Grossfield, 2014; Tolokh et al., 2009). However, difficulties also exist in the
statistical convergence of US simulations, mainly because reorganization of the lipid bilayer that takes place due to binding/insertion can potentially occur on the microsecond timescale, which is beyond the capabilities of most computing resources. Nevertheless, it is possible to overcome this problem, either by evaluating the time-dependence of a key observable, such as the binding free energy, or conducting multiple US simulations with different initial conditions (Neale et al., 2014).

Considering that the process of pHLIP binding and insertion occurs over large timescales and multiple steps from State I to State III, we have chosen to focus on the initial step, uncoiled pHLIP binding to a lipid bilayer surface (State I → State II). We have quantified this process by computing the free energy change upon binding of pHLIP-1 over a series of orientations to model lipid bilayers. In addition, our studies provide atomistic details about the binding behavior of pHLIP-1 to a lipid bilayer that can be used to further our fundamental understanding of the pHLIP mechanism.
METHODS

pHLIP-POPC system construction

In this study, we used pHLIP-1, a C-terminal truncated version of wild-type pHLIP. The amino acid sequence of pHLIP-1 is AEDQNDPYWARYADWLFTTPLLLDLALLVG.

To get the peptide in an unstructured form, a MD simulation was conducted on the α-helical conformation of pHLIP-1, modified from helix C of the crystal structure of bacteriorhodopsin (pdb code: 2NTU). The simulation was carried out in the $NVT$ ensemble ($T = 700K$) in vacuum for 2 ns using a 2 fs time step. The CHARMM36 force field (Best et al., 2012) was used to describe the molecular interactions. After the simulation completed, we saved the final snapshot from the trajectory as our starting structure for binding simulations.

A membrane bilayer consisting of 166 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) phospholipid molecules, the most commonly used lipid in pHLIP experiments, was also used. The bilayer was constructed using the membrane builder of CHARMM-GUI (Jo et al., 2009) and brought to an equilibrated state with 50ns molecular dynamics simulation in NAMD.

The free energy of binding was estimated by modeling the process of pHLIP-1 unbinding from the POPC bilayer. This decision was based on the fact that it is difficult to maintain a well-equilibrated bilayer for the process of pHLIP-1 binding from bulk solvent. The unstructured pHLIP-1 peptide was placed ~20 Å from the center-of-mass (COM) of the bilayer. Five equally spaced orientations (0, 72, 144, 216, and 288 deg) along the $x$ axis of the $xy$ plane formed by the POPC bilayer surface were generated (Fig. 3). Next, we added 15,861 TIP3 water molecules, and 8 sodium and 4 chloride ions such that the system was electrically neutral with a
free salt concentration of 10 mM. The lipids within 0.6 Å of peptide were removed to avoid clashes between the lipids headgroups and the peptide. Each system was approximately 87,000 atoms.

**Theoretical background**

**All-atom MD simulations.** One of the principal tools in the theoretical study of biological molecules is the method of molecular dynamics. This computational method calculates the time-dependent behavior of a molecular system based on Newton's second law of motion \( F = ma \). MD simulations are able to provide detailed information on the fluctuations and conformational changes of microscopic molecular systems. These methods are now routinely used to investigate the structure, dynamics, and thermodynamics of biological molecules and their complexes. The potential energy function (also referred to as the force field) is the central term used to describe the atomic interactions. A typical atomistic force field has the form

\[
V(R) = \sum_{\text{bond}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedral}} K_\varphi(1 - \cos(n\varphi)) + \sum_{\text{vdW}} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} \right) + \sum_{\text{electrostatic}} \frac{q_i q_j}{Dr_{ij}}
\]

(1)

where each term describes bond stretching, angle bending, bond rotation, van der Waals interactions, and electrostatic interactions, respectively. The \( K \) variables are a spring constant that is applied to each interaction, \( r_{ij} \) is the distance between two atoms, \( q_{ij} \) is the partial atomic charge, \( A_{ij} \) and \( C_{ij} \) are empirically derived constants, and \( D \) is the scaled electrical permittivity of space. Each of these energy terms is empirically derived from either experimental data or quantum mechanical calculations on model molecules.
**Steered MD simulations.** Steered molecular dynamics simulation (SMD) is a technique in which the center of mass of one or more atoms is harmonically constrained with a force constant \( k \) to move with velocity \( v \) in the direction \( \vec{n} \). Thus the following additional potential is applied to the system:

\[
U (\vec{r}_1, \vec{r}_2, ..., t) = \frac{1}{2} k [vt - (\vec{R}(t) - \vec{R}_0) \cdot \vec{n}]^2,
\]  

where \( t = N_{ts} dt \), and \( N_{ts} \) is the number of elapsed timesteps in the simulation and \( dt \) is the size of the timestep in femtoseconds. Also, \( \vec{R}(t) \) is the current center of mass of the SMD atoms and \( \vec{R}_0 \) is the initial center of mass as defined by the coordinates in SMD File. The main purpose of SMD is to accelerate a biophysical process in order to study the behavior of a system under various conditions (e.g., unbinding of ligands and large-scale conformational changes of proteins) when equilibrium MD simulations are unable to adequately sample configurational space of the potential energy surface.

**Umbrella sampling (US) simulations.** The potential of mean force (PMF) along a reaction coordinate is a key property that can be obtained from statistical mechanics. In a more practical way, the PMF can be used to characterize how the free energy changes as a function of a reaction coordinate of the system. (The reaction coordinate (RC) is any process of interest (one-dimensional in our case) that can be monitored by following the evolution of the system between two defined endpoints.) However, it is impractical to compute the PMF or the distribution function directly from an unbiased molecular dynamics simulation. The reason for this is because rare biophysical events (e.g., peptide binding to a membrane surface) possess large energy barriers which are typically insurmountable in an equilibrium simulation and subsequently lead to poor sampling of the potential energy surface. Umbrella sampling (US) is a
technique in which the microscopic system of interest is simulated in the presence of an artificial biasing potential to enhance the sampling within the chosen region of the reaction coordinate (Kumar et al., 1992; Roux, 1995). Typically, the biasing potential serves to restrain the variations of the coordinate within a small interval of the reaction coordinate (i.e., the “window”), helping to achieve sufficient configurational sampling in this region (Roux, 1995). The potential energy for the biased simulation has the form,

\[ E = U(R) + W(\xi) , \]  

where the biasing potential \( W(\xi) \) has a harmonic function of the form,

\[ W_i(\xi) = \frac{1}{2} K(\xi - \xi_i)^2, \]  

where \( \xi_i \) is the center of the reaction coordinate for each window.

**The weighted histogram analysis method (WHAM).** Upon completion of the umbrella sampling simulations, it is necessary to transform the data points of coordinates to PMF values. The weighted histogram analysis method (WHAM) is now routinely used to calculate the PMF along a reaction coordinate (Kumar et al., 1992). The central idea consists in constructing an optimal estimate of the unbiased distribution function as a weighted sum over the data extracted from all the US simulations and determining the functional form of the weighted factors that minimizes the statistical error. The WHAM equation can be summarized as,

\[
\langle \rho(\xi) \rangle = \sum_{i=1}^{N} n_i \langle \rho(\xi) \rangle_{(i)} \times \left[ \sum_{j=1}^{N} n_j e^{-[W_j(\xi) - F_j]/k_B T} \right]^{-1},
\]

where the reaction coordinate \( \xi \) is a function of the reaction coordinates, \( \rho(\xi) \) is computed as a normalized histogram of the values of \( \xi \) occurring during the simulations, \( n_i \) is the number of
independent data points used to construct the biased distribution function, and \( N_w \) is the number of biased distribution functions. \( W(\xi) \) is the biasing potential applied to each window, and the free energy constants \( F_j \) are determined using the optimal estimate of the distribution function as,

\[
e^{-F_i/k_BT} = \int d\xi \ e^{-W_i(\xi)/k_BT} \langle \rho(\xi) \rangle.
\] (6)

The last step is to calculate the PMF from the average distribution function of \( \rho(\xi) \) using,

\[
W(\xi) = W(\xi^*) - k_BT \ln \left[ \frac{\langle \rho(\xi) \rangle}{\langle \rho(\xi^*) \rangle} \right],
\] (7)

where \( \xi^* \) and \( W(\xi^*) \) are arbitrary constants based on the reference state we want to compare with (Grossfield, 2011; Kumar et al., 1992; Roux, 1995).

**Simulation details**

*Steered molecular dynamics simulations.* The system was equilibrated for 2 ns and the pulling force was applied to the peptide away from the bilayer surface on the \( xy \) plane. The process was achieved by conducting a steered MD simulation with a force along the \( z \) coordinate applied to all atoms of the pHLIP-1 peptide. A force constant of 4 kcal/mol/Å\(^2\) was used with a pulling velocity of \( 5 \times 10^{-5} \) Å/timestep. The unbinding simulations were conducted starting with the five orientations of pHLIP-1 as detailed above (Fig. 3).

*Umbrella sampling simulations.* After we completed the pulling simulation, we selected our starting coordinates for each US window, based on the distance between the COM of the peptide and COM of the bilayer. The starting configurations from the SMD trajectory were located at the center of each window along the reaction coordinate. Using the \( z \) axis as the reaction coordinate, 50 windows were chosen spanning from 21-70 Å with a 1 Å interval. The
For each US simulation, we applied a harmonic force with a force constant of 20 kcal/mol/Å$^2$ on the reaction coordinate to the center of each window. Simulations for each window were run for 32 ns, with the first 2 ns of simulation treated as an equilibration phase and excluded from analysis. The 1 Å spacing interval and the harmonic force constant of 20 kcal/mol/Å$^2$ were chosen to ensure significant overlap of the average coordinate distribution between adjacent windows, in order to provide sufficient sampling within a given region along the reaction coordinate.

All simulations were performed using NAMD2.9 (Phillips et al., 2005) in the isothermal-isobaric ($NPT$) ensemble with Nose´-Hoover temperature coupling (Hoover, 1985; Nosé and Klein, 1983) and the Parrinello-Rahman barostat (Parrinello and Rahman, 1981), set to 300 K and 1 bar, respectively. 2 fs was chosen as the time step. The CHARMM36 force field (Best et
was used to describe the peptide and bilayer interactions. Electrostatics were accounted for using the shift function with a Coulomb cutoff of 12 Å. Shift was used for van der Waals as well, with a switch distance of 10 Å and a pairlist distance of 13.5 Å. We used the particle mesh Ewald method (PME) for the computation of the electrostatic forces, and a grid spacing of 1.0 Å was applied.

**Analysis**

**Free energy calculation.** The potentials of mean force (PMFs) of binding of pHLIP-1 peptide to the membrane bilayer were calculated by applying the weighted histogram analysis method (WHAM) to US trajectories (Grossfield, 2011; Kumar et al., 1992; Roux, 1995). 500 bins were chosen, with a convergence tolerance of 0.00001 for the WHAM analysis. The statistical uncertainty at each bin was estimated using bootstrapping, with 100 bootstrap trials for each PMF.

To compute the free energy of binding from the PMF, we need to divide the system into bound and unbound states, after which the free energy change upon binding can be calculated as,

$$
\Delta G = W_{\text{bound}} - W_{\text{unbound}} \text{,} \quad (8)
$$

where $W_{\text{state}}$ is the PMF value of a particular RC. $W_{\text{bound}}$ is the PMF of the bound (reference) state and is defined as 0, so equation (8) can be simplified to $\Delta G = -W_{\text{unbound}}$.

**Details of pHLIP-1 binding.** All analyses was carried out on the bound state (the first window). Hydrogen bond occupancies were calculated using the hydrogen bonds plugin of VMD (Humphrey et al., 1996), with a distance cutoff of 3.0 Å and an angle cutoff of 60°. Contact analysis was carried out using a tcl script in VMD, counting the number of specific atoms within a sphere of radius 5 Å.
Electron density is defined as the measure of the probability of an electron being present at a specific location, which is usually used to locate atoms of different types. The “density-dist” tool in LOOS (Romo et al., 2014) was used to compute the electron density distribution for each component in our system along the z axis.

The molecular order parameter provides relevant information related to the disorder of the hydrocarbon region in the interior of the lipid bilayers by measuring the orientation of the hydrogen dipole of the methylene groups with respect to the bilayer normal. The equation can be described as,

\[ S_{CD} = \left( \frac{3 \cos^2 \alpha - 1}{2} \right), \]  

where \( \alpha \) is the angle formed between the acyl chain vector and the bilayer normal. The “dibmops” tool in LOOS (Romo et al., 2014) was used to obtain the molecular order parameters binned by lateral distance from the target (pHLIP-1), in order to obtain the distribution of lipid distortion upon peptide binding.
RESULTS AND DISCUSSION

_Potentials of mean force._ Initially PMFs were computed for the five orientations using 2 ns-long trajectories per window (Fig. 4). All of the PMFs possess a similar behavior, with binding leading to a free energy release, although the height of the plateau with respect to the bound state varies among the orientations. The PMF of the orientation with 0 deg has the highest value (~ 40 kcal/mol), while the 288 deg orientation has the lowest, and the 72, 144, and 216 deg orientations have similar shapes and maxima. However, the calculated PMFs were found to be unstable; a simulation time of 2 ns per window is too short to obtain convergence of the PMF in our system (Fig. 5). For example, for the 0 deg orientation, when the simulation time per window is decreased from 2 ns to 1.8, 1.6, 1.4, 1.2, and 1.0 ns, the PMF significantly increases. In a converged system the PMF will remain stable with an increase in simulation time. Therefore we extended the simulation in each window to 32 ns. Due to our limited computational resources, we were able to extend only three orientations (72, 144, and 288 deg) to the specified timescale.

![Figure 4. pHLIP-1 binding to bilayer surface is orientation-dependent. PMF for different orientations of pHLIP with respect to the principal axis of the peptide with simulation time of 2 ns/window. Distance is defined as the distance between COM of peptide and COM of lipids. The free energy barrier of binding is interpreted as the point at which the PMF plateaus (approximately 50 Å).](image-url)
These orientations were chosen because they represent three distinct binding configurations: 1) half embedded with the C-terminus in the lipid bilayer, 2) lying on the surface of lipids, and 3) half embedded with the N-terminus in the lipid bilayer.

After extending the simulation time for the US windows to 32 ns, the behavior and maxima of the PMFs changed for each orientation (Fig. 6). The windows beyond 58 Å from the COM of

Figure 5. PMF does not converge with 2 ns/window. Time-dependent PMFs of the orientation with 0 deg were obtained by calculating the PMF using different time lengths (1, 1.2, 1.4, 1.6, 1.8, 2 ns) for each window.

Figure 6. PMFs of extended simulations have lower barriers but still possess distinct behavior. PMF plots for three different orientations with the simulation time of 32 ns/window (first 2 ns were discarded for analysis).
the lipid bilayer were discarded because the PMFs had already plateaued (stabilized) at distances below 50 Å. The peptides with an orientation of 72 deg and 144 deg with have similar maxima around 13 kcal/mol, while the 288 deg orientation has the lowest at 10.8 kcal/mol. The different behavior at the lower distances of the 288 deg orientation may be the result of fewer constraining interactions with the bilayer, leading to a sudden break of the interactions between peptide and lipids.

Convergence analysis of each respective peptide orientation produced different results (Fig. 7). For the orientations of 72 deg and 144 deg, the PMF values continuously decrease as the simulation time increases, meaning the peptide-bilayer system has not sufficiently converged. These simulations are currently being extended to longer timescales in order to obtain

Figure 7. Only the 288 deg orientation of pHLIP-1 converges within 32 ns. The time dependent PMFs were obtained by plotting the PMF from US simulations with different time lengths (25, 26, 27, 28, 29, 30 ns) in each window for 72 deg (A), 144 deg (B), and 288 deg (C) orientations.
convergence. However, for the 288 deg orientation, the PMF has converged within 32 ns because there is no significant variation over a range of timescales (Fig. 7c).

**Free energy of binding.** To compute the free energy of binding from the PMF, the system needs to be separated into discrete bound and unbound states. In general, this unbound state is defined as the point where the PMF stabilizes at a maximum value, and the bound state is the first window with a PMF of 0 kcal/mol. This would lead to a free energy of binding of the 288 deg orientation of $\Delta G_{\text{bind}} = -10.8$ kcal/mol. If one assumes a similar release in free energy for insertion of pHLIP-1 compared to the wild-type pHLIP (-1.8 kcal/mol, (Reshetnyak et al., 2008)), then the value of the free energy of binding that we obtained for the 288 deg orientation is in excellent agreement with the thermodynamically determined value (-13.2 – (-1.8) = -11.4 kcal/mol, (Karabadzhak et al., 2012). This allows us to extract details of the atomistic interactions between pHLIP-1 and POPC that leads to peptide binding.

**Atomistic details of pHLIP-1 binding.** To determine the characteristics of pHLIP-1 binding to the lipid bilayer surface for each orientation, we analyzed the trajectories of the first window (window 0 = binding state). A specific focus was made on the interaction between pHLIP-1 and lipids/water, the location of the peptide and residues interacting with the lipid headgroup region, and the effect on lipids due to peptide binding.

**Table 1. Characteristics of hydrogen bonding$^a$ of pHLIP-1 bound to the POPC bilayer surface.**

<table>
<thead>
<tr>
<th>Hydrogen bond pair</th>
<th>72 deg</th>
<th>144 deg</th>
<th>288 deg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.70% (102)</td>
<td>0.03% (144)</td>
<td>23.83% (94)</td>
</tr>
<tr>
<td></td>
<td>0.20% (151)</td>
<td>0.72% (128)</td>
<td>2.20% (129)</td>
</tr>
<tr>
<td></td>
<td>0.21% (142)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln4-lipids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr8-lipids</td>
<td>1.03% (131)</td>
<td>10.51% (144)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.37% (155)</td>
<td>35.69% (128)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.37% (158)</td>
</tr>
</tbody>
</table>

---

$^a$ Values rounded to two decimal places.
<table>
<thead>
<tr>
<th>Residue</th>
<th>Lipids</th>
<th>Water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trp9</td>
<td>1.03% (131)</td>
<td>0.12% (143)</td>
</tr>
<tr>
<td></td>
<td>0.37% (155)</td>
<td>18.7% (89)</td>
</tr>
<tr>
<td>Arg11</td>
<td>n/a</td>
<td>28.17% (164)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>31.52% (104)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td>Tyr12</td>
<td>9.23% (89)</td>
<td>26.27% (142)</td>
</tr>
<tr>
<td></td>
<td>1.00% (144)</td>
<td>38.20% (158)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.80% (148)</td>
</tr>
<tr>
<td>Ala13</td>
<td>n/a</td>
<td>62.17% (128)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td>Trp15</td>
<td>16.5% (142)</td>
<td>0.06% (102)</td>
</tr>
<tr>
<td></td>
<td>13.97% (128)</td>
<td>2.0% (158)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1% (129)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.07% (144)</td>
</tr>
<tr>
<td>Leu16</td>
<td>34.8% (148)</td>
<td>n/a</td>
</tr>
<tr>
<td>Ala27</td>
<td>5.53% (131)</td>
<td>18.90% (137)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td>Trp9-water</td>
<td>31.37%</td>
<td>29.37%</td>
</tr>
<tr>
<td>Trp15-water</td>
<td>22.00%</td>
<td>0.87%</td>
</tr>
<tr>
<td>Asp14-water</td>
<td>88.97% / 88.97%</td>
<td>98.90% / 89.93%</td>
</tr>
<tr>
<td>Asp25-water</td>
<td>73.23% / 89.13%</td>
<td>97.97% / 97.80%</td>
</tr>
</tbody>
</table>

*a* Percent of time hydrogen bond is formed over the trajectory. 3 Å and 60 deg are chosen as the cutoff distance and angle to define the hydrogen bond.

*b* Numbers in parentheses are the residue IDs of POPC lipids interacting with tryptophan.

*c* Significant interactions between residues and lipids are highlighted in gray.

*d* The two values represent hydrogen bond occupancies for O1 and O2 in the aspartic acid, respectively, with water.

Hydrogen bonds play a significant role in stabilizing the interactions between pHLIP-1 and the bilayer surface. Significant hydrogen bonding interactions were identified between pHLIP-1 and lipids. For all three orientations, the region from residues 8 to 16 are highly involved in these interactions, especially the two tryptophan and tyrosine residues. Gln4 had significantly more interactions in the 288 deg orientation, due to the fully buried N-terminus, whereas Ala27 had more interactions for the 144 deg and 72 deg orientations, indicating that the C-terminus was buried in both 72 deg and 144 deg orientations. However, since the Try8-Leu16 region made major contributions to the binding of all three orientations, we chose to focus on the
two tryptophan residues Trp9 and Trp15, since they possess both hydrophilic and hydrophobic characteristics that can be exploited by different regions of the lipid bilayer.

Interactions of the tryptophan residues of pHLIP-1 with water are highly variable between the different orientations. Both Trp15 in the 144 deg orientation and Trp9 in the 288 deg orientation have minimal interactions with water, due to being deeply buried in the headgroup region of the lipids (Fig. 8). Other tryptophan residues are either close to the lipid interface or exposed to bulk solution. This distribution can be seen from the electron density of the system (Fig. 9), which can describe the one-dimensional location of each component in the system. Both Trp9 in the 288 deg orientation and Trp15 in the 144 deg orientation are most deeply embedded in the lipid bilayer. For the 72 deg orientation, the tryptophan residues are much closer to the bilayer interface and exposed to water. This solvent exposure is a direct result of the binding configuration of the peptide, in which the C-terminus of pHLIP-1 is bound to the bilayer surface.

To further investigate the local environment of pHLIP-1-lipid interactions, we analyzed the contacts that were formed between pHLIP-1 and the surrounding environment. Results were obtained by calculating the number of heavy atoms (lipids or water) within a sphere of radius 5 Å from the center of the body of interest (peptide or tryptophan residues) and classifying the fractional contribution of contacts between lipids and water. (Fig. 10). The contacts of pHLIP-1
Figure 9. Tryptophan binds to the headgroup region of the bilayer surface in an orientation-dependent manner. The electron density profile of the pHLIP-1-lipid system provides information about the location of each component. Electron density is shown for 72 deg (A), 144 deg (B), 288 deg (C) orientations. Inset (D): zoomed-in view of tryptophan residues for all pHLIP-1 orientations. $z$ is the distance along the $z$ axis, where $z = 0$ is the midplane of the bilayer.

did not show significant changes for both the 72 and 288 deg orientations, whereas the 144 deg orientation possessed a gradual transition from water-dominant interactions to lipid-dominant ones with C terminal half of peptide. Over the course of the simulation, the peptide tends to bury more residues within the headgroup region of the bilayer. In the context of the tryptophan residues, Trp9 gradually moved away from the bilayer surface whereas Trp15 was partitioning into the headgroup region. The 144 deg orientation initially had most residues lying on the surface of the bilayer, but as the simulation progressed, the Tyr8-Leu16 region gradually inserted into the lipids, leading to a deeply buried C-terminus. This behavior was also consistent for the 288 deg orientation.
Previous thermodynamics studies were able to investigate the distortion of the lipid bilayer upon binding and insertion of pHLIP by quantifying the number of lipid molecules that were directly or indirectly perturbed in States II and III (Reshetnyak et al., 2008). The disorder of lipids in our US simulations can be quantified by the molecular order parameter, which measures the orientation of the hydrogen dipole of the methylene groups of lipids with respect to the lipid bilayer normal. The molecular order parameter was calculated as a function of the lateral distance distribution along the surface of the bilayer (the $xy$ plane) from 0–30 Å to the COM of peptide (Fig. 11). Lipids that were proximal to pHLIP-1 were more disordered. For all three orientations, the lipid bilayer underwent significant distortion within 10 Å of the peptide. By comparing the different orientations, one can see that pHLIP-1 in the 288 deg orientation had the weakest distortion of the bilayer. This result has a direct correlation to the PMF of binding—lipid
distortion is a key factor that contributes to the difficulty in stabilization of the membrane-peptide interaction and the energetic cost of binding.

**CONCLUSION**

Based on our current studies, the 288 deg orientation of pHLIP-1 binding to POPC has the weakest distortion in the bound complex and gives a reliable estimate of the binding energy of -10.8 kcal/mol. Longer US simulations will be needed to accurately model the other orientations. However, the atomistic details of each of these binding events reveal a complex relationship between the three components of pHLIP-1 (N-terminus, interior, and C-terminus) that lead to distinct interactions with the lipid bilayer surface. Specifically, the region Try8-Leu16 significant contributions to the binding of pHLIP-1 for all three orientations, whereas binding of the C-terminus may lead to a peptide-lipid complex that is more conducive to the folding and insertion that takes place in pHLIP when the pH is lowered to acidic levels. We anticipate formulating more definitive conclusions upon extension of each of our US simulations.

**Figure 11. Lipid bilayer distortion occurs by varying degrees for different orientations of pHLIP-1 binding.** The molecular order parameters (MOPS) of the lipids show the extent of lipid distortion upon binding. MOPS were calculated according to the lateral distance along the surface of the bilayer from the center of peptide. Greater distortion occurs as the peptide lies closer to the bilayer surface.
REFERENCES


Grossfield, A. (2011). WHAM, Ver. 2.05: an implementation of the weighted histogram analysis method.


Acknowledgement

Foremost, I would like to express my deepest thanks to my advisor, Professor Blake Mertz. His patience, encouragement, and immense knowledge were key motivations throughout my master. His guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Justin Legleiter and Prof. Stephen Valentine, for their encouragement, insightful comments, and hard questions. My thanks to my friends and colleagues for the great time I had in our group. I enjoyed the atmosphere, their friendship, and their support. My thanks to, Dr. Jun Feng, Dr. Choongkeun Lee, Sadegh Faramarzi, and Chitrak Gupta. It was a pleasure to work with all these people and to benefit from their knowledge. Especially, I would like to thank Dr. Jun Feng for the helpful instructions to me on different problems, specifically during my learning process on computational chemistry.

A special mention for Dejun Lin, for his insightful and detailed discussions with me on the difficulties I had in my project.

I thank HPC at West Virginia University and Stampede at Texas Advanced Computing Center for their support.

Last but not least, I wish to thank my family who have always supported me, my parents Rongqi Ren, Fengling Wang, and most of all Xiang Gao for enjoying life together with me.